



**Universidade Federal de Santa Catarina
Programa de Pós-Graduação em
Engenharia de Produção**

**APLICAÇÃO DE UM MODELO DE
DESDOBRAMENTO GRADUADO
GENERALIZADO DA TEORIA DA RESPOSTA
AO ITEM – TRI**

Silvana Ligia Vincenzi Bortolotti

Orientador: Prof. Miguel Verdinelli, Dr

Florianópolis - 2003

Silvana Ligia Vincenzi Bortolotti

**Aplicação de um modelo de Desdobramento
Graduado Generalizado da Teoria da Resposta ao
Item - TRI**

Dissertação apresentada ao
Programa de Pós-Graduação em
Engenharia de Produção da
Universidade Federal de Santa Catarina
como requisito parcial para a obtenção
do grau de Mestre em
Engenharia de Produção

Orientador: Prof. Miguel Verdinelli, Dr.

Florianópolis
2003

Ficha Catalográfica

B739a Bortolotti, Silvana Ligia Vincenzi
Aplicação de um modelo de desdobramento graduado generalizado da teoria da resposta ao item - TRI / Silvana Ligia Vincenzi Bortolotti. – Florianópolis : [s.n.], 2003.
107 f. : il. ; 30 cm

Orientador : Prof. Dr. Miguel Verdinelli
Dissertação (Mestrado) - UFSC. Programa de Pós-Graduação em Engenharia de Produção. Florianópolis, 2003.
Bibliografia : f. 69-76

1. Testes e medidas educacionais. 2. Avaliação educacional – Modelos estatísticos. 3. Teoria clássica de medidas. 4. GGUM (Modelo de desdobramento graduado generalizado). I. Verdinelli, Miguel, orient. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia de Produção. III. Título.

CDD : 371.26
CDU : 37

Silvana Ligia Vincenzi Bortolotti

**Aplicação de um modelo de Desdobramento
Graduado Generalizado da Teoria da Resposta ao
Item - TRI**

Esta dissertação foi julgada e aprovada para
obtenção do grau de **Mestre em Engenharia de
Produção e Negócios com Ênfase em Estatística Aplicada no
Programa de Pós-Graduação em
Engenharia de Produção da
Universidade Federal de Santa Catarina**

Florianópolis, 25 de julho de 2003.

Prof. Edson Pacheco Paladini
Coordenador do Programa

BANCA EXAMINADORA

Prof. Dalton F. de Andrade, PhD
Universidade Federal de Santa Catarina

Prof. Miguel Verdinelli, Dr.
Universidade Federal de Santa Catarina
Orientador

Prof. Sandré G. Macedo - Dr
Universidade do Vale do Itajaí

Aos meus pais Vander e Edvirges
pela força, afeto e por todos os ensinamentos
Recebidos.
Ao meu esposo Valdair, companheiro de todos os
momentos.
Aos meus filhos Yuri Luiz e Ana Ligia
pela paciência e compreensão que tiveram nas
horas de ausência.

Agradecimentos

A Deus, que me concedeu a vida e a inspiração necessária
para chegar ao final deste trabalho.

À Universidade Federal de Santa Catarina – UFSC.

Ao orientador Prof. Miguel Verdinelli, pela orientação, apoio
e amizade, recebida durante esta dissertação de mestrado.

Aos professores: Dalton F. de Andrade, Sandré G. Macedo
integrantes da banca.

Ao Prof. James Roberts pelo material fornecido bem como
as valiosas sugestões e críticas que muito contribuíram
para a versão final deste trabalho.

Aos colegas Décio Antonio Baraviera e Rosely Antunes de
Souza pela amizade, apoio e companheirismo ao longo de
minha carreira acadêmica e principalmente durante a
realização deste curso.

À Universidade Estadual de Maringá.

A todos os colegas do curso de Mestrado que foram
amigos e companheiros os quais dividi bons e difíceis
momentos.

Ao Centro Federal de Educação Tecnológica do Paraná –
Medianeira pela pesquisa feita nessa instituição.

.

.

...

A todos que direta ou indiretamente
Contribuíram para a realização
desta pesquisa.

Resumo

BORTOLOTTI, Silvana Ligia Vincenzi. Aplicação de um Modelo de Desdobramento Graduado Generalizado da Teoria de Resposta ao Item – TRI.

2003. 107 f. Dissertação (Mestrado em Engenharia de Produção com ênfase em Estatística) – Programa de Pós – Graduação em Engenharia de Produção, UFSC, Florianópolis.

Palavras – chave: Teoria da Resposta ao Item, modelo de desdobramento GGUM, teoria clássica, grau de satisfação.

A avaliação como instrumento de medida ocupa um importante espaço nas organizações atentas às exigências crescentes do mundo atual. A Teoria da Resposta ao Item (TRI) fornece modelos para a avaliação. Esta Teoria originou-se na metade da década de 30, contudo só recentemente está sendo utilizada em diversas áreas. Aqui no Brasil está sendo aplicada na área educacional. A TRI constitui um conjunto de modelos que representam a relação entre a probabilidade de dar uma certa resposta a um item e o traço (s) latente (s), habilidade (s) ou proficiência (s) do indivíduo. A presente dissertação tem como objetivo principal apresentar o GGUM – Modelo de Desdobramento Graduado Generalizado para a avaliação de medidas. Este modelo foi projetado para analisar respostas binárias e graduadas baseadas numa relação de proximidade. Apresentou-se o conceito, suposições básicas os métodos de estimação dos parâmetros dos itens e as habilidades das pessoas. Por fim apresentou-se uma aplicação do modelo GGUM com os alunos do Curso de Tecnologia do CEFET – MD – PR para avaliar o grau de satisfação destes quanto ao curso. Os resultados obtidos na análise ilustram o grande potencial desse modelo, pois, com um número pequeno de itens, os resultados foram consistentes.

Abstract

BORTOLOTTI, Silvana Ligia Vincenzi. Aplicação de um Modelo de Desdobramento Graduado Generalizado da Teoria de Resposta ao Item – TRI.

2003. 107 f. Dissertação (Mestrado em Engenharia de Produção com ênfase em Estatística) – Programa de Pós – Graduação em Engenharia de Produção, UFSC, Florianópolis.

Words key: Item Response Theory, unfolding model, classical theory, satisfaction degree

The evaluation as a measure instrument has an important space in the attentive organization to growing demands of our current world. The Item Response Theory (IRT) gives models for the evaluation. The theory comes from the half of the 30 decade. However only recently it has been used in several areas. IRT appoints to a group of models that shows the relation between the probability of giving a correct answer to an item and one latent (s) trait (s), ability (s) or proficiency (s) of the person. This dissertation has as main objective to show the GGUM (The generalized graded unfolding model). This model has been projected to analyze binary and graduate answers based on a proximity relation. The concept has been showed, with the basic suppositions with estimate methods of parameters items about people's abilities. Finally it has shown one application of the model GGUM with the students from a Technology Course from CEFET – MD – PR to evaluate their satisfaction degree and relation about the course. Then the results obtained in the analyses can illustrate the great potential of this model. Therefore with a small number of items, the results were consistent.

Índice

Resumo	vii
Abstract	viii
Lista de Tabelas	xii
Lista de Figuras	xiii
1 INTRODUÇÃO	1
1.1 Considerações iniciais.....	1
1.2 Justificativa.....	2
1.3 Objetivos.....	4
1.1.1 Objetivo Geral.....	4
1.1.2 Objetivos Específicos.....	4
1.4 Estrutura da Dissertação.....	4
2 REVISÃO DE LITERATURA	6
2.1 Introdução.....	6
2.2 Teoria Clássica de Medidas.....	7
2.3 Teoria de Resposta ao Item – TRI.....	9
2.3.1 Introdução.....	9
2.3.2 Função de Informação do Item.....	10
2.3.3 Função de Informação do Teste.....	11
2.3.4 Escala de Medidas.....	11
2.3.5 Suposições do Modelo.....	12
2.3.6 Estimação.....	13
2.4 Modelo de Desdobramento Graduado Generalizado – GGUM.....	14
2.4.1 Introdução.....	14
2.4.2 Modelo de Desdobramento Graduado Generalizado – GGUM.....	17
2.4.3 Função de Informação do Teste e do Item.....	23
2.4.4 Suposições do Modelo.....	25
2.4.5 Ajuste do Modelo.....	25

2.4.6 Os Efeitos dos Parâmetros de discriminação e dos limiares em função do Valor Esperado.....	26
2.4.7 Estimação.....	28
2.4.7.1 Estimação dos Parâmetros de Item.....	28
2.4.7.2 Valores de Parâmetros de Item iniciais.....	34
2.4.7.3 Estimação do Parâmetro do Indivíduo.....	34
2.4.7.4 Recuperação dos Parâmetros do Modelo.....	35
2.4.7.5 Erros Padrões das Estimativas dos Parâmetros	35
2.4.7.6 Sensibilidade das estimativas de Parâmetros quanto à suposição de uma distribuição <i>priori</i> e o número de quadratura.....	38
2.4.8 Equalização de estimativas de Parâmetros do GGUM.....	38
2.5 CEFET – Centro Federal de Educação Tecnológica do Paraná.....	39
3 RECURSOS COMPUTACIONAIS	41
3.1 Introdução.....	41
3.2 GGUM2000.....	42
3.2.1 Processo de Estimação do GGUM2000.....	42
3.2.2 Pontos de Quadratura.....	42
3.2.3 Erro Padrão na estimação dos Parâmetros.....	43
3.2.4 Limitações do Programa.....	43
3.2.5 Comandos do Programa.....	43
3.2.6 Produção do GGUM2000.....	44
4 METODOLOGIA	46
4.1 Introdução.....	46
4.2 Coleta de Dados.....	46
4.3 Desenvolvimento do Banco de Itens.....	48
4.3.1 Dimensionalidade.....	48
4.3.2 Seleção de Itens para a escala final.....	50
4.4 Estimativas dos Parâmetros dos Itens.....	51
4.5 Estimativas do Parâmetro θ_j	52
4.6 Ajuste do Modelo.....	53
4.7 Interpretação dos Parâmetros dos Itens.....	56

5 CONCLUSÃO E RECOMENDAÇÕES	66
------------------------------------	-----------

REFERÊNCIAS.....	69
-------------------------	-----------

APÊNDICES

APÊNDICE A – Questionário

**APÊNDICE B – Alguns Resultados dos Parâmetros dos Itens do modelo
GGUM segundo o software GGUM2000**

APÊNDICE C – Gráficos das Funções de Probabilidade dos Itens

Lista de Tabelas

4.1 Descrição dos alunos do CEFET - MD que freqüentam os cursos de Tecnologia quanto à cidade de procedência.....	47
4.2 Extração dos autovalores através dos componentes principais.....	48
4.3 Matriz dos dois primeiros componentes e comunalidades dos itens.....	49
4.4 Distribuição de Freqüência das respostas observadas na escala final.....	50
4.5 Estimação dos parâmetros dos itens ($\hat{\delta}_i, \hat{\alpha}_i, \hat{\tau}_{ik}$) para 25 itens e a correlação (r_i) entre a respostas observadas e esperadas para cada item.....	51

Lista de figuras

2.1 Exemplo de um item segundo o modelo de desdobramento (processo de ponto ideal).....	15
2.2 Representação gráfica de categorias de respostas subjetivas.....	17
2.3 Função de probabilidade de um item de quatro categorias de resposta subjetiva em função de $\theta_j - \delta_i$ ($\alpha_i=1,0$; $\tau_{ik}=-1,3$; $-0,7$; $-0,3$; $0,0$; $0,3$; $0,7$; $1,3$).....	18
2.4 Função de probabilidade de um item hipotético de 4 categorias de respostas observáveis em função de $\theta_j - \delta_i$	21
2.5 Valor esperado de uma resposta observável para um item hipotético de categoria 4 em função de $\theta_j - \delta_j$	22
2.6 Função de informação do item do GGUM em função de $\theta_j - \delta_i$, α_i e τ_{ik}	24
2.7 Valor esperado de uma resposta observável para um item de três categorias de resposta em função de $\theta_j - \delta_i$ e α_i	27
2.8 Valor esperado de uma resposta observável para um item de três categorias de resposta em função $\theta_j - \delta_i$ e τ_{ik}	28
4.1 Descrição dos alunos do CEFET - MD que freqüentam os cursos de Tecnologia quanto ao sexo.....	47
4.2 Histograma das idades dos alunos do CEFET - MD.....	48
4.3 Gráfico de Normalidade dos parâmetros θ_j através do software MINITAB....	52
4.4 Médias observadas e esperadas das respostas dos itens em função das médias $\hat{\theta}_j - \hat{\delta}_i$	53
4.5 Médias observadas e esperadas das respostas aos itens em função das médias de $\hat{\theta}_j - \hat{\delta}_i$ do item 17.....	54
4.6 Médias observadas e esperadas das respostas aos itens em função das médias de $\hat{\theta}_j - \hat{\delta}_i$ do item 18.....	55
4.7 Médias observadas e esperadas das respostas aos itens em função das médias de $\hat{\theta}_j - \hat{\delta}_i$ do item 34.....	55

4.8 Função de Probabilidade das categorias de resposta observável do item	
4 em função de $\hat{\theta}_j - \hat{\delta}_i$	57
4.9 Função de Probabilidade das categorias de resposta observável do item	
23 em função de $\hat{\theta}_j - \hat{\delta}_i$	59
4.10 Função de Probabilidade das categorias de resposta observável do item	
28 em função de $\hat{\theta}_j - \hat{\delta}_i$	61
4.11 Função de Probabilidade das categorias de resposta observável do item	
15 em função de $\hat{\theta}_j - \hat{\delta}_i$	63
4.12 Distribuição dos parâmetros dos alunos e dos itens.....	64

1 INTRODUÇÃO

1.1 Considerações iniciais

Somos expectadores e protagonistas de uma extraordinária velocidade no desenvolvimento dos mais variados ramos do conhecimento humano, rapidez esta representada especialmente pelos sucessivos aprimoramentos e inovações nos campos científico e tecnológico.

Inseridos nesse contexto de mudanças e transformações técnicas, sociais e econômicas, acentua-se a importância de descobrir novas metodologias que forneçam condições para que essas áreas se desenvolvam e a avaliação, certamente, tem sido um instrumento que representa a maior eficácia nessas transformações.

Avaliação, segundo Klein & Fontanive (2001), é um sistema de informações que tem como objetivo fornecer um diagnóstico e subsídios para manutenção e para prover um contínuo monitoramento de um sistema, como por exemplo, sistema educacional, sistema econômico, sistema religioso, sistema político, com vistas a detectar os efeitos positivos ou negativos de políticas adotadas.

Avaliar, no sentido de medir, abrange não tão somente a avaliação de desempenho, mas também medir a satisfação por um serviço, a preferência por determinado produto ou mesmo a avaliação de pessoas, por exemplo, através de atitudes, etc.

Uma destas metodologias que surgiu é a Teoria da Resposta ao Item – TRI que se desenvolveu para suprir deficiências da Teoria Clássica de Medidas – TCM.

Essa Teoria teve origem entre os anos 1935 e 1940, contudo, somente nos últimos trinta anos passou a ser explorada e aperfeiçoada na pesquisa de medida.

A Teoria da Resposta ao Item propõe modelos que representam a relação entre a probabilidade de dar uma certa resposta a um item e a habilidade (traço latente) de um indivíduo (VAN DER LINDEN & HAMBLETON, 1997).

Esta dissertação vem ao encontro com essa necessidade de introduzir novas metodologias no país, propondo apresentar um modelo estatístico para avaliação (medida) baseado na Teoria da Resposta ao Item – TRI com ênfase no modelo de desdobramento.

E como aplicação propõe-se utilizar o modelo de desdobramento da TRI, para avaliar a satisfação de alunos do CEFET – Medianeira - PR com relação aos cursos de Tecnologia oferecidos por esta instituição.

1.2 Justificativa

Pesquisadores da área de medida utilizam questionários ou testes com um conjunto de itens cujas respostas podem obedecer a uma escala segundo uma abordagem clássica de Thurstone (1928) ou com a popular escala de Likert (1932), que usa uma escala graduada de concordância, como, por exemplo, discordo fortemente, discordo, nem discordo nem concordo, concordo, concordo fortemente. Após terem sido coletadas as respostas dos indivíduos, elas são utilizadas para desenvolver estimativas da atitude destes.

Desta forma, isso permite a profissionais na área de medida ou psicólogos medir traços latentes de indivíduos definidos por um conjunto de itens, por exemplo, habilidade, atitude, satisfação, qualidade de vida, preferência por tipos de instrução, etc.

Geralmente essas respostas têm sido analisadas com modelos da TRI de natureza cumulativa, tais como modelos de um parâmetro (RASCH, 1960), de dois parâmetros (BIRNBAUM, 1968), de três parâmetros (LORD, 1980, ANDRICH, 1978), modelo de crédito parcial generalizado (MURAKI, 1992), entre outros.

Modelos cumulativos da TRI, implicam que em níveis mais altos de traços latentes, por exemplo, de níveis mais altos de habilidade, níveis mais altos de satisfação, deveriam, com toda probabilidade, gerar escores de itens mais altos, os quais, por outro lado, atuariam sobre os escores totais do teste, isto é, modelos onde a probabilidade de dar uma resposta de categoria mais alta aumenta quando a habilidade (traço latente) do respondente aumenta. Todavia, quando são estudadas respostas binárias e graduadas isso não ocorre freqüentemente.

Roberts (1995) em simulações de pesquisa verificaram que estes modelos não têm fornecido estimativas consistentes e que os modelos de desdobramentos nesses estudos têm sido mais adequados.

Os modelos de desdobramento são modelos de proximidade, onde escores de itens mais altos são mais prováveis (indicativo de níveis mais forte de concordância) quando a distância entre um indivíduo e um item num *continuum* latente subjacente diminui (COOMBS, 1964).

O uso de modelos de desdobramentos para uma medida de atitude se dá ao acreditar que um indivíduo concorda com uma declaração de atitude na medida em que o sentimento transmitido pela declaração combina com a emissão da própria opinião do indivíduo. Em termos de modelos de Teoria de Resposta ao Item, o grau de relação entre um indivíduo e um item é representado pela proximidade de um indivíduo ao item numa atitude latente hipotética num *continuum* que se estende de um ponto não favorável até um ponto favorável, assim, se as categorias de respostas forem codificadas progressivamente com sucessivos números inteiros (por exemplo, 1, 2, 3, 4, 5, 6), de acordo com níveis mais fortes de discordância até níveis mais fortes de concordância, então um modelo de desdobramento irá prever escores de itens maiores quando o item e o indivíduo em questão estejam perto um do outro, num *continuum* latente (ROBERTS *et al.*, 2000).

Foram desenvolvidos alguns modelos de desdobramentos tais como modelo Cosseno Hiperbólico de (ANDRICH & LUO, 1993), modelo Parella (HOIYTINK 1990), o Modelo de Desdobramento Generalizado Graduado conhecido por GGUM desenvolvido por Roberts *et al.* (2000), entre outros.

O GGUM tem como aplicação típica situações de medidas onde os entrevistados são solicitados a indicarem seu nível de concordância a um conjunto de itens que se situam num intervalo *continuum* bipolar, ou seja, que variam o seu conteúdo, do negativo, para o positivo, passando pelo neutro.

O GGUM é uma generalização do modelo de desdobramento graduado GUM de Roberts (1995) e Roberts & Laughlin (1996ab) e foi desenvolvido para respostas de questionários ou testes do tipo binárias e graduadas. Este modelo implementa relações de proximidades que predizem escores de itens e, conseqüentemente, escores totais baseando-se na distância entre um dado indivíduo e cada item em questão.

Tendo em vista que os modelos da TRI tornaram-se nos últimos 35 anos uma crescente ferramenta de medida, úteis na estimação de habilidade, e que vários modelos foram desenvolvidos para a estimação de habilidades pode-se, então, justificar o tema da pesquisa, procurando responder à seguinte questão:

...SERÁ QUE A APLICAÇÃO DO MODELO DE DESDOBRAMENTO DA TRI FORNECERÁ RESULTADOS CONSISTENTES QUANDO APLICADOS A UM ESTUDO DE SATISFAÇÃO?...

Para a aplicação da Teoria, usaremos o modelo para medir o grau de satisfação dos alunos do CEFET – MD - PR com relação ao curso de Tecnologia que freqüentam.

A utilidade do modelo GGUM apóia-se na habilidade de estimar corretamente os parâmetros do modelo, e, portanto, é importante para o avanço desta classe de modelos na pesquisa educacional e de outras áreas.

Neste trabalho, na aplicação do GGUM, considerar-se-á apenas o modelo unidimensional.

1.3 Objetivos

1.3.1 Objetivo Geral

Contribuir para o melhor entendimento das metodologias do Modelo de Teoria de Resposta ao Item, através da apresentação e aplicação do Modelo de Desdobramento Graduado Generalizado.

1.3.2 Objetivos Específicos

Fornecer embasamento teórico referente ao modelo de desdobramento da TRI.

Aplicar o modelo de desdobramento Graduado Generalizado com os alunos dos cursos de Tecnologia do CEFET – MD – PR.

Verificar relação entre a satisfação dos alunos e os cursos de Tecnologia oferecidos pelo CEFET- MD – PR.

Estimar os parâmetros do modelo de desdobramento e avaliar o grau de satisfação dos alunos.

1.4 Estrutura da Dissertação

A introdução apresentada no Capítulo I reflete questão relacionada com a problemática em que se baseou esta dissertação, além da justificativa que levou à investigação deste tema. Ainda no Capítulo I definem-se os objetivos, geral e específico, estabelecendo as intenções e os limites no desenvolvimento desta dissertação.

A seguir, no Capítulo II, apresenta-se a revisão de literatura em que se mostram os modelos matemáticos da TRI bem como o modelo de desdobramento graduado generalizado GGUM, no que diz respeito ao conceito, suposições básicas, os

métodos de estimação dos parâmetros dos itens e habilidades dos respondentes e a forma iterativa dos mesmos.

O Capítulo III contém uma descrição dos recursos computacionais existentes que viabilizam a aplicação da TRI e GGUM.

O Capítulo IV apresenta o material e a metodologia e resultados utilizados neste trabalho.

Os materiais utilizados para a leitura, armazenamento, conferência e resultados da pesquisa são:

- a) questionário da pesquisa
- b) arquivos de dados
- c) programa GGUM2000, Excel e MINITAB.
- d) Tabelas, gráficos e Figuras dos resultados da pesquisa.

No Capítulo V encontra-se a conclusão, recomendações bem como sugestões de futuras pesquisas. Na seqüência, apresentam-se as referências bibliográficas, na parte final do trabalho, expõem-se os apêndices.

2 REVISÃO DE LITERATURA

Neste Capítulo apresenta-se um breve histórico, as características da Teoria Clássica de Medidas e da Teoria da Resposta ao Item - TRI, tais como, função de informação do item, escala de medidas, processos de estimação dos parâmetros, o modelo de crédito parcial generalizado utilizado para determinar o modelo de desdobramento graduado generalizado GGUM, as características do modelo de desdobramento, as suposições básicas que norteiam o modelo GGUM, a estimação dos parâmetros do modelo GGUM, ajuste do modelo e equalização de estimativa dos parâmetros do modelo.

2.1 Introdução

Em muitos estudos, algumas variáveis de interesse não podem ser medidas diretamente, tais como: habilidade em determinado conteúdo na avaliação educacional, grau de satisfação do consumidor em determinado produto, a predisposição a uma determinada anomalia genética. Essas variáveis são denominadas de variáveis latentes e são referidas por habilidade, proficiência em avaliação educacional ou por traço latente (TAVARES, 2001).

Essas variáveis apresentam características que não podem ser observadas diretamente, e, portanto, esses tipos de variáveis devem ser inferidos a partir da observação de variáveis secundárias que estejam relacionadas a ela (VALLE, 1999).

Desta forma podemos estabelecer relações entre o desempenho de um teste formado por vários itens (questões, dado pelo score) e a habilidade a ser medida (TAVARES, 2001).

O objetivo principal, na maioria dos questionários e testes educacionais, é inferir sobre a habilidade ou traço latente do examinando em determinada área de conhecimento. Segundo Lord (1980), é preciso que se tenha alguma informação sobre como essa habilidade determina sua resposta a um item. Assim, um modelo de resposta ao item se ajusta bem a esta situação, pois expressa a relação probabilística entre a performance de um examinando em um teste (dado pelo conjunto de itens) e sua habilidade, que é uma característica individual do examinando, não diretamente observável.

Segundo Assunção, (1999),

... esses modelos quantificam a probabilidade de um examinando acertar (ou errar) um item específico, baseado na sua habilidade, o que se dá através de uma relação que é descrita por uma função matemática chamada Curva Característica do Item (CCI) ou Função Resposta ao Item. Em geral, os dados que se aplicam aos modelos de TRI consistem de uma matriz de respostas a um teste, onde cada elemento dessa matriz corresponde à resposta de um examinando a um determinado item do teste.

Assim, alguns dos modelos propostos da TRI funcionam como se a resposta do aluno a cada item do teste fosse ou verdadeira ou falsa, estabelecendo, através de uma função logística, relação entre a habilidade do aluno e sua probabilidade de acertar o item (ASSUNÇÃO, 1999).

2.2 Teoria Clássica de Medidas

Um dos primeiros trabalhos que surgiram no sentido de inserir uma modelagem estatística para estimação das habilidades tiveram por base os escores individuais (total de pontos em testes) e foram feitos por Sperman (1904), com forma axiomática final devida a Novick (1966) ficando conhecida como Análise Clássica de Itens (TAVARES, 2001).

No modelo clássico são introduzidos dois construtos: um escore verdadeiro e um erro de medida. O escore verdadeiro para um indivíduo pode ser definido como um valor esperado dos seus escores em vários testes. O erro de medida pode ser definido como a diferença entre o escore verdadeiro e o observado. O modelo clássico assume que: os erros de medida são aleatórios com média zero e não correlacionados entre si e com os escores verdadeiros. Os escores observados e os erros de medida são linearmente relacionados (NOJOSA 2001).

A equação básica da TCM descreve a relação entre os escores observados do indivíduo, escores verdadeiros e o erro:

$$X = T + E$$

onde X é o escore observado, T é o escore verdadeiro (a habilidade) e E é um erro de medida.

Essa teoria utiliza, basicamente, estatísticas descritivas, coeficientes de correlação e proporções, para medir a qualidade dos itens, e quase nenhuma estatística inferencial, contudo fornece resultados úteis como a fórmula de Sperman-Brow e a fórmula-20 de Kuder-Richardson, ambas utilizadas para calcular a

fidedignidade de um teste (fidedignidade refere-se à estabilidade dos seus resultados, se um teste é aplicado inúmeras vezes ao mesmo grupo de indivíduos espera-se que os resultados sejam os mesmos) (LORD, 1980; VIANNA, 1987). Embora a Teoria Clássica tenha sido muito útil. Hambleton & Swaminatan (1985) citam várias limitações, como por exemplo: todas as suas medidas são dependentes das características dos examinandos que se submetem ao teste ou ao questionário; a dificuldade do item (proporção de indivíduos que acertam ao item) e a discriminação do item, que são usados para caracterizar a qualidade dos itens de um teste, dependem do grupo de indivíduos do qual elas foram obtidas e, portanto, tem seu uso restringido se os examinandos no pré-teste não são representativos da população (NOJOSA, 2001).

Outro problema é que os escores, o observado e o verdadeiro aumentam e diminuem dependendo da dificuldade do teste. Isto é, testes diferentes, com dificuldades e discriminação diferentes, produzem estimativas das habilidades diferentes (ASSUNÇÃO, 1999).

Outras duas limitações, descritas em Hambleton & Swaminathan (1985), do modelo clássico dizem respeito à suposição de erros padrão de medida e a definição de confiança ou fidedignidade de testes paralelos. Na verdade, a Teoria Clássica de Medidas tem como fundamento, para a avaliação e a seleção de indivíduos, o teste como um todo, não prevê nenhuma consideração de como os indivíduos respondem a um determinado item (NOJOSA, 1997).

Por essas razões, especialistas em medidas, principalmente na área de psicologia e educação, buscaram outras teorias alternativas para obter um modelo que atendesse aos seguintes quesitos:

- a) estatísticas de itens não dependentes do grupo;
- b) escores que não dependessem da dificuldade do teste para descrever as habilidades dos indivíduos;
- c) modelos que não requeiram testes estritamente paralelos para avaliar a confiança ou fidedignidade dos indivíduos;
- d) modelos que expressem antes o nível do item do que o nível do teste.

Esses e outros anseios foram resolvidos por uma outra estrutura de teoria de medida, conhecida como Teoria de Resposta ao Item (LORD 1980; HAMBLETON & SWAMINATHAN, 1985).

2.3 Teoria de Resposta ao Item – TRI

2.3.1 Introdução

Atualmente, em várias áreas do conhecimento, vem crescendo o interesse na aplicação de técnicas derivadas da Teoria de Resposta ao Item – TRI como, por exemplo, na área educacional (ANDRADE, 1999; ANDRADE *et al.*, 2000); medicinal (DeROOS & MEARES, 1998); psicossocial (GRANGER & DEUTSCH, 1998); marketing (BAYLEY, 2001); área de serviços (COSTA, 2001); na gestão pela qualidade total (ALEXANDRE *et al.*, 2001). Essa teoria surgiu como uma forma de considerar cada item particularmente, sem relevar os escores totais, assim, as conclusões não dependem propriamente do teste, mas de cada item que o compõe (TAVARES, 2001). A Teoria de Resposta ao Item sugere modelos para os traços latentes, propondo formas de representar a relação entre a probabilidade de um indivíduo dar uma resposta a um item e seus traços latentes ou habilidades, na área de conhecimento a ser avaliada ou verificada, os quais não podem ser observados diretamente (ANDRADE *et al.*, 2000).

Uma das grandes vantagens da Teoria de Resposta ao Item é que ela possibilita fazer comparações entre habilidade de indivíduos de populações diferentes quando são submetidos a testes que tenham alguns itens comuns ou permite, ainda, a comparação de indivíduos de mesma população submetidos a testes totalmente diferentes. Isso é possível porque a TRI tem como elementos centrais os itens e não a prova como um todo (VALLE, 1999).

Costa (2001) traz um relato histórico do desenvolvimento da TRI desde 1936 até os tempos de hoje.

Os diversos modelos de respostas ao item existentes se distinguem na forma matemática da função característica do item e/ou no número de parâmetros especificados no modelo. Todos os modelos podem conter um ou mais parâmetros relacionados ao indivíduo. Podem ser vistos detalhes de diversos modelos existentes em Van Der Linden & Hambleton (1997) e Andrade, *et al.*, (2000).

De uma forma geral, os modelos da TRI podem ser organizados na seguinte maneira (NOJOSA, 2001):

- Modelos lineares ou não lineares;
- Modelos unidimensionais ou multidimensionais em relação ao traço latente;

- Modelos de resposta dicotômica ou politômica;
- Modelos para uma ou mais de uma população.

A TRI vem tornando-se numa técnica muito utilizada no campo de testes em vários países. No Brasil, a TRI foi aplicada pela primeira vez em 1995 na análise dos dados do Sistema Nacional de Ensino Básico – SAEB (ALEXANDRE *et al.*, 2001). Uma lista de aplicações da TRI no Brasil em avaliações educacionais pode ser visto em Andrade & Klein (1999).

Esta teoria inovadora baseada nos itens se apóia em dois pilares (NOJOSA, 2001):

- O desempenho de um indivíduo em um teste pode ser explicado por um conjunto de fatores chamados de traços latentes ou habilidades;
- O relacionamento entre as respostas dos indivíduos a cada item e a habilidade medida pelo teste pode ser representada por uma função monótona crescente, chamada Função Característica do Item (CCI). Esta função fornece a probabilidade de indivíduos de vários níveis de habilidade darem uma certa resposta a um determinado item.

Para se aplicar a TRI a um conjunto de dados, inicialmente, deve-se escolher um melhor modelo que seja adequado ao tipo de questionário; a seguir, deve-se obedecer a alguns passos e, entre eles verificar as suposições básicas e se há necessidade de equalização (ver Seção 2.4.8, pág.38) e fazer a estimação dos parâmetros. Para isso deve-se escolher um software apropriado para obter as estimativas e, após a fase de importância primordial, que é a interpretação dos resultados obtidos, discutir a validade dos resultados obtidos com a aplicação da TRI.

2.3.2 Função de Informação do Item

A função de informação do item é muito utilizada em conjunto com a CCI. Ela permite analisar quanto um item (ou pesquisa) traz de informação para a medida de habilidade (VALLE, 1999). A função de informação de um item é dada por:

$$I_i(\theta_j) = \frac{\left[\frac{d}{d\theta_j} P_i(\theta_j) \right]^2}{P_i(\theta_j) Q_i(\theta_j)}$$

onde:

$I_i(\theta_j)$ é a “informação” fornecida pelo item i no nível de habilidade (θ_j) ;

$$P_i(\theta_j) = P(X_{ij} = 1 | \theta_j) \quad \text{e}$$

$$Q_i(\theta_j) = 1 - P_i(\theta_j).$$

2.3.3 Função de informação do Teste

A informação fornecida pelo teste (ou pesquisa) é a soma das informações fornecidas pelos itens que compõem o mesmo:

$$I(\theta_j) = \sum_{i=1}^n I_i(\theta_j)$$

Outra maneira de representar esta função de informação do teste (ou pesquisa) é através do erro padrão de medida, chamado na TRI de erro padrão de estimação (EP). A $I(\theta_j)$, na verdade, é o quadrado do inverso desse erro:

$$I(\theta_j) = \left[\frac{1}{EP(\theta_j)} \right]^2$$

Isolando EP,

$$EP(\theta_j) = \frac{1}{\sqrt{I(\theta_j)}}$$

Analogamente ao erro padrão de medida da Teoria Clássica, o EP permite estabelecer intervalos de confiança em torno das habilidades (θ_j) dos indivíduos, contudo na TRI o EP (θ_j) é dependente de θ_j

2.3.4 Escala de Medidas

A habilidade pode teoricamente assumir qualquer valor real entre $-\infty$ e $+\infty$ ao contrário da medida escore em um teste (ou pesquisa) com n questões do tipo dicotômico (certo/errado), que assume valores inteiros entre 0 e n . Desta forma, é necessário estabelecer antes do processo de estimação uma origem que representará a média das habilidades e uma unidade de medida que representará o desvio padrão das habilidades dos indivíduos da população em estudo. Na prática utiliza-se uma escala com média $\mu = 0$ e desvio padrão $\sigma = 1$, denotada por $(0; 1)$.

Um problema que surge quando não é fixada uma escala para estimar as habilidades e os parâmetros dos itens é a Indeterminância ou não-identificabilidade. Este problema acontece principalmente devido à geração de um mesmo valor para a

função característica do item a partir de diferentes conjuntos de parâmetros. Maiores detalhes encontram-se em Nojosa (2001).

Quando se deseja a estimação de parâmetros em modelos envolvendo mais de uma população, é necessário para contornar o problema de indeterminância, além de fixar escalas de medidas ou distribuição *a priori* para as habilidades, definir também uma das populações como população de referência, assim todos os parâmetros serão medidos na mesma escala da população de referência (MATOS, 2001). Essa exigência constitui um dos importantes conceitos da TRI que é denominado de Equalização. Isso significa colocar os parâmetros de itens vindos de testes diferentes ou habilidades de indivíduos de diferentes populações em uma mesma escala, numa mesma métrica.

Um detalhe importante na TRI é que as habilidades são estimadas a partir das respostas a um conjunto de itens. É uma característica primordial da TRI conhecida como Princípio da Invariância, que diz: as estimativas das habilidades não se modificam mediante a apresentação de um novo conjunto de itens e, por outro lado, as estimativas dos parâmetros dos itens não se alteram ao se aplicar os mesmos itens a populações distintas (HAMBLETON & SWAMINATAN, 1991). Desta forma, fica evidente que a habilidade não depende do conjunto de itens ao qual o indivíduo é submetido, possibilitando, assim, comparação de itens, testes ou desempenhos de diferentes grupos de indivíduos.

2.3.5 Suposições do modelo

Unidimensionalidade

O modelo proposto pressupõe a unidimensionalidade do teste (ou da pesquisa), ou seja, a homogeneidade do conjunto de itens que supostamente deve estar medindo um único traço latente, uma só habilidade deve ser responsável pela realização de todos os itens do teste ou da pesquisa.

Entretanto, para satisfazer o postulado da unidimensionalidade, é suficiente admitir que haja uma habilidade dominante (um fator dominante), responsável pelo conjunto dos itens. Este fator é o que se supõe estar sendo medido pelo teste ou pesquisa (VALLE, 1999).

Geralmente, a dimensionalidade da pesquisa pode ser verificada através da Análise Fatorial, feita a partir da matriz de correlações tetracóricas. Quando se aplica

a Análise Fatorial, técnica estatística multivariada, o que se pretende é identificar os fatores mais importantes ou com predominância mais forte, de forma que se possa atuar nesse sentido quanto à habilidade do indivíduo. Tipicamente, a Análise Fatorial é feita a partir da matriz de correlação de Pearson, contudo quando as variáveis envolvidas na análise são do tipo dicotômicas então se usa a correlação tetracórica (ANDRADE *et al.*, 2000; MATOS, 2001).

Pode-se usar a proposta por Reckase (1979), segundo a qual se o primeiro autovalor da matriz de correlação tetracórica é dominante, então apenas uma habilidade é medida pelo teste.

Mislevy & Bock (1990) discute as deficiências da aplicação desse procedimento e sugere um outro procedimento, baseado no método da máxima verossimilhança.

Independência Local

Uma outra suposição do modelo é a chamada independência local ou independência condicional, a qual assume que, para uma dada habilidade as respostas aos diferentes itens da prova são independentes. Essa suposição é fundamental para o processo de estimação dos parâmetros do modelo. Na realidade, a independência é vista como conseqüência da correta determinação da dimensionalidade dos dados, então, tem-se somente uma e não duas suposições a serem verificadas, conforme Lord (1980). Assim, os itens devem ser elaborados de modo a satisfazer a suposição de unidimensionalidade (VALLE, 1999).

2.3.6 Estimação

Uma das etapas mais importantes da TRI é a estimação dos parâmetros dos itens e das habilidades dos indivíduos. Vimos que a probabilidade de uma resposta correta a um determinado item depende da habilidade do indivíduo e dos parâmetros que caracterizam o item. Pesquisas têm revelado que cada indivíduo deve ser submetido a, pelo menos, 30 itens e o número de indivíduos para cada item deve ser de, pelo menos, 300, para se obter estimativas para os parâmetros do modelo com erros-padrão pequenos.

Nesse processo, depara-se com 3 situações de estimação:

- Conhecem-se as habilidades dos indivíduos e deseja-se estimar os parâmetros dos itens.
- Conhecem-se os parâmetros dos itens e deseja-se estimar as habilidades dos indivíduos.

- Não se conhece as habilidades dos indivíduos e nem os parâmetros dos itens; deseja-se então estimar ambos.

Em todas estas situações assume-se como verdadeiro o modelo proposto. E para cada situação existe um procedimento para a estimação dos parâmetros dos itens e/ou habilidades e esses procedimentos podem ser feitos a partir de métodos de máxima verossimilhança marginal (MVM) ou conjunta (MVC) e bayesianos EAP (Esperança a *Posteriori*) com a aplicação de métodos iterativos tais como de Newton-Raphson ou *Scoring* de Fisher para a resolução das equações.

2.4 Modelo de desdobramento Graduado Generalizado – GGUM

2.4.1 Introdução

Segundo Coombs (1964) um indivíduo escolhe uma determinada categoria de resposta a um item se a posição dele num *continuum* latente θ está perto do item, caracterizando assim, uma relação de proximidade entre o indivíduo e o item num *continuum* latente (processo de ponto ideal).

Muitos pesquisadores tais como Andrich (1996); Roberts *et al.*, (1999) e Van Schuur & Kiers (1994) verificaram que os modelos de desdobramentos são mais adequados para avaliar um conjunto de respostas binárias ou graduadas de questionários de atitudes, pois perceberam que elas geralmente resultam de processos de ponto ideal (COOMBS, 1964 *apud* ROBERTS *et al.*,1998), onde um indivíduo defende uma declarada atitude na medida que o sentimento expressado pela declaração se ajusta adequadamente com a opinião do indivíduo (ROBERTS *et al.*, 2001).

Segundo Roberts *et al.* (1998)

...este argumento significa que as respostas do tipo discordo-concordo são melhores analisadas através de modelos de desdobramento (isto é, de proximidade) que implementa uma função de resposta de pico simples("single-peaked").

De acordo com os modelos de desdobramentos, a probabilidade de um indivíduo endossar um item de atitude está em função da distância entre a locação da atitude do indivíduo e a locação do item num *continuum* de atitude latente subjacente.

A Figura 2.1 mostra um exemplo de um item hipotético: "Eu gosto de conversar tranqüilamente com um amigo em um bar". Um indivíduo quando questionado sobre

este item pode discordar deste item ou porque ele é muito introvertido (incomoda-se em lugares públicos), ou é muito extrovertido (para ele conversar em um bar é tedioso),

O modelo de desdobramento coloca juntamente num *continuum*, indivíduos e itens que variam de uma posição negativa, neutra e positiva. Os indivíduos são localizados nesse *continuum* de acordo com sua opinião e os itens são localizados nesse mesmo *continuum* de acordo com seu conteúdo.

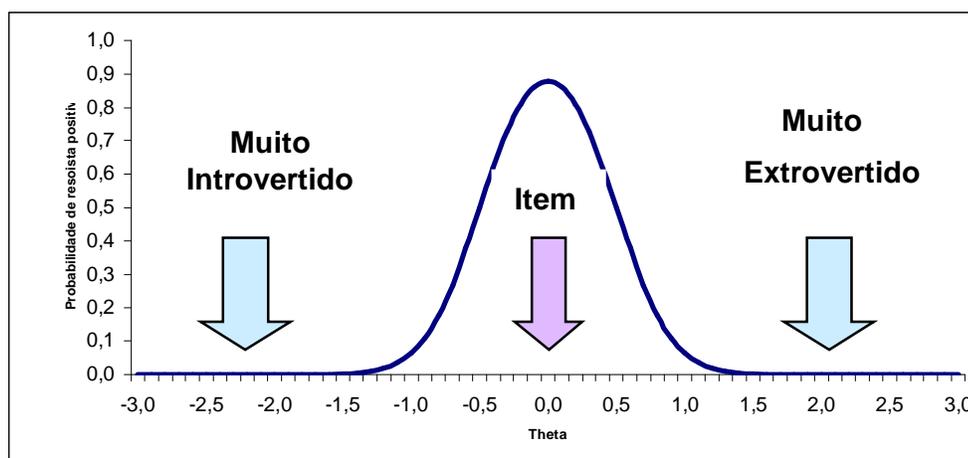


Figura 2.1 – Exemplo de um item segundo o modelo de desdobramento (processo de ponto ideal).

Vários modelos de desdobramentos de resposta ao item unidimensional foram desenvolvidos para medida de atitude.

Dentre esses modelos alguns são adequados para respostas binárias enquanto outros são apropriados para respostas graduadas. Os modelos para dados binários podem ter tanto abordagem paramétrica (ANDRICH, 1988; ANDRICH & LUO, 1993; DESARBO & HOFFMAN, 1987; VEHLEST & VERSTRALEN, 1993) quanto abordagens não paramétricas (CLIFF *et al.*, 1988; VAN SCHUUR, 1984). Há também modelos paramétricos para dados graduados (ROBERTS & LAUGHLIN, 1996ab) e modelos não paramétricos (CLIFF *et al.*, 1988; VAN SCHUUR, 1993).

As estimativas de atitudes de modelos paramétricos são invariantes aos itens utilizados para calibrar as estimativas e as estimativas de itens de locação são também invariantes para distribuição de atitudes na amostra, assim facilitando a aplicação de medidas como criação de banco de itens aspecto esse precioso dos modelos paramétricos de Hambleton *et al.*, 1991 e Lord, 1980 (ROBERTS *et al.*, 1998).

Essa dissertação enfoca modelos paramétricos de desdobramentos de itens para dados graduados.

2.4.2 Modelo de Desdobramento Graduado Generalizado – GGUM

O modelo de desdobramento graduado generalizado GGUM é um modelo de Teoria de Resposta ao Item unidimensional, desenvolvido para analisar respostas binárias como as graduadas baseadas numa relação de proximidade (COOMBS, 1964 *apud* ROBERTS *et al.*, 1998). Ele generaliza modelos de desdobramentos de resposta ao item em dois modos:

- Primeiro, ele implementa um parâmetro de discriminação que varia através dos itens e logo os itens são utilizados para discriminar entre os respondentes em modos diferentes.
- Segundo, o GGUM permite o uso de categoria de resposta diferencial através dos itens. Dessa forma ele age por implementação no limiar da categoria resposta que varia através dos itens.

Suponha que um indivíduo é solicitado a responder um questionário que visa avaliar uma atitude deste, como por exemplo, uma satisfação. Este questionário é dado por um conjunto de itens, e estes por sua vez são compostos por uma escala de respostas graduadas (por exemplo, discordo fortemente, discordo, concordo e concordo fortemente). Essas respostas são denominadas de categorias de respostas observáveis.

Porém, há um outro tipo de categoria de resposta. Segundo Roberts *et al.* (2001) um indivíduo pode responder a uma determinada categoria de resposta observável por dois motivos distintos. Por exemplo, um indivíduo pode discordar com um item porque seu conteúdo é tão negativo ou tão positivo em relação a sua própria opinião. Assim sendo, se o conteúdo do item é tão negativo, o indivíduo “discorda acima do item”, porque sua locação (posição) no *continuum* θ é acima da locação do item. Contudo, se o conteúdo do item for tão positivo, o indivíduo “discorda abaixo do item”, porque sua locação é abaixo do item. Em ambos os casos, estas categorias de respostas são denominadas de categorias de respostas subjetivas que o indivíduo pode usar.

Por exemplo, um item contendo duas categorias de respostas (discordo e concordo), logo há 4 categorias de respostas subjetivas:

- discordo abaixo;
- concordo abaixo;
- concordo acima;
- discordo acima.

A Figura 2.2 ilustra esta situação.

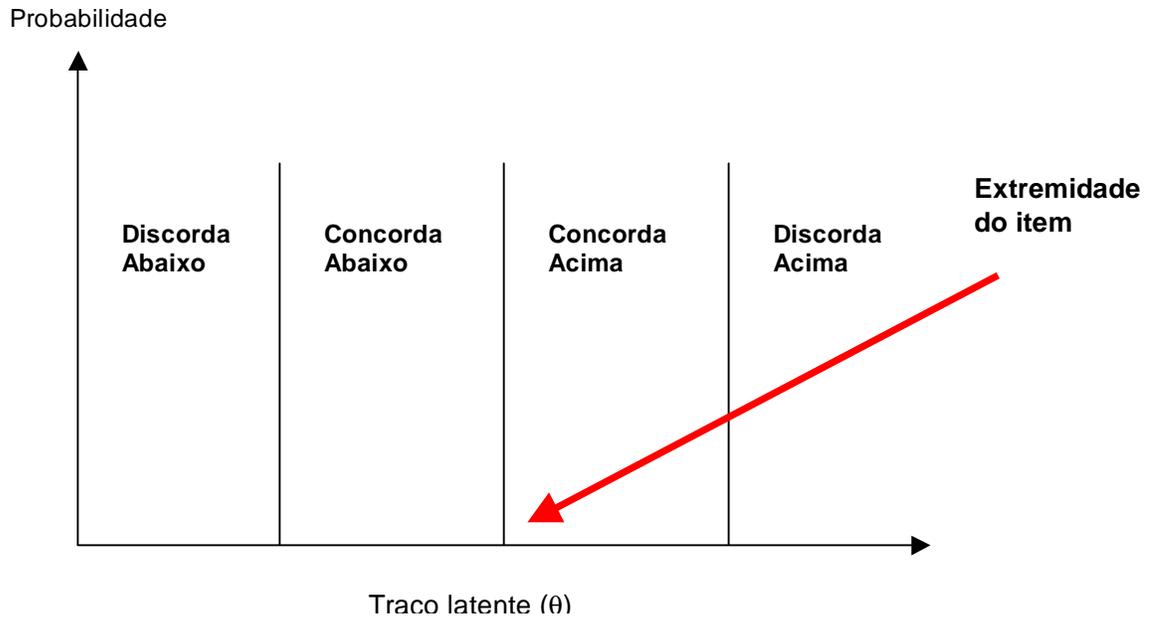


Figura 2.2 – Representação gráfica de categorias de respostas subjetivas

O modelo de desdobramento generalizado graduado se baseia no modelo de Credito Parcial Generalizado. Este modelo foi proposto por Muraki (1992) que se fundamentou no modelo de créditos parciais de Masters (1982), em que os parâmetros são de locação. Quando é aplicado às respostas subjetivas, o modelo de crédito parcial generalizado é dado por:

$$P[Y_i = y | \theta_j] = \frac{\exp \left[\alpha_i \left(y(\theta_j - \delta_i) - \sum_{k=0}^y \tau_{ik} \right) \right]}{\sum_{v=0}^M \exp \left[\alpha_i \left(v(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik} \right) \right]} \quad (2.1)$$

Com a seguinte restrição:

$$\sum_{k=0}^M \tau_{ik} = 0 \quad (2.2)$$

onde:

Y_i = uma resposta subjetiva à declaração de atitude i ;

$y = 0, 1, 2, 3, \dots, M$; $y = 0$ corresponde ao nível mais forte de discordância abaixo do item, enquanto que, $y = M$ corresponde ao nível mais forte de discordância acima do item (ver Figura 2.1, tomada de ROBERTS *et al.*, 2000);

θ_j = parâmetro de locação do indivíduo j num *continuum* de atitude;

δ_i = parâmetro de locação de um item i num *continuum* de atitude;

α_i = parâmetro de discriminação de um item i ;

τ_{ik} = parâmetro de locação do limiar de categoria de resposta subjetiva k num *continuum* de atitude relativa à posição do item i ;

M é o numero das categorias de respostas subjetivas menos 1.

Define-se como ψ distância entre os parâmetros limiaries.

O valor de τ_{i0} é arbitrário definido para ser zero na Equação 2.1, mas poderia ser ajustado para qualquer constante sem afetar o resultado das probabilidades (MURAKI, 1992).

A Figura 2.3 apresenta um item com quatro categorias de respostas observáveis: “discordo fortemente”, “discordo”, “concordo”, “concordo fortemente”.

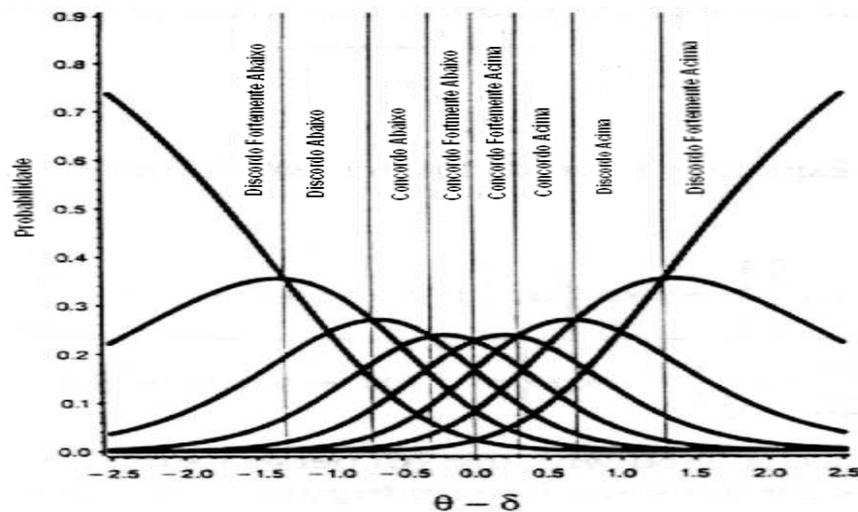


Figura 2.3– Função de Probabilidade de um Item com quatro categorias de resposta em função de $\theta_j - \delta_i$ ($\alpha_i=1,0$; $\tau_{ik}=-1,3$; $-0,7$; $-0,3$; $0,0$; $0,3$; $0,7$; $1,3$).

Fonte: Modificada de Roberts *et al.*, 2000.

Observa-se na abscissa da Figura 2.3 uma escala *continuum* dada pela distância entre a posição de atitude de um indivíduo j e a locação do item i ($\theta_j - \delta_i$). A ordenada representa as probabilidades que uma resposta do indivíduo cairá em uma

dos oitos categorias de respostas subjetivas, pois, há oito categorias de respostas subjetivas e funções de probabilidades associadas às respostas, devido ao fato de que um indivíduo possa responder quaisquer das 4 categorias de resposta observável porque sua posição de atitude está ora abaixo ou acima da posição (localidade) do item. As sete linhas verticais designam as locações onde as sucessivas funções de probabilidades das categorias de respostas subjetivas se interceptam. Essas locações são os limiares das categorias de respostas subjetivas. Neste exemplo, os sete limiares das categorias subjetivas são ordenados num *continuum* latente. Logo esses limiares dividem *continuum* latente em 8 intervalos nos quais uma resposta subjetiva diferente é mais provável. Dado um conjunto de parâmetros limiares, a predominância de uma resposta subjetiva mais provável dentro de cada intervalo é determinada pelo parâmetro de discriminação α_i , (ROBERTS *et al.*, 2000).

O GGUM foi desenvolvido a partir de quatro proposições básicas sobre o processo de resposta.

A primeira salienta que quando um indivíduo é solicitado para expressar a sua opinião de aceitação em uma declaração de atitude, o indivíduo tende a concordar com o item à medida que ele é localizado próximo de sua posição pessoal num *continuum* de atitude latente unidimensional. Dessa forma, o grau para o qual o sentimento de um item reflete a opinião de um indivíduo é dada pela proximidade do indivíduo ao item num *continuum* de atitude (ROBERTS *et al.*, 2000).

Se δ_i denotar a locação (posição) do item i num *continuum* e θ_j denotar a locação do indivíduo j no mesmo *continuum* então o indivíduo é mais tendente a concordar com o item à medida que a distância entre θ_j e δ_i se aproxima de zero. Isso é uma característica fundamental de um processo de ponto ideal (COOMBS, 1964 *apud* ROBERTS *et al.*, 1998).

A segunda proposição do modelo é que um indivíduo pode responder uma determinada categoria de resposta por dois motivos distintos. De acordo com Roberts *et al.* (1998), por exemplo, considere um indivíduo com uma opinião neutra em relação a “aborto”, então, este indivíduo quando interrogado sobre o assunto pode discordar fortemente com o item que retrata o assunto de aborto tanto de uma maneira muito negativa como também de uma maneira muito positiva. Se o item

está situado distante, debaixo da posição do indivíduo num *continuum* de atitude (isto é, o conteúdo do item é muito mais negativo que a atitude do indivíduo), então se diz que o indivíduo “discorda fortemente acima” do item. Em contrapartida, se o item estiver localizado muito acima da posição do indivíduo (ou seja, o conteúdo do item é muito mais positivo do que a atitude do indivíduo), então, diz-se que o indivíduo “discorda fortemente abaixo” do item.

Dessa forma, pode se dizer que há duas possíveis respostas subjetivas “discordo fortemente acima” e “discordo fortemente abaixo” associada com a única resposta observável de “fortemente discordo”.

O GGUM desdobra, então, duas respostas subjetivas para cada resposta observável numa escala de avaliação.

A terceira proposição do GGUM é que as respostas subjetivas às declarações de atitudes seguem um modelo de resposta ao item cumulativo (ANDRICH & LUO, 1993).

Neste trabalho assumiu-se o Modelo de Crédito Parcial Generalizado - MCPG de Muraki (1992) para as respostas subjetivas devido a sua generalidade, e que já foi apresentado anteriormente.

A Equação 2.1 do modelo de Muraki define um modelo de resposta ao item para níveis de respostas subjetivas. Cada categoria de resposta observável está associada com duas possibilidades de resposta subjetiva (isto é, um abaixo do item e outro acima do item).

As duas categorias de respostas subjetivas correspondentes a uma dada categoria de resposta observável são mutuamente exclusivas. Portanto, a probabilidade de um indivíduo responder usando uma categoria observável particular é simplesmente a soma das probabilidades associadas com as duas respostas subjetivas correspondentes:

$$P(Z_i = z | \theta_j) = P(Y_i = z | \theta_j) + P(Y_i = (M - z) | \theta_j) \quad (2.3)$$

Onde:

Z_i = uma resposta observável à declaração de atitude i ;

$z = 0, 1, 2, 3, \dots, D$; $z = 0$ corresponde ao nível de discordância mais forte e $z = D$ corresponde ao nível de concordância mais forte;

D = número de categorias de respostas observáveis menos 1. $M = 2 \times D + 1$.

A quarta e última proposição do GGUM é que os limiares das categorias subjetivas são simétricos em torno do ponto $(\theta_j - \delta_i) = 0$, a qual implica que:

$$\tau_{i(D+1)} = 0, e \quad (2.4)$$

$$\tau_{iz} = -\tau_{i(M-z+1)}, \text{ para } z \neq 0 \quad (2.5)$$

Essa proposição implica que os indivíduos estão bem propensos a concordar com um item localizado tanto em unidades $-p$ ou unidades $+p$ da posição do indivíduo num *continuum* de atitude. Essa proposição conduz à identidade:

$$\sum_{k=0}^z \tau_{ik} = \sum_{k=0}^{M-z} \tau_{ik} \quad (2.6)$$

Incorporando esta identidade na Equação 2.3 resulta na função de probabilidade do GGUM (ROBERTS *et al.*, 1998, ver pág. 67).

$$P(Z = z | \theta_j) = \frac{\exp\left[\alpha_i \left(z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik} \right)\right] + \exp\left[\alpha_i \left((M-z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik} \right)\right]}{\sum_{v=0}^D \left[\exp\left[\alpha_i \left(v(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik} \right)\right] + \exp\left[\alpha_i \left((M-v)(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik} \right)\right] \right]} \quad (2.7)$$

Assim o GGUM define uma função de probabilidade das categorias de resposta observável relacionadas com a resposta observável do indivíduo j ao item i . A Figura 2.4 mostra estas funções de probabilidade de respostas observáveis das categorias de respostas para um mesmo item referenciado na Figura 2.3. Cada uma dessas funções de probabilidade é a soma de duas funções de probabilidades das respostas subjetivas correspondentes mostrados na Figura 2.3 (ROBERTS *et al.*, 2000).

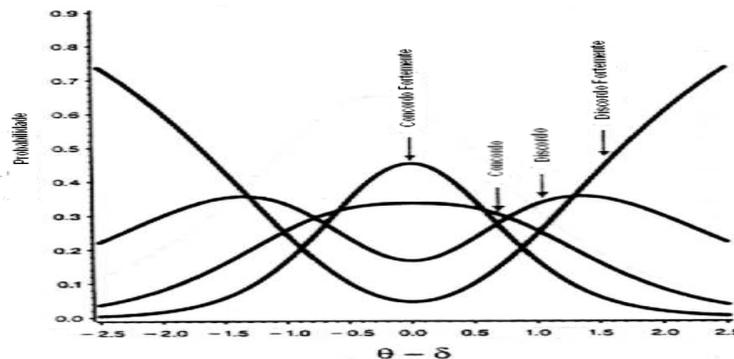


Figura 2.4 – Função de probabilidade de um item com quatro categorias de respostas observáveis em função de $\theta_j - \delta_i$.

Fonte: Modificada de Roberts *et al.*, 2000.

Observe que as funções de probabilidades das respostas observáveis não se interceptam em τ_{i1} , τ_{i2} , $\tau_{i3}, \dots, \tau_{iD}$. Assim, os parâmetros τ_{ik} perdem sua simples interpretação ao nível de resposta observável. Da mesma maneira, o parâmetro α_j indexa a discriminação para um nível de resposta subjetiva. Em contrapartida, a média de θ_j e δ_i não mudam quando se desloca de um nível de resposta subjetiva para observável. Um outro ponto com respeito aos parâmetros τ_{ik} é que eles não precisam ser sucessivamente ordenados num *continuum* latente. Esses valores simplesmente indicam onde a função de probabilidade de categoria de resposta subjetiva sucessiva se intercepta. É necessário, entretanto, que os valores máximos associados com a função de probabilidade de categoria de resposta subjetiva estejam ordenados na seqüência e este aspecto é garantido pelo modelo cumulativo associado com respostas subjetivas. Na prática, limiares desordenados poderão ocorrer sempre que uma ou mais categoria de respostas observáveis não sejam usadas freqüentemente pelos indivíduos (ROBERTS *et al.*, 2000).

O GGUM é um modelo de desdobramento de processo de resposta. Isto pode ser visto calculando-se o valor esperado de uma resposta observável a vários valores de $\theta_j - \delta_i$ usando a função probabilidade mostrada na Equação 2.7. A Figura 2.5 mostra o valor esperado de uma resposta observável para um item com 4 categorias de respostas. As categorias são codificadas com os valores inteiros de 0 a 3 onde cada número corresponde à resposta “discordo fortemente”, “discordo”, “concordo”, “concordo fortemente”, respectivamente.

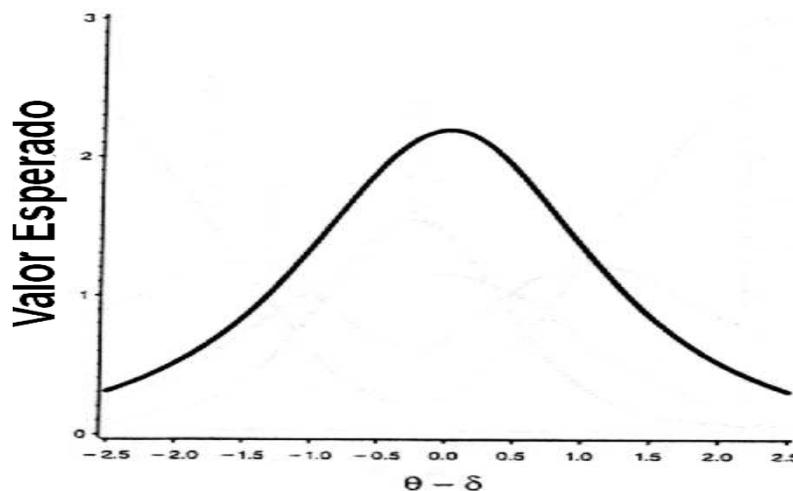


Figura 2.5 – Valor esperado de uma resposta observável de um item de quatro categorias de resposta em função de $\theta_j - \delta_j$.

Fonte: Modificada de Roberts *et al.*, 2000.

A Figura 2.5 mostra que o item apresenta níveis maiores de concordância à medida que a distância entre o indivíduo e o item num *continuum* de atitude diminui (ROBERTS *et al.*, 2000). A função do valor esperado do GGUM tem natureza de pico simples e esse pico se eleva intuitivamente porque os picos das funções de probabilidade das categorias de respostas subjetivas são ordenados num *continuum* latente e τ_{ik} s são simétricos em torno da origem $\theta_j - \delta_i = 0$. Isso foi confirmado por Donoghue (1999) que mostrou analiticamente que a função do valor esperado para o GGUM é sempre de pico simples e simétrico quanto ao ponto $\theta_j - \delta_i = 0$, exceto no caso em que $\alpha_i = 0$.

2.4.3 Função de Informação do Item e do Teste

A função de informação do item para o GGUM é igual a:

$$I_i(\theta_j) = -E \left[\frac{\partial^2 \ln(L)}{\partial \theta_j^2} \right]$$

$$I_i(\theta_j) = \alpha_i^2 \left[\left(\sum_{z=0}^D P(Z_i = z) \sigma_{Y_i|\theta_j, z}^2 \right) - \sigma_{Y_i|\theta_j}^2 \right] \quad \dots \quad (2.8)$$

e a função de informação do teste é igual a:

$$I(\theta_j) = \sum_{i=1}^I I_i(\theta_j)$$

$$= \sum_{i=1}^I \alpha_i^2 \left[\left(\sum_{z=0}^D P(Z_i = z) \sigma_{Y_i|\theta_j, z}^2 \right) - \sigma_{Y_i|\theta_j}^2 \right], \quad (2.9)$$

onde $\sigma_{Y_i|\theta_j, z}^2$ é variância condicional da resposta subjetiva do indivíduo j ao item i dado a resposta observável do indivíduo ao item i, e $\sigma_{Y_i|\theta_j}^2$ é a variância da resposta subjetiva ao item i do indivíduo j.

Mais detalhes da derivação das equações acima se encontram em Roberts *et al.*, (2000).

A Figura 2.6 mostra a função de informação do item para um item com seis categorias de respostas. O item apresenta $\delta_i = 0$ e tem como $\alpha_i = 1.0$. O gráfico superior da Figura 2.6 mostra como a informação de item varia com mudanças correspondentes na distância entre valores de τ_{ik} igualmente espaçados. A distância entre os valores sucessivos τ_{ik} , (isto é, a distância interlimiar foi definida como ψ) é mantida constante dentro de uma curva dada, mas o valor dessa distância muda

através das curvas. Especificamente, o gráfico mostra as funções de informação do item associadas com valores de ψ de 0,2; 0,4; 0,6 e 1,0. Observa-se que as funções de informação do item são todas bimodais e simétricas sobre a origem. Elas se aproximam de zero sempre que $|\theta_j - \delta_i|$ é igual a 0 ou é infinitamente grande. As informações são mais altas para valores menores de ψ , e os pontos (ou intervalos) sobre os eixos $\theta_j - \delta_i$ nas quais cada máximo ocorre torna-se mais distante da origem quando ψ aumenta. Além disso, as funções de informações tornam-se com pico menor quando ψ aumenta, e, dessa maneira, a distância dos valores de $|\theta_j - \delta_i|$ produzem informações próximas de quantidades máximas que ficam maiores com o aumento de ψ (ROBERTS *et al.*, 2000)

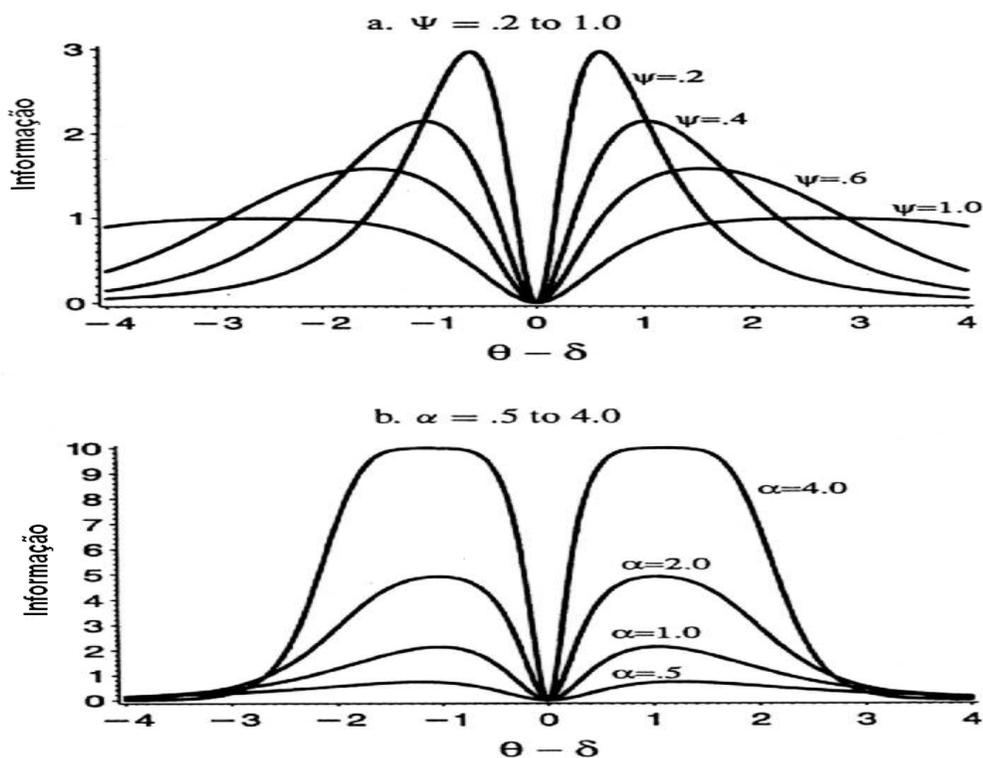


Figura 2.6 – Função de informação do item do GGUM em função de $\theta_j - \delta_i$, α_i e τ_{ik} . 4a. α_i é mantido constante enquanto a distância entre os sucessivos valores de τ_{ik} são variados. 4b. ψ é mantido constante enquanto os valores de α_i são variados.

Fonte: Modificada de Roberts *et al.*, 2000.

O gráfico inferior da Figura 2.6 mostra como a função de informação do item varia com mudanças ocorridas em α_i . O item nesta Figura tem também $\delta_i = 0$ e $\psi = 0.4$. As

curvas correspondem para valores α_i de 0,5; 1,0; 2,0 e 4,0. Como na parte superior, as funções de informação de item são bimodais, simétricos em relação à origem e aproxima-se de zero sempre que $|\theta_j - \delta_i|$ é igual a zero ou infinitamente grande. A quantidade máxima de informação é conseguida com valores maiores de α_i , e as funções de informação tornam-se mais pontuda quando α_i aumenta (ROBERTS *et al.*, 2000).

Quando se comparam os gráficos, superior e inferior da Figura 2.6 é visível que α_i , e τ_{ik} afetam a função de informação em modos distintamente diferentes. A função de informação torna-se maior e mais em pico quando α_i aumenta e torna-se menor e menor em pico quando ψ aumenta. Portanto, a precisão de medida será conseguida em dois pontos ou regiões simétricas num *continuum* latente e os itens com índices de discriminação grande e distância de interlimiaries pequenos produzirão mais precisão nestes pontos (ROBERTS *et al.*, 2000).

2.4.4 Suposições do Modelo - Unidimensionalidade

Esse modelo pressupõe a unidimensionalidade, isto é, o conjunto de itens deve estar medindo apenas uma habilidade. Para verificar a unidimensionalidade, como salientado em 2.3.5, a análise fatorial pode ser empregada verificar se o teste é unidimensional e também para identificar os itens que são menos propensos para ajustar à suposição de unidimensionalidade do GGUM.

Para verificar se um teste é unidimensional pode se utilizar o critério de Reckase (1979), segundo ele se o primeiro fator der conta de 20% da variância o teste pode ser considerado unidimensional.

Ao nível de item, Roberts *et al.* (2000) salienta que é necessário verificar as comunalidades dos dois primeiros componentes principais extraídos na análise fatorial. Um item é considerado unidimensional (para finalidade de análise) se sua comunalidade baseada nestes dois componentes for maior que ou igual a 0.3.

2.4.5 Ajuste do Modelo

Para verificar se os resultados obtidos através do GGUM se ajustam ao modelo, pode-se fazer, uma avaliação gráfica, onde respostas observadas e esperadas são plotadas em função da diferença entre θ_j e δ_i estimados, ou seja, em função de $\hat{\theta}_j - \hat{\delta}_i$ para determinar pontos num *continuum* latente em que os dois tipos de

respostas divergem (HAMBLETON & SWAMINATHAN, 1985). Para confeccionar este gráfico, deve-se computar todo par item-pessoa e distribuir as distâncias em grupos homogêneos de tamanhos iguais. Calculam-se, então, as médias das respostas observadas e esperadas para cada item do grupo. Essas médias são plotadas em função da média de $\hat{\theta}_j - \hat{\delta}_i$ para cada grupo.

A ocorrência de grandes discrepâncias no gráfico significa que o modelo não se ajusta aos dados.

Pode-se ainda calcular as correlações de Pearson entre as respostas observadas e esperadas para cada par de itens-pessoa ou entre suas médias, através de cada grupo $\hat{\theta}_j - \hat{\delta}_i$.

Além disso, pode-se utilizar o erro quadrático médio chamado de *infit* ou *outfit* para cada item (LINACRE & WRIGHT, 1994). Devem ser investigados os itens que possuem *infit* e/ou *outfit* maiores que 1 e que desviam substancialmente dos quadrados médios associados com a maioria dos outros itens (ROBERTS *et al.*, 2000).

2.4.6 Os Efeitos dos Parâmetros de discriminação e dos Limiares em Função do Valor Esperado

A forma da função do valor esperado no GGUM é determinada conjuntamente pelos parâmetros α_i e τ_{ik} . A Figura 2.7 mostra os efeitos de α_i numa função de valor esperado para as respostas de três categorias enquanto os valores de τ_{ik} permanecem constantes ($\tau_{i1} = -2$ e $\tau_{i2} = -1$). Cada gráfico da Figura 2.7 descreve a função para diferentes valores de α_i , isto é, de 0,5; 1,0; 1,5; 2,0; 10 e 30. Enquanto α_i aumenta, o valor máximo da função se aproxima de seu ponto mais alto e a função simultaneamente torna-se mais pontuda. Quando o α_i cresce sem um limite, a função do valor esperado da função se aproxima da função semelhante à de Guttman na qual a resposta torna-se totalmente determinada pela distância entre θ_j e δ_i (ROBERTS *et al.*, 2000).

Os efeitos dos valores de τ_{ik} em função do valor esperado são mostrados na Figura 2.8 e corresponde a uma função cuja escala de resposta é de três categorias. Os parâmetros τ_{ik} foram escolhidos de modo que a distância interlimiar seja igual

através de todo o *continuum* latente, e essa distância varia de 0,25 para 1,5 através das quatro ilustrações na Figura 2.8. O valor de α_i foi = 1 em todos os gráficos da Figura 2.8. Assim que a distância interlimiar aumenta, o máximo da função do valor esperado se aproxima do limite superior, embora a função fique menos íngreme. Este segundo aspecto é oposto ao encontrado quando α_i aumentou (ROBERTS *et al.* 2000).

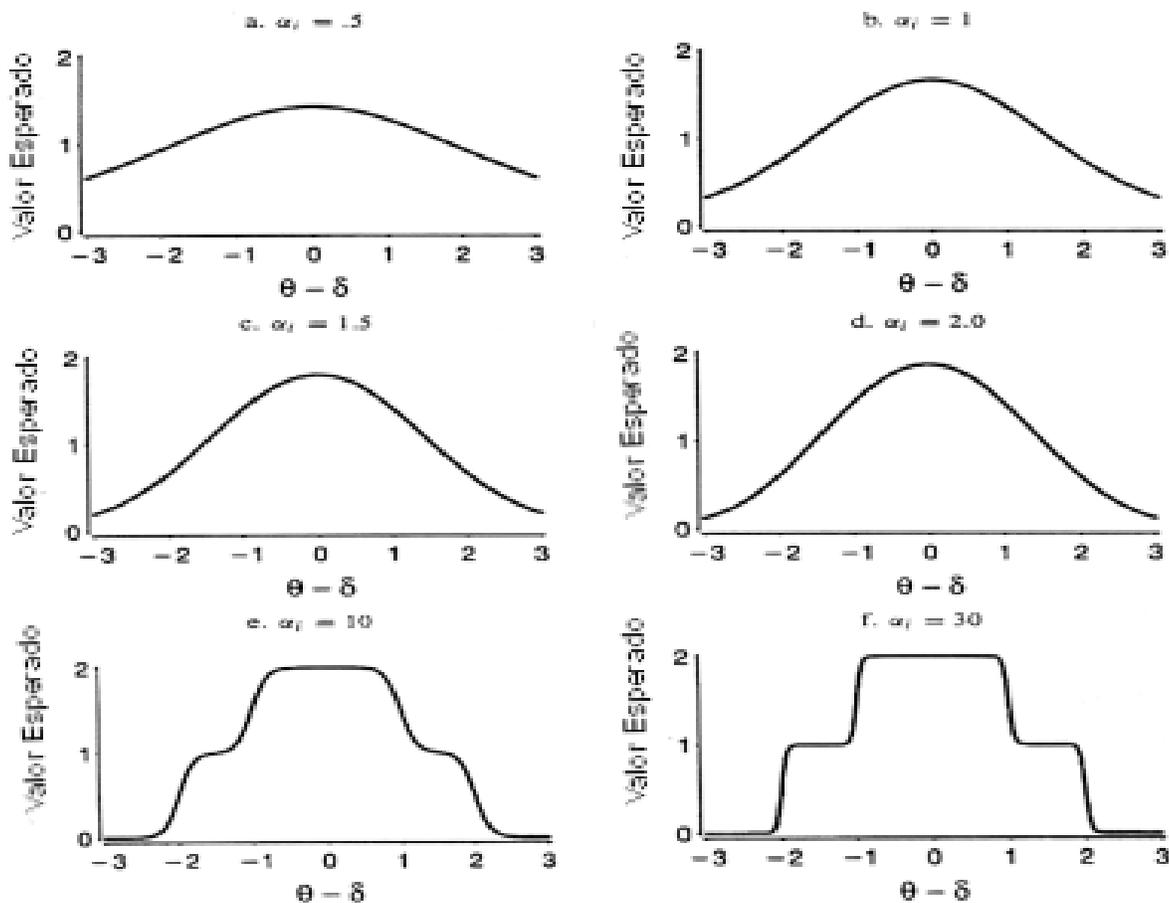


Figura 2.7 – – Valor esperado de uma resposta observável para um item de três categorias de resposta em função de $\theta_j - \delta_i$ e α_i

Fonte: Modificada de Roberts *et al.*, 2000

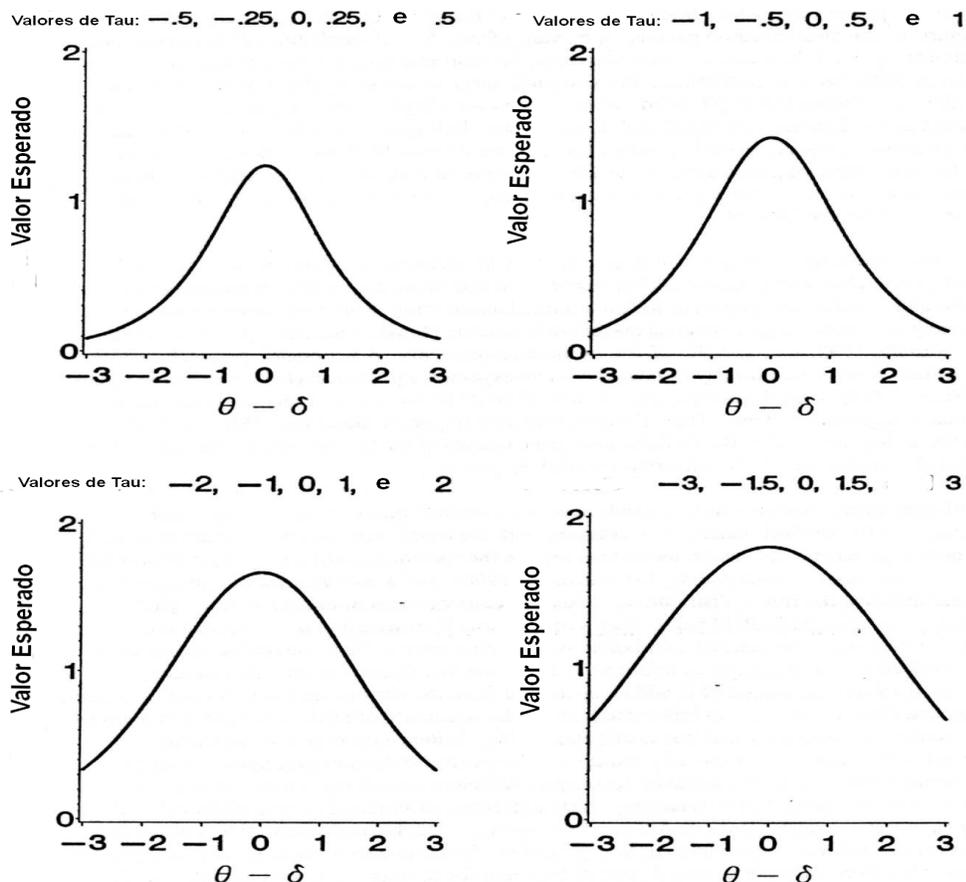


Figura 2.8 - Valor esperado de uma resposta observável para um item de três categorias de resposta em função $\theta_j - \delta_i$ e τ_{ik} .

Fonte: Modificada de Roberts *et al.*, 2000.

2.4.7 Estimação

2.4.7.1 Estimação dos Parâmetros de Item

Os parâmetros do item do GGUM são estimados usando a máxima verossimilhança marginal (MVM) (BOCK & LIEBERMAN, 1970; MURAKI, 1992 *apud* ROBERTS *et al.*, 2000). O algoritmo de solução faz um paralelo com o procedimento de Muraki (1992) baseado no algoritmo EM (Esperança e Maximização).

Seja X_s um vetor resposta, distinto de S para um dado conjunto de dados com $s = 1, 2, 3, \dots, S$, onde x_{si} é um elemento i de X_s . Sob a suposição de independência local, a probabilidade condicional observada em um vetor resposta particular X_s dado θ é igual a:

$$P(X_s | \theta) = \prod_{i=1}^I P(Z_i = x_{si} | \theta) \quad (2.10)$$

Se os indivíduos de uma população são amostrados com uma distribuição contínua denotada por $g(\theta)$, então a probabilidade marginal observada em um vetor resposta particular X_s é igual a:

$$P(X_s) = \int_{-\infty}^{+\infty} P(X_s | \theta) g(\theta) d\theta \quad (2.11)$$

Seja r_s o número de indivíduos com vetor resposta X_s , e N é o número de indivíduos na amostra. Nesta situação, r_s é distribuído multinomialmente com parâmetros N e $P(X_s)$, e a função de verossimilhança é igual a:

$$L = \frac{N!}{\prod_{s=1}^S r_s!} \prod_{s=1}^S [P(X_s)]^{r_s} \quad \dots \quad (2.12)$$

O logaritmo da função de verossimilhança é dado por:

$$\ln(L) = \ln(N!) - \sum_{s=1}^S \ln(r_s!) + \sum_{s=1}^S r_s \ln[P(X_s)] \quad \dots \quad (2.13)$$

As equações de verossimilhança para determinar α_i , δ_i , e τ_{ik} são obtidas calculando a derivada parcial de primeira ordem da Equação 2.13 com respeito a cada parâmetro e após fazendo as derivadas iguais a zero. Os valores de α_i , δ_i , e τ_{ik} que resolvem essas equações são as estimativas da máxima verossimilhança marginal (ROBERTS *et al.*, 1998).

A forma geral da derivada de primeira ordem do logaritmo da função de verossimilhança com respeito a um parâmetro de item particular, Φ_i , é dado por:

$$\begin{aligned} \frac{\partial \ln(L)}{\partial \Phi_i} &= \sum_{s=1}^S \frac{r_s}{P(X_s)} \frac{\partial P(X_s)}{\partial \Phi_i} \\ &= \sum_{s=1}^S \frac{r_s}{P(X_s)} \int_{-\infty}^{+\infty} \frac{\partial P(Z_i = x_{si} | \theta)}{\partial \Phi_i} \prod_{\substack{i'=1 \\ i' \neq i}}^I P(Z_{i'} = x_{si'} | \theta) g(\theta) d\theta \end{aligned}$$

$$\begin{aligned}
&= \sum_{s=1}^S \frac{r_s}{P(X_s)} \int_{-\infty}^{+\infty} \frac{\partial P(Z_i = x_{si} | \theta)}{\partial \Phi_i} \prod_{i=1}^I P(Z_i = x_{si} | \theta) \frac{g(\theta)}{P[Z_i = x_{si} | \theta]} d\theta \\
&= \sum_{s=1}^S \frac{r_s}{P(X_s)} \int_{-\infty}^{+\infty} \frac{\partial P(Z_i = x_{si} | \theta)}{\partial \Phi_i} P(X_s | \theta) \frac{g(\theta)}{P[Z_i = x_{si} | \theta]} d\theta
\end{aligned} \tag{2.14}$$

A Equação 2.14 pode ser aproximada usando quadratura de Gauss-Hermite como segue:

$$\begin{aligned}
\frac{\partial \ln(L)}{\partial \Phi_i} &= \sum_{f=1}^F \sum_{s=1}^S \frac{r_s L_s(W_f) A(W_f)}{\tilde{P}_s} \frac{\partial P(Z_i = x_{si} | W_f)}{\partial \Phi_i} \frac{1}{P(Z_i = x_{si} | W_f)} \\
&= \sum_{f=1}^F \sum_{z=0}^D \sum_{s=1}^S \frac{H_{siz} r_s L_s(W_f) A(W_f)}{\tilde{P}_s} \frac{\partial P(Z_i = z | W_f)}{\partial \Phi_i} \frac{1}{P(Z_i = z | W_f)} \\
&= \sum_{f=1}^F \sum_{z=0}^D \frac{\bar{r}_{izf}}{P(Z_i = z | W_f)} \frac{\partial P(Z_i = z | W_f)}{\partial \Phi_i}
\end{aligned} \tag{2.15}$$

onde:

$$L_s(W_f) = \prod_{i=1}^I P(Z_i = x_{si} | W_f), \tag{2.16}$$

$$\tilde{P}_s = \sum_{f=1}^F L_s(W_f) A(W_f), \tag{2.17}$$

$$\bar{r}_{izf} = \sum_{s=1}^S \frac{H_{siz} r_s L_s(W_f) A(W_f)}{\tilde{P}_s} \tag{2.18}$$

e H_{siz} é uma variável dummy que é igual a 1 quando $z = x_{si}$, em caso contrário é igual a 0. Na Equação 2.15, W_f é um ponto de quadratura (STROUD & SECREST, 1966 *apud* ROBERTS *et al.*, 1998), e $A(W_f)$ é re-escalado de $g(\theta)$ em W_f . A escala dos valores $A(W_f)$ é tal que:

$$\sum_f A(W_f) = 1 \tag{2.19}$$

$L_s(W_f)$ é a probabilidade condicional do vetor de resposta padrão X_s no ponto de quadratura W_f , \tilde{P}_s é a probabilidade marginal do vetor de resposta padrão X_s e \bar{r}_{izf} é a frequência esperada da resposta z para o item i no ponto de quadratura W_f . A Equação 2.15 inclui um componente específico do parâmetro, $\partial P(Z_i = z | W_f) / \partial \Phi_i$, que deve ser calculado separadamente em cada parâmetro para computar a derivada parcial de primeira ordem. A derivação desse componente encontra-se no apêndice do artigo de Roberts *et al.*, (2000).

A solução para as Equações 2.14 e 2.15 exige a especificação de uma distribuição *a priori* para os parâmetros θ_j , isto é $g(\theta)$, o qual permite que os θ_j sejam integrados fora da equação de solução quando se resolvem os parâmetros dos itens. A distribuição *a priori* fixada restringe a locação e a escala do *continuum* latente de modo que a solução matemática seja possível. Nenhuma outra restrição aos parâmetros de item é imposta. Para os modelos de Teoria de Resposta ao Item cumulativo, os pesquisadores selecionaram a distribuição normal como uma razoável aproximação para $g(\theta)$ quando não se conhece a forma da distribuição θ_j (BOCK & AITKIN, 1981; BOCK & LIEBERMAN, 1970; MURAKI, 1992; MISLEVY & BOCK, 1990). Entretanto, se uma informação *a priori* sobre a distribuição de θ_j é conhecida, então essa informação poderia ser usada para construir uma distribuição *a priori* baseada empiricamente (ROBERTS *et al.*, 2000).

O algoritmo EM descrito por Muraki (1992) é um procedimento iterativo para encontrar estimativas de máxima verossimilhança de parâmetros de modelos de probabilidade na presença de variáveis aleatórias não observáveis, chamadas de variáveis latentes. O E representa o passo em que se calcula a Esperança e o M representa o passo de Maximização. No modelo GGUM o EM é usado para resolver equação de verossimilhança para determinar os parâmetros α_i , δ_i e τ_{ik} . No estágio E do algoritmo, as estimativas das frequências esperadas de resposta z para o item i no ponto de quadratura W_f (\bar{r}_{izf}) são calculadas a partir das respostas observadas e as estimativas dos parâmetros de item temporário. No estágio M do algoritmo, as estimativas de \bar{r}_{izf} são tratadas como constantes conhecidas, e as equações de verossimilhança são, então, resolvidas. Pelas estimativas de \bar{r}_{izf} serem fixas, é possível resolver as equações de verossimilhança individualmente para cada item. O

estágio da maximização continua até que as estimativas de parâmetros de item para todos os itens tenham sido computados para um dado \bar{r}_{izf} . A complementação de um estágio da esperança simples seguido pelo estágio da maximização simples constitui um ciclo do algoritmo EM. Esse processo se repetirá até que algum critério de convergência seja satisfeito (sugestão, menos que 0,0005) (ROBERTS *et al.*, 2000).

O estágio da maximização do algoritmo EM acontece em duas etapas. Na primeira etapa, as equações de verossimilhança associadas com os parâmetros τ_{ik} são resolvidas para cada item individualmente. A solução é computada usando-se método de *scoring* de Fisher e, assim, uma matriz de informação para os parâmetros τ_{ik} é solicitada para cada item (ROBERTS *et al.*, 1998).

A matriz informação para um item i dado é::

$$\bar{I}_{\tau(i)} = \begin{bmatrix} \bar{I}_{\tau_{11}\tau_{11}} & \bar{I}_{\tau_{11}\tau_{12}} & \dots & \bar{I}_{\tau_{11}\tau_{1D}} \\ \bar{I}_{\tau_{12}\tau_{11}} & \bar{I}_{\tau_{12}\tau_{12}} & \dots & \bar{I}_{\tau_{12}\tau_{1D}} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{I}_{\tau_{1D}\tau_{11}} & \bar{I}_{\tau_{1D}\tau_{12}} & \dots & \bar{I}_{\tau_{1D}\tau_{1D}} \end{bmatrix} \quad (2.20)$$

Os elementos da matriz de informação são derivados em Rao (1973) são iguais a:

$$\bar{I}_{\tau_{ik}\tau_{ik'}} = \sum_{f=1}^F \bar{N}_{if} \sum_{z=0}^D \frac{1}{P(Z_i = z | W_f)} \frac{\partial P(Z_i = z | W_f)}{\partial \tau_{ik}} \frac{\partial P(Z_i = z | W_f)}{\partial \tau_{ik'}} \quad \dots \quad (2.21)$$

Onde \bar{N}_{if} é o número esperado de indivíduos no pontos de quadratura W_f que responderam ao item i :

$$\bar{N}_{if} = \sum_{z=0}^D \bar{r}_{izf} \quad (2.22)$$

O valor de \bar{N}_{if} é calculado no estágio da esperança do algoritmo e é mantido constante durante o estágio de maximização.

No método de *scoring* de Fisher, a função atualizada (*update*) usada para calcular os parâmetros τ_{ik} em q interações é dada por (ROBERTS *et al.*, 1998):

$$\begin{bmatrix} \tau_{i1} \\ \tau_{i2} \\ \vdots \\ \tau_{iD} \end{bmatrix}_q = \begin{bmatrix} \tau_{i1} \\ \tau_{i2} \\ \vdots \\ \tau_{iD} \end{bmatrix}_{q-1} + \left[\tilde{I}_{\tau(i)} \right]^{-1} \begin{bmatrix} \frac{\partial \ln(L)}{\partial \tau_{i1}} \\ \frac{\partial \ln(L)}{\partial \tau_{i2}} \\ \vdots \\ \frac{\partial \ln(L)}{\partial \tau_{iD}} \end{bmatrix} \quad (2.23)$$

Os parâmetros τ_{ik} para um dado item são atualizados de um modo iterativo até que haja pouca mudança nos parâmetros de uma interação para outra próxima ou até que algum limite máximo de interações tenha sido atingido (ex. 30 interações).

No segundo passo do estágio de maximização, as equações de verossimilhança para os parâmetros α_i e δ_i são resolvidos para cada item individualmente. A solução é, de novo, computada usando-se o método de Fisher e a matriz de informação exigida na solução é demonstrada assim:

$$\tilde{I}_{\alpha\delta(i)} = \begin{bmatrix} \tilde{I}_{\alpha,\alpha_i} & \tilde{I}_{\alpha,\delta_i} \\ \tilde{I}_{\delta,\alpha_i} & \tilde{I}_{\delta,\delta_i} \end{bmatrix} \quad (2.24)$$

Os elementos desta matriz são derivados em Rao (1973) e é igual a:

$$\tilde{I}_{\Phi_i\Phi_i^!} = \sum_{f=1}^F N_{if} \sum_{z=0}^D \frac{1}{P(Z_i = z | W_f)} \frac{\partial P(Z_i = z | W_f)}{\partial \Phi_i} \frac{\partial P(Z_i = z | W_f)}{\partial \Phi_i^!} \quad (2.25)$$

onde Φ_i e $\Phi_i^!$ representa tanto α_i e δ_i de uma maneira geral. Os parâmetros são atualizados de um modo iterativo, e a função de atualização para uma q interação é dada por:

$$\begin{bmatrix} \alpha_i \\ \delta_i \end{bmatrix}_q = \begin{bmatrix} \alpha_i \\ \delta_i \end{bmatrix}_{q-1} + \left[\tilde{I}_{\alpha\delta(i)} \right]^{-1} \begin{bmatrix} \frac{\partial \ln(L)}{\partial \alpha_i} \\ \frac{\partial \ln(L)}{\partial \delta_i} \end{bmatrix} \quad (2.26)$$

Os parâmetros α_i e δ_i para um dado item são atualizados iterativamente até que haja pouca mudança nos parâmetros de uma interação para outra ou até que algum

limite máximo de interação tenha sido atingido. Os dois passos do estágio de maximização são executados repetidamente até que haja pouca mudança em qualquer estimativa do parâmetro de um item de uma repetição até a outra repetição ou até que 10 repetições tenham sido executadas. Este procedimento de maximização de dois passos é essencial quando um conjunto de constantes de limiares da categoria subjetiva τ_k é estimado através de todos os itens (ROBERTS, 1995; ROBERTS & LAUGHLIN, 1996ab). Todavia, no GGUM, as equações de verossimilhança associadas com todos os parâmetros para um dado item poderiam facilmente ser resolvidas num passo de maximização simples. No entanto, o procedimento de dois passos é mantido aqui para promover a consistência do modelo através do algoritmo de solução (ROBERTS *et al.*, 2000).

A conclusão do estágio de maximização constitui o fim de um ciclo EM. A estabilidade das estimativas é calculada ao final de cada ciclo EM, e ciclos adicionais são executados se necessários.

2.4.7.2 Valores de Parâmetro de Item Inicial

O algoritmo EM exige uma seleção cuidadosa dos valores de parâmetros de item inicial. Na prática, esses valores “iniciais” são obtidos estimando os parâmetros de item de versões restritas do GGUM. Por exemplo, o modelo de desdobramento graduado pode ser usado para produzir estimativas de τ_k e δ_i sob suposição de que limiares de categorias subjetivas sejam iguais através dos itens e que todos os α_i sejam iguais a 1. As estimativas de τ_k e δ_i produzidas com o GUM podem ser usadas como valores iniciais em modelos menos restritos, permitindo ao parâmetro τ_k variar através dos itens. Essas estimativas de τ_k e δ_i podem então ser usados como valores iniciais.

Simulações têm indicado que esta produção de valores iniciais fornece um “input” adequado para o algoritmo de estimativas (ROBERTS *et al.*, 1998).

2.4.7.3 Estimação do Parâmetro do indivíduo

A estimação de máxima verossimilhança marginal do parâmetro do item é usada em conjunto com as respostas observadas para derivar as estimativas do parâmetro do indivíduo. Essas estimativas do parâmetro do indivíduo constitui as “estimativas

de atitude” do indivíduo. Neste estudo, as estimativas dos parâmetros do indivíduo são obtidas, usando procedimento esperado *a posteriori* (EAP), na qual a estimativa para um j indivíduo é calculada com:

$$\hat{\theta}_j = \frac{\sum_{f=1}^F W_f L_f A(W_f)}{\sum_{f=1}^F L_j(W_f) A(W_f)} \quad (2.27)$$

onde $L_f(W_f)$ é a verossimilhança condicional para observar o vetor resposta do indivíduo j , dado que o indivíduo esteja localizado no ponto de quadratura W_f . A estimativa EAP, $\hat{\theta}_j$, é a média posterior da distribuição θ para um j indivíduo dado o vetor resposta dele. A estimativa EAP é vantajosa porque ela existe para qualquer padrão de resposta, e o erro médio na população especificada pela distribuição *priori* é menor do que outro estimador (BOCK & MISLEVY, 1982).

2.4.7.4 Recuperação dos Parâmetros do Modelo

Roberts *et al.* (1998) desenvolveram um estudo de simulação para examinar a habilidade dos procedimentos de estimativas para recuperar os parâmetros do modelo GGUM quando o θ_j é extraído de uma distribuição *a priori* normal e quanto ao ajuste dos dados. Os resultados sugeriram que estimativas exatas de parâmetros de item poderiam ser obtidas com o tamanho da amostra de 750 indivíduos e 15 a 20 itens com 6 categorias de respostas por item.

2.4.7.5 Erros padrões das estimativas dos Parâmetros

O desvio padrão *a posteriori* aproximado da estimativa de θ de EAP foi dado por Bock & Mislevy (1982) como:

$$\sigma_{\hat{\theta}_j} = \sqrt{\frac{\sum_{f=1}^F (W_f - \theta_j)^2 L_f(W_f) A(W_f)}{\sum_{f=1}^F L_j(W_f) A(W_f)}} \quad (2.28)$$

onde W_f , $L_f(W_f)$ e $A(W_f)$ são definidos na Equação 2.27. Essas aproximações são facilmente computadas ao final do algoritmo EM.

As estimativas dos erros padrões dos parâmetros do item são derivadas seguindo a lógica implementada no software BILOG (MISLEVY & BOCK, 1990). Primeiro, a matriz de informação do parâmetro de item descrito em Bock & Lieberman (1970) é aproximada, considerando-se somente os padrões de respostas observadas na amostra. Esta aproximação é baseada em que todos os parâmetros de item e contêm $l \times (D + 2) \times l \times (D + 2)$ elementos. O elemento da linha p e coluna q da matriz informação é dado por:

$$\hat{I}_{pq} = N \sum_{s=1}^S \frac{P_{OBS_s}}{P_s^2} \left(\begin{array}{cc} \frac{\partial \tilde{P}_s}{\partial \Phi_p} & \frac{\partial \tilde{P}_s}{\partial \Phi_q} \\ \frac{\partial \tilde{P}_s}{\partial \Phi_p} & \frac{\partial \tilde{P}_s}{\partial \Phi_q} \end{array} \right), \text{ para } \begin{array}{l} p = 1, 2, \dots, l(D + 2) \\ q = 1, 2, \dots, l(D + 2) \end{array} \quad (2.29)$$

onde

P_{OBS} é a freqüência relativa do vetor de resposta observada X_s ;

Φ_p é um parâmetro arbitrário $(\alpha_i, \delta_i, \tau_i)$ para um dos l itens;

Φ_q é um parâmetro arbitrário $(\alpha_i, \delta_i, \tau_i)$ para um dos l itens, e

\tilde{P}_s é verossimilhança marginal de observar X_s o vetor resposta como a definida na Equação 2.17 (ROBERTS *et al.*, 2000).

A matriz definida pela Equação 2.29 pode ficar muito grande assim que o número de itens aumenta, e, portanto pode ser difícil em termos de computação e demorado. Uma aproximação mais eficiente da matriz de informação pode ser derivada considerando somente aquelas derivadas que correspondem a um item simples (isto é, os elementos diagonais do bloco da matriz de informação quando a linha p da matriz informação está ordenada por item):

$$\hat{I}_{\alpha\delta\tau(i)} = \begin{bmatrix} \hat{I}_{\alpha_i\alpha_i} & \hat{I}_{\delta_i\delta_i} & \hat{I}_{\alpha_i\tau_{i1}} & \dots & \hat{I}_{\alpha_i\tau_{iD}} \\ \hat{I}_{\delta_i\alpha_i} & \hat{I}_{\delta_i\delta_i} & \hat{I}_{\delta_i\tau_{i1}} & \dots & \hat{I}_{\delta_i\tau_{iD}} \\ \hat{I}_{\tau_{i1}\alpha_i} & \hat{I}_{\tau_{i1}\delta_i} & \hat{I}_{\tau_{i1}\tau_{i1}} & \dots & \hat{I}_{\tau_{i1}\tau_{iD}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{I}_{\tau_{iD}\alpha_i} & \hat{I}_{\tau_{iD}\delta_i} & \hat{I}_{\tau_{iD}\tau_{i1}} & \dots & \hat{I}_{\tau_{iD}\tau_{iD}} \end{bmatrix} \quad \dots \quad (2.30)$$

Essa aproximação é razoável sempre que o número de itens é grande (maior que 20) devido ao fato de que as derivadas cruzadas entre os itens serão geralmente menores (MISLEVY & BOCK, 1990).

Os elementos do indivíduo de $I_{\alpha\delta\tau(i)}$ são dados por:

$$\hat{I}_{\Phi_i\omega_i} = N \sum_{s=1}^S \frac{P_{OBS_s}}{\tilde{P}^2} \left(\frac{\partial \tilde{P}_s}{\partial \Phi_i} \frac{\partial \tilde{P}_s}{\partial \omega_i} \right), \quad \dots \quad (2.31)$$

Onde Φ_i e ω_i referem-se geralmente a pares distintos dos parâmetros de item para item i . A derivada geral na Equação 2.31 é igual a:

$$\begin{aligned} \frac{\partial \tilde{P}_s}{\partial \Phi_i} &= \sum_{f=1}^F A(W_f) \frac{\partial L_s(W_f)}{\partial \Phi_i} \\ \frac{\partial \tilde{P}_s}{\partial \Phi_i} &= \sum_{f=1}^F A(W_f) \frac{\partial \left(\prod_{i=1}^I P[Z_i = x_{si} | W_f] \right)}{\partial \Phi_i} \\ \frac{\partial \tilde{P}_s}{\partial \Phi_i} &= \sum_{f=1}^F A(W_f) \frac{\partial P[Z_i = x_{si} | W_f]}{\partial \Phi_i} \left(\prod_{\substack{i=1 \\ i \neq i}}^I P[Z_i = x_{si} | W_f] \right) \\ \frac{\partial \tilde{P}_s}{\partial \Phi_i} &= \sum_{f=1}^F A(W_f) \frac{\partial P(Z_i = x_{si} | W_f)}{\partial \Phi_i} \left(\frac{\prod_{i=1}^I P[Z_i = x_{si} | W_f]}{P[Z_i = x_{si} | W_f]} \right) \\ \frac{\partial \tilde{P}_s}{\partial \Phi_i} &= \sum_{f=1}^F \frac{L_s(W_f)}{P(Z_i = x_{si} | W_f)} A(W_f) \frac{\partial P(Z_i = x_{si} | W_f)}{\partial \Phi_i} \end{aligned} \quad (2.32)$$

A quantidade $\partial P(Z_i = x_{si} | W_f) / \partial \Phi_i$ referido na Equação 2.32, pode ser calculada para cada parâmetro de item de interesse usando uma equação dada no apêndice de Roberts *et al.* (2000). Uma vez que a matriz de informação aproximada para um dado conjunto de parâmetros de item tenha sido calculados, pode-se, então, derivar os erros padrões aproximados para aqueles parâmetros como:

$$\hat{\sigma}_{\Phi_i} = \left[\text{DIAG} \left(\hat{I}_{\alpha\delta\tau(i)}^{-1} \right) \right]^{1/2} \quad (2.33)$$

2.4.7.6 A Sensibilidade das estimativas dos Parâmetros quanto à suposição de uma distribuição *priori* e o Numero de pontos de Quadratura

Atualmente pouco se sabe sobre a sensibilidade das estimativas dos parâmetros do modelo GGUM quanto à suposição de uma distribuição a *priori*, usada na técnica MML/EAP. Todavia, se considerarmos os resultados de pesquisa de sensibilidade obtidos nos modelos cumulativos da TRI, então as estimativas dos parâmetros de item são robustas e podem se esperar discrepâncias moderadas entre as distribuições a *priori* presumidas e verdadeira de θ_j (BARTHOLOMEW, 1988; BOCK e AITKIN, 1981; SEONG, 1990; SWINDERMAN & VAN DEN WOLLENBERG, 1990 *apud* ROBERTS *et al.* 2000). As estimativas de EAP de θ_j parecem ser mais sensíveis para tais discrepâncias do que são as estimativas dos parâmetros do item, embora possam encontrar estimativas inexatas de θ_j devido às *priori* mal ajustadas ainda que pareçam ser modestas (SEONG, 1990 *apud* ROBERTS *et al.* 2000). A influência da distribuição nas estimativas EAP torna-se menor assim que o número de respostas de item aumenta, dadas as estimativas exatas de parâmetro de item (BOCK & MISLEVY, 1982 *apud* ROBERTS *et al.* 2000). Portanto, a distribuição das estimativas EAP pode corresponder muito bem à verdadeira distribuição de θ_j mesmo quando a verdadeira distribuição difere da *priori* presumida. Esta é uma força da abordagem MML/EAP (ROBERTS *et al.* 2000).

2.4.8 Equalização de estimativas de Parâmetros do GGUM

Em primeiro lugar vamos definir o que significa equalizar. Equalizar significa equiparar, tornar comparável, o que na TRI significa colocar parâmetros de itens vindos de testes, provas distintas ou habilidades de indivíduos de diferentes grupos, na mesma métrica, ou seja, numa escala comum, tornando os itens e/ou as habilidades comparáveis (VALLE 1999)

São utilizados três métodos para equalização de estimativas de parâmetros para generalizar o GGUM: Média - Desvio, Média -Média e Método da Curva Característica do Item. Os dois primeiros métodos são implementados usando-se duas estratégias diferentes baseadas na parametrização alternativa GGUM. Todos esses métodos tentam estimar uma escala constante A e a locação constante B, que possam tornar comparável a métrica do parâmetro de modelo de resposta do item provenientes de calibrações separadas.

Logo, uma média confiável para equalizar às estimativas de parâmetro do GGUM através de calibrações múltiplas é necessária antes que tais aplicações sejam possíveis.

Como já foi mencionado o GGUM é consistente com processo de resposta baseado na proximidade, ele produz CCI que são de pico simples e não monótonas.

O GGUM fornece uma média para equalizar testes de uma perspectiva da TRI que incorpora este relacionamento não-monotônico entre resposta de item e o traço latente. Da perspectiva da TRI, se as respostas para dois conjuntos de itens de teste forem analisados separadamente usando dois GGUM, então a equalização do teste é simplesmente uma questão de se colocar as estimativas de parâmetros das duas calibrações sobre a mesma métrica (ROBERTS *et al.* 2001).

Mais detalhes sobre métodos de equalização pode-se ver em Robert *et al.* 2001.

2.5 CEFET – Centro Federal de Educação Tecnológica do Paraná

O CEFET – PR teve início em 1909 quando foram criadas as escolas de Aprendizes e Artífices no Brasil, por decreto do então Presidente Nilo Peçanha.

Em 1910 foi implantada a Escola do Paraná. Em 1942 devido à unificação do ensino industrial ela passou a ser denominada Escola Técnica de Curitiba. Em 1944 foi criado o primeiro curso de segundo ciclo na instituição. Em 1959 a escola ganhou autonomia com a reforma do ensino industrial e passou a chamar-se Escola Técnica Federal do Paraná. Em 1973 passou a ofertar cursos superiores de curta duração de Engenharia de Operações. Todavia foi 1978 que houve um grande avanço da instituição, pois passou a ministrar ensino superior de duração plena, transformando em Centro Federal de Educação Tecnológica do Paraná (CEFET – PR).

E, desde então o ensino no CEFET – PR cresceu gradativamente, sempre seguindo o desenvolvimento da sua área de atuação e o seu contexto sócio-cultural-econômico (ROMANO 2000).

De 1986 a 1995, através de um Programa de Expansão e Melhoria do Ensino Técnico, desenvolvido pelo Governo Federal, o CEFET – PR ampliou sua atuação para todo o estado do Paraná, implantando Unidades Descentralizadas em mais de 5 cidades do estado: Medianeira, Cornélio Procópio, Pato Branco, Ponta Grossa e Campo Mourão. O CEFET – PR atualmente se presta em desenvolver sua ação educacional através de uma política consolidada de integração escola-empresa,

além de estender sua competência a atividades de ensino, extensão e pesquisa científico – tecnológica à comunidade (ROMANO, 2000)

Com a Nova Legislação Brasileira nº9394 em 20 de dezembro de 1996 o CEFET-PR implantou novas modalidades de ensino superior; os chamados cursos de Tecnologia.

Esses cursos foram estruturados deixando evidente a formação diferenciada exigida pela mudança nos cenários tecnológicos e profissionais, principalmente, porque foi implantada no currículo uma disciplina obrigatória, no oitavo e último período, denominada Trabalho de Diplomação, que tem por objetivo o desenvolvimento de um produto ou um processo inovador e, também, porque todas as atividades educacionais foram apoiadas em três pilares curriculares: a ciência, a tecnologia e a gestão (ROMANO 2000).

Essas foram as características que nortearam os cursos de Tecnologia do CEFET-PR.

Após esta visão geral sobre a fundamentação teórica que norteia o modelo GGUM. O próximo capítulo apresenta o recurso computacional disponível que se propõem a resolver na prática os métodos de estimação dos parâmetros do modelo GGUM da TRI. Métodos estes que exigem cálculos avançados de derivação e integração, assim sendo os recursos computacionais auxiliam em muito a utilização e sua viabilização.

3 RECURSOS COMPUTACIONAIS

Neste capítulo apresenta-se o software GGUM2000 um recurso computacional empregado na estimação dos parâmetros do GGUM, os processos de estimação do GGUM2000, os pontos de quadratura usados, os erros padrão obtidos na estimação dos parâmetros, as limitações do programa. Por fim o arquivo de produção do software.

3.1 Introdução

O crescimento e a divulgação da TRI sempre estiveram intimamente atrelados ao desenvolvimento paralelo de recursos computacionais que viabilizassem sua utilização. Isto porque os instrumentos matemáticos imprescindíveis para sua aplicação são muito mais complexos do que as técnicas empregadas na TCM.

Desde o início de suas aplicações, alguns pesquisadores desenvolveram seus próprios programas computacionais, mas é evidente que seu uso em larga escala depende diretamente da disponibilidade de programas computacionais comerciais no mercado.

Nos Estados Unidos e na Europa, desde a década de 70, foram lançados alguns programas específicos para análises via TRI como, por exemplo, TESTFACT, MULTILog (DOS) v. 6,0, PARSCALE (DOS) v. 3,3, BILOG W (BILOG for Windows v.309 , BILOG-MG v.1,0 DOS Extender) e ConTEST v.2,0 (DOS). Aqui no Brasil, onde a utilização da TRI é bem mais recente, há uma variedade bem menor de programas computacionais comerciais sendo usados.

Dos programas disponíveis no mercado, os que são atualmente mais utilizados nas análises envolvendo a TRI - aqui no Brasil - são o BILOG (MISLEVY & BOCK, 1990) e o BILOG-MG (ZIMOWSKI *et al.*, 1996). Estes dois programas são específicos para análises via TRI de itens dicotômicos e ambos têm implementado os modelos unidimensionais logísticos de 1, 2 e 3 parâmetros. A diferença básica entre eles é que o BILOG-MG permite a análise de mais de um grupo de respondentes, enquanto que o BILOG permite apenas analisar respondentes considerados como provenientes de uma única população.

Neste capítulo, vamos comentar apenas sobre GGUM2000 (DOS), programa computacional que se propõem obter estimativas para os parâmetros do modelo de

desdobramento GGUM e que se obtém, sem custos em www.education.umd.edu/EDMS/tutorials/index.html.

3.2 GGUM2000

Os modelos da TRI de desdobramentos que são implementados no programa GGUM2000 são apropriados para medir uma grande variedade de construtos, tais como atitudes de indivíduos (nas quais se utilizam questionários segundo a escala de Likert ou Thurstone), preferências e também processos de desenvolvimento (NÖEL, 1999 *apud* ROBERTS *et al.*, 2001) que ocorrem em estágios.

O programa GGUM2000 estima parâmetros dos oito modelos da família do GGUM2000. No capítulo 2 apresentou-se o modelo graduado generalizado GGUM. Pode-se, a partir desse modelo, obter 7 modelos através de restrições feitas nos parâmetros em diferentes modos.

3.2.1 Processo de Estimação do GGUM 2000

O método implementado para estimação dos parâmetros dos itens utiliza uma aproximação por máxima verossimilhança marginal (MVM) e esperança a *posteriori* (EAP).

Para que os parâmetros possam ser estimados é necessário o uso de uma distribuição de probabilidade *a priori* para θ dos respondentes. Esse programa assume que a *a priori* segue uma distribuição normal padrão com média 0 e desvio padrão 1, isto é, $N(0; 1)$. Ele utiliza duas formas de resolver as equações de verossimilhança marginal, o método EM e o método de *scoring* de Fischer como descrito no Capítulo 2. Para a execução dos passos de maximização, o GGUM2000 denomina de ciclos internos. O estágio esperança é denominado de ciclos externos.

Ele utiliza as respostas observadas juntamente com a máxima verossimilhança marginal da estimação dos parâmetros dos itens para estimar os parâmetros do indivíduo, mais precisamente ele utiliza procedimento EAP para esta estimação.

3.2.2 Pontos de Quadratura

As equações de verossimilhança envolvem integrações que são solucionadas através de pontos de quadratura discretos. E é o usuário que especifica o número de pontos de quadratura, esse número, que pode chegar até um máximo de 50

pontos de quadratura, são sempre igualmente espaçados num intervalo de $-4,0$ a $+4,0$.

3.2.3 O erro padrão na estimação dos parâmetros

Os erros padrão produzidos pelo GGUM2000 são aproximados e como tal devem ser interpretados assim. A aproximação considera somente a resposta padrão na amostra de dados. Sendo os elementos da porção diagonal fora do bloco da matriz de informação de todos os parâmetros de item muitos pequenos, dessa forma permite que cada item seja trabalhado independentemente.

Roberts *et al.*, (2000) em simulações feitas, verificaram que questionários com 20 itens de 6 categorias por item sugerem erros padrões razoáveis.

3.2.4 Limitações do Programa

O número de respondentes que o programa GGUM2000 suporta é de, no máximo, 2000 indivíduos e, no máximo 100 itens com, no máximo, 10 categorias de respostas por item, sendo que os itens devem ter o mesmo número de categoria de resposta, e as respostas dos itens não podem conter dados perdidos. O número de categorias de respostas deve ser igual para cada item. O número máximo que deve ser especificado para os pontos de quadratura é de 50 e a distribuição *a priori* a ser utilizada é a distribuição normal padrão.

3.2.5 Comandos do programa

Este programa necessita de uma série de comandos para se obter as estimativas dos parâmetros do modelo GGUM. Esses comandos podem ser iniciados interativamente com o usuário ou agrupados através de um arquivo no formato ASCII. Existem dois tipos de comandos: comandos primários e comandos secundários, que são opcionais, dependem dos comandos primários.

Quando os comandos de programa forem agrupados, então cada comando primário e secundário é listado em um registro separado no arquivo de grupo (mais detalhes ver ROBERTS *et al.*, 2000).

O programa foi desenvolvido em MS-DOS sob Windows 95/98.

Ao instalar o programa, ele automaticamente criará três subdiretórios chamados de C:\GGUM2000, C:\GGUM2000\TEMPFILE e C:\GGUM2000\EXAMPLES.

O programa interage com o usuário e a cada comando solicitado existe uma opção de comando resposta em que o usuário vai estabelecendo o tipo e modelo desejado, a quantidade de categorias de respostas que dispõe, a quantidade de item de arquivo, o número de ciclos EM, o número de interações de *scoring* de Fischer, critério de convergência, se algum indivíduo deve ser eliminado ou algum item, se sim, especificar quantos, se deseja estatística de ajuste, o valor de corte (*cutoff*), nível de α para o qui-quadrado e assim o software vai interagindo com o usuário e este vai especificando o modelo e as condições que almeja para a estimação dos parâmetros do modelo.

No programa há vários exemplos encontrados no subdiretório C:\GGUM/EXAMPLES.

3.2.6. Produção do GGUM 2000

Este programa cria vários arquivos de produção e todos eles são armazenados no subdiretório C:\GGUM2000\TEMPFILE\

Dentre estes arquivos há alguns que são operacionais, isto é, arquivos requeridos pelo programa e há outros que são de interesse do usuário.

Os arquivos de interesse do usuário são:

- . FT16F001 – que contém todas as estimativas de todos os parâmetros dos itens.
- . FT17F001 – que contém estimativas de parâmetros theta. A estimativa de cada indivíduo é fornecida em um registro separado e cada registro contém o número ID do indivíduo seguido pela estimativa EAP e o erro padrão de estimativa.
- . FT06F001 – este arquivo descreve tudo o que acontece durante a execução do programa GGUM2000.
- . FT15F001 – neste arquivo são descritos as estatísticas de ajustes se estas forem solicitadas.

Este sistema oferece uma variedade de estatísticas de ajuste para o item, indivíduo e o modelo de ajustes. Numa primeira seção do arquivo, pode-se encontrar, parâmetros de item para cada item classificado pelo δ_i junto com diversas estatísticas de ajuste de item, incluindo média de quadrados de *infit* e *outfit*, qui-quadrado e valor de probabilidades associados, valores de t *infit* e *outfit* e valores de t para *infit* e *outfit* localizados. Na seção seguinte ele fornece informação sobre *infit* do indivíduo e índice *outfit*.

Há uma outra seção intitulada “ANDRICH’S ITEM-LEVEL CHI SQUARE VALUES”. Essa seção está baseada num índice qui-quadrado de item adaptado para dados politômicos da TRI.

Na seção final, lista-se a razão qui-quadrado da verossimilhança para cada item e para o modelo como um todo (mais detalhes ver Roberts *et al.*, 2000).

Finalmente, este capítulo abordou de forma sucinta o sistema computacional GGUM2000, programa desenvolvido para fazer as estimativas dos parâmetros do modelo GGUM.

O próximo capítulo apresenta a metodologia empregada na aplicação do modelo GGUM.

4 METODOLOGIA

Nos capítulos anteriores foram apresentados o modelo de desdobramento graduado generalizado - GGUM da TRI, suposições e como é feita a estimação dos parâmetros. Já neste, far-se-á uma aplicação do modelo GGUM da TRI onde serão descritos todas as etapas para tal aplicação.

4.1 Introdução

A elaboração de uma escala de medida diretamente aplicada à avaliação de ensino para verificar a satisfação de alunos quanto ao curso é primordial para as instituições de ensino, pois proporciona uma reflexão sobre as metodologias utilizadas, bem como fornece subsídios para a melhoria da qualidade do ensino dos cursos oferecidos pela instituição.

4.2 Coleta de Dados

Para a aplicação do modelo GGUM da TRI foi utilizado um questionário de avaliação de ensino da UNIOESTE – Cascavel – PR que foi adaptado, obedecendo aos objetivos propostos.

A pesquisa foi aplicada a 524 alunos dos cursos de Tecnologia do CEFET – MD – PR.

Os alunos responderam um questionário contendo 35 itens onde cada item continha seis categorias de resposta, e os alunos foram convidados a assinalar uma das seis categorias: totalmente insatisfeito, insatisfeito, pouco insatisfeito, pouco satisfeito, satisfeito, totalmente satisfeito. Os alunos responderam o questionário em sala de aula, após uma breve explanação sobre o objetivo do mesmo e foi aplicado pela própria autora.

Nesta pesquisa, além destes 35 itens, foram incluídos também mais 5 itens com as seguintes características:

- 1) Sexo
- 2) Curso
- 3) Trabalha
- 4) Cidade de Procedência
- 5) Idade

Estes itens foram incluídos para se ter um perfil dos alunos que responderam o questionário.

As características principais dos alunos são demonstradas nas figuras abaixo.

A Figura 4.1 mostra a descrição dos alunos do CEFET - MD que freqüentam os cursos de Tecnologia quanto ao sexo.

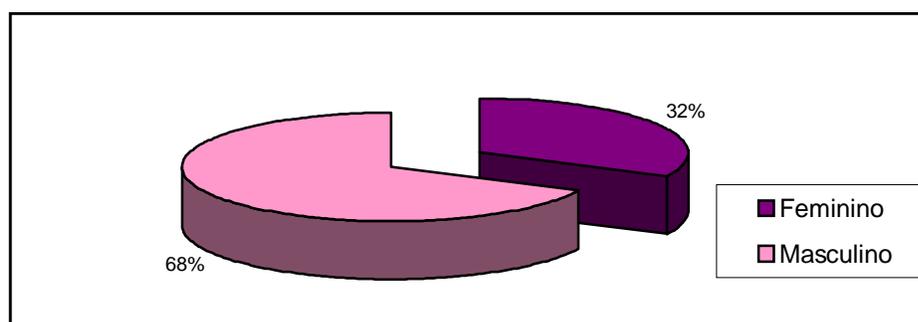


Figura 4.1 - Descrição dos alunos do CEFET - MD que freqüentam os cursos de Tecnologia quanto ao sexo.

Observa-se na Figura 4.1 que o número de alunos do sexo masculino é maior do que a do sexo feminino.

A Tabela 4.1 apresenta a descrição dos alunos quanto à cidade de procedência.

Tabela 4.1 - Descrição dos alunos do CEFET - MD que freqüentam os cursos de Tecnologia quanto à cidade de procedência.

Cidade de Procedência	Nº de alunos	%
Assis Chateaubriand	1	0,2
Cascavel	25	4,8
Ceu Azul	2	0,4
Céu Azul	7	1,3
Corbelia	1	0,2
Foz do Iguaçu	40	7,6
Iporã	3	0,6
Itaipulândia	11	2,1
Marechal C. Rondon	2	0,4
Matelandia	28	5,3
Medianeira	277	52,9
Missal	14	2,7
Pato Bragado	1	0,2
Santa Helena	23	4,4
Santa Terezinha de Itaipu	4	0,8
São Miguel do Iguaçu	39	7,4
Serranópolis do Iguaçu	12	2,3
Toledo	27	5,2
Vera Cruz do Oeste	7	1,3

Visualiza-se que 53% dos alunos que freqüentam os cursos de Tecnologia são de Medianeira, cidade onde se localiza o CEFET, e em número menor alunos oriundos de cidades próximas de Medianeira.

A Figura 4.2 apresenta as idades dos alunos que freqüentam os cursos de Tecnologia do CEFET - MD.

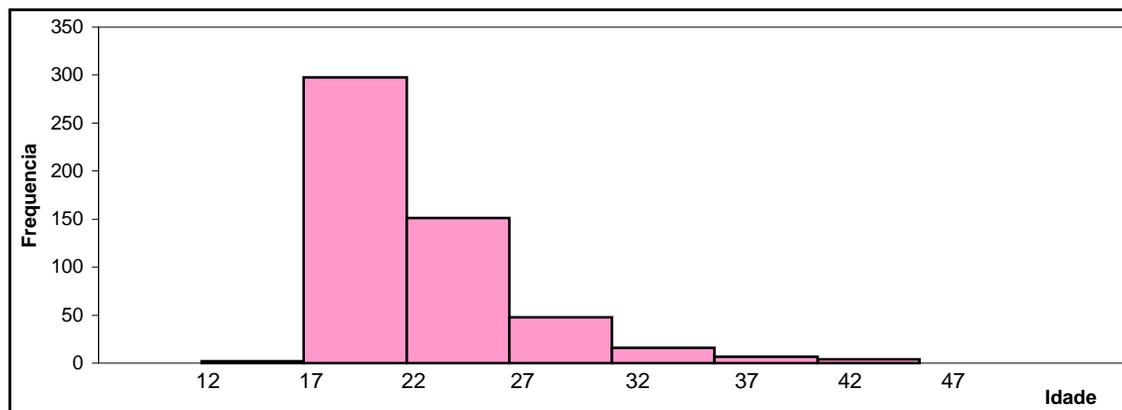


Figura 4.2: Histograma das idades dos alunos do CEFET-MD.

Nota-se que a faixa etária dos alunos que freqüentam os cursos de Tecnologia é maior dos 17 aos 22 anos.

4.3 Desenvolvimento do Banco de Itens

4.3.1 Dimensionalidade

A suposição de unidimensionalidade é primordial no processo de estimação, como foi salientado anteriormente, uma vez que sua violação pode invalidar as estimativas dos parâmetros.

Para a verificação dessa suposição desenvolveu-se uma Análise Fatorial.

A Tabela 4.2 mostra o resultado da extração dos autovalores da análise fatorial pelo método dos componentes principais. O item 30 foi eliminado antes de fazer a análise, pois alguns alunos não responderam a este item.

Tabela 4.2: Extração dos autovalores através dos componentes principais

Fator	Autovalor	Variância Total %	Autovalor Cumul.	Cumul %
1	11,29467	33,21961	11,29467	33,21961
2	2,050988	6,032316	13,34565	39,25192

Segundo Reckase (1979) os resultados podem indicar uma unidimensionalidade se o primeiro fator da conta de 20% da variância total. Como se observa na Tabela 4.2, o primeiro fator explica 33% da variância, satisfazendo este critério, logo se pode considerar que o teste é unidimensional.

Foram calculadas também na análise fatorial as comunalidades de cada item para avaliar quais itens eram menos favoráveis com a suposição de unidimensionalidade nos dois primeiros componentes. Na Tabela 4.3 observa-se o resultado da cargas fatoriais dos dois primeiros componentes e as comunalidades de cada item.

Tabela 4.3: Matriz dos dois primeiros componentes e comunalidades dos itens

Item	Fator 1	Fator 2	Comunalidade
1	-0,529	-0,066	0,284
2	-0,519	-0,132	0,287
3	-0,614	-0,394	0,532
4	-0,576	-0,406	0,497
5	-0,544	-0,337	0,409
6	-0,511	-0,184	0,295
7	-0,474	0,089	0,233
8	-0,621	-0,278	0,463
9	-0,473	0,072	0,229
10	-0,590	-0,150	0,370
11	-0,588	0,019	0,346
12	-0,507	0,322	0,361
13	-0,557	0,366	0,444
14	-0,608	0,395	0,526
15	-0,651	0,423	0,603
16	-0,525	0,389	0,427
17	-0,609	0,183	0,404
18	-0,548	0,187	0,336
19	-0,441	0,084	0,202
20	-0,664	-0,046	0,443
21	-0,676	-0,036	0,458
22	-0,701	-0,127	0,507
23	-0,680	-0,085	0,470
24	-0,727	0,046	0,531
25	-0,591	0,195	0,387
26	-0,504	0,293	0,340
27	-0,601	0,341	0,478
28	-0,661	0,038	0,438
29	-0,689	-0,141	0,494
31	-0,556	-0,051	0,311
32	-0,240	-0,223	0,108
33	-0,361	-0,370	0,267
34	-0,515	-0,356	0,392
35	-0,657	-0,209	0,475

Nota: A análise fatorial se realizou com o método dos componentes principais extraíndo-se dois fatores, utilizando as cargas fatoriais sem nenhuma rotação..

Para que um item se ajuste à suposição de unidimensionalidade é necessário que sua comunalidade seja maior que 0,3, ou seja, que o item tenha mais do que 30% de aspectos em comuns com outros itens. Nesta análise com todos os itens, menos o item 30, nota-se que oito itens apresentam comunalidade menor que 0,3, por conseguinte estes itens foram removidos da análise.

4.3.2 Seleção de itens para a escala final

Foram realizadas várias calibrações, na primeira calibração observou-se que 10 alunos apresentaram o desvio padrão no parâmetro θ elevado, assim optou-se por eliminar esses 10 alunos na calibração final. As calibrações foram feitas nos 26 itens restantes usando a máxima verossimilhança marginal com uma distribuição $N(0;1)$ a priori para os θ_j . Foram usados 50 pontos de quadratura igualmente espaçados entre $-4,0$ e $+4,0$. Para o ajuste dos itens utilizou-se tanto o *infit* como o *outfit*. Na análise dos resultados obtidos, nenhum item apresentou falta de ajuste (*misfit*), assim esses 26 itens ficaram retidos na escala final. As estimativas dos parâmetros dos alunos também foram obtidas através do procedimento EAP.

A Tabela 4.4 mostra a distribuição de freqüências das respostas observadas dos alunos aos 25 itens da escala final.

Tabela 4.4: Distribuição de Freqüência das respostas observadas na escala final.

Item	1	2	3	4	5	6
3	3	30	70	192	200	19
4	2	17	57	178	235	25
5	4	18	65	193	211	23
8	8	36	72	196	180	22
10	5	22	49	121	254	63
11	7	23	62	137	247	38
12	8	39	79	149	202	37
13	27	61	114	175	127	10
14	13	56	116	184	133	12
15	26	48	93	198	141	8
16	12	46	91	168	171	26
17	28	52	101	172	138	23
18	24	50	89	120	180	51
20	29	84	98	174	116	13
21	35	67	128	170	103	11
22	20	49	115	189	132	9
23	16	44	95	203	151	5
24	4	32	93	197	175	13

Item	1	2	3	4	5	6
25	11	37	106	166	172	22
26	24	67	91	161	146	26
27	12	47	93	154	188	20
28	26	46	85	160	165	32
29	13	51	75	163	186	26
31	12	31	42	116	209	104
34	3	20	38	124	254	75
35	20	45	92	177	155	25

Nota-se que a grande parte dos alunos escolhem a categoria 4 e 5 em relação aos itens caracterizando que estão pouco satisfeitos e satisfeitos com relação aos itens questionados.

4.4 Estimativas dos Parâmetros dos Itens

A Tabela 4.5 mostra os itens retidos na escala final seguido das estimativas dos itens através do modelo GGUM. Os itens estão listados de acordo com a ordem crescente do parâmetro δ_i . O GGUM ordena as declarações de uma maneira lógica correspondente a conteúdo afetivo do negativo, passando pelo neutro e positivo. Entretanto os 26 itens que ficaram na escala final tiveram conteúdo positivo variando de moderado a extremo, não foi verificado nenhum item negativo e nem neutro.

Tabela 4.5: Estimação dos parâmetros dos itens ($\hat{\delta}_i, \hat{\alpha}_i, \hat{\tau}_{ik}$) para 26 itens e a correlação (r_i) entre a respostas observadas e esperadas para cada item.

Item	$\hat{\delta}_i$	$\hat{\alpha}_i$	$\hat{\tau}_{i1}$	$\hat{\tau}_{i2}$	$\hat{\tau}_{i3}$	$\hat{\tau}_{i4}$	$\hat{\tau}_{i5}$	r_i
4	1,08	1,3	-4,1	-3,1	-2,6	-1,4	1,2	0,975
34	1,10	1,0	-4,4	-2,8	-3,0	-2,0	0,6	0,991
5	1,16	1,0	-3,9	-3,3	-2,7	-1,3	1,5	0,960
3	1,23	1,3	-4,3	-2,8	-2,5	-1,3	1,2	0,986
23	1,32	1,7	-3,1	-2,6	-2,1	-1,0	1,5	0,995
35	1,34	1,2	-3,1	-2,7	-2,2	-1,1	0,8	0,982
8	1,35	1,2	-3,8	-2,8	-2,6	-1,2	1,0	0,989
29	1,42	1,5	-3,5	-2,5	-2,4	-1,4	0,6	0,995
22	1,42	1,9	-3,1	-2,7	-2,0	-1,0	0,9	0,996
21	1,50	1,4	-2,9	-2,6	-1,9	-0,8	1,0	0,992
20	1,53	1,2	-3,4	-2,3	-2,2	-0,9	1,2	0,992
24	1,61	2,0	-4,0	-3,1	-2,5	-1,4	0,6	0,996
10	1,64	1,0	-4,4	-3,4	-3,1	-2,5	0,3	0,986
31	1,71	0,8	-4,0	-2,9	-3,5	-2,6	-0,3	0,985
11	1,87	1,0	-4,3	-3,8	-3,2	-2,5	0,7	0,970
26	1,90	0,6	-4,4	-2,9	-3,1	-1,5	1,9	0,990
28	1,94	1,2	-3,5	-3,2	-2,8	-1,8	0,2	0,983
17	1,96	1,0	-3,5	-3,3	-2,7	-1,5	0,7	0,990

Item	$\hat{\delta}_i$	$\hat{\alpha}_i$	$\hat{\tau}_{i1}$	$\hat{\tau}_{i2}$	$\hat{\tau}_{i3}$	$\hat{\tau}_{i4}$	$\hat{\tau}_{i5}$	r_i
27	2,06	1,0	-4,4	-3,5	-2,9	-2,1	-0,9	0,979
18	2,19	0,7	-4,1	-3,6	-2,9	-2,7	0,4	0,986
13	2,40	0,8	-4,3	-3,7	-3,1	-1,7	1,9	0,982
16	2,82	0,8	-5,5	-4,4	-3,9	-2,7	0,5	0,998
12	3,02	0,7	-6,2	-4,7	-4,3	-3,4	0,1	0,984
14	3,13	1,1	-5,5	-4,5	-3,8	-2,5	0,3	0,979
15	3,13	1,4	-4,6	-4,3	-4,0	-2,6	0,3	0,992

Os parâmetros de discriminação α_i apresentaram uma variação entre 0,6 e 2,0.

A estimação foi obtida através do software GGUM2000, que emprega a máxima verossimilhança marginal para estimar os parâmetros dos itens, utilizando o algoritmo EM nas estimativas. No processo de maximização EM, foram requeridos 10 ciclos internos e 200 ciclos externos. Foram usados 50 pontos de quadratura igualmente espaçados entre $-4,0$ e $+4,0$. Para a solução do estágio de maximização do algoritmo EM usou-se 30 interações de *scoring* de Fischer tanto para a estimação dos parâmetros dos itens como para a estimação dos valores de τ_{ik} .

4.5 Estimativas do Parâmetro θ_j

Foi empregada a máxima verossimilhança marginal dos parâmetros dos itens juntamente com as respostas observadas para obter as estimativas dos parâmetros dos alunos. Essas estimativas são obtidas utilizando o procedimento EAP. Utilizou-se o software GGUM2000 para essas estimativas.

A média de distribuição de θ_j foi de -0.02269 com um desvio padrão de $0,927$. A Figura 4.3 ilustra o teste de normalidade de Kolmogorov –Smirnov.

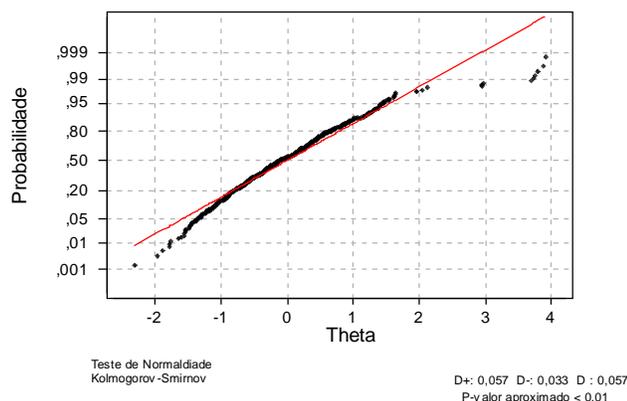


Figura 4.3: Gráfico de Normalidade dos parâmetros θ_j através do software MINITAB.

A distribuição dos parâmetros θ_j difere significativamente de uma distribuição normal de acordo com o teste Kolmogorov –Smirnov (D:0,057, $p < 0.01$). Essa diferença foi devido a uma assimetria positiva de 0,9724 e coeficiente de curtose de 2.438. O valor mediano de θ_j foi de -0.1219 . O intervalo $[-0.103; 0.0577]$ contém a verdadeira média dos parâmetros θ_j , com 95% de confiança.

4.6 Ajuste do Modelo

O ajuste do modelo foi feito de forma heurística. Primeiramente calculou-se as diferenças entre os pares $(\hat{\theta}_j, \hat{\delta}_i)$ e essas diferenças foram divididas em aproximadamente 10 grupos de tamanhos, 6 de 52 e 4 de 51. Foram calculados dentro da cada grupo a média observada e o escore esperado.

A Figura 4.4 mostra a relação entre a média dos escores observados e esperados em função da média $\hat{\theta}_j - \hat{\delta}_i$ para cada grupo.

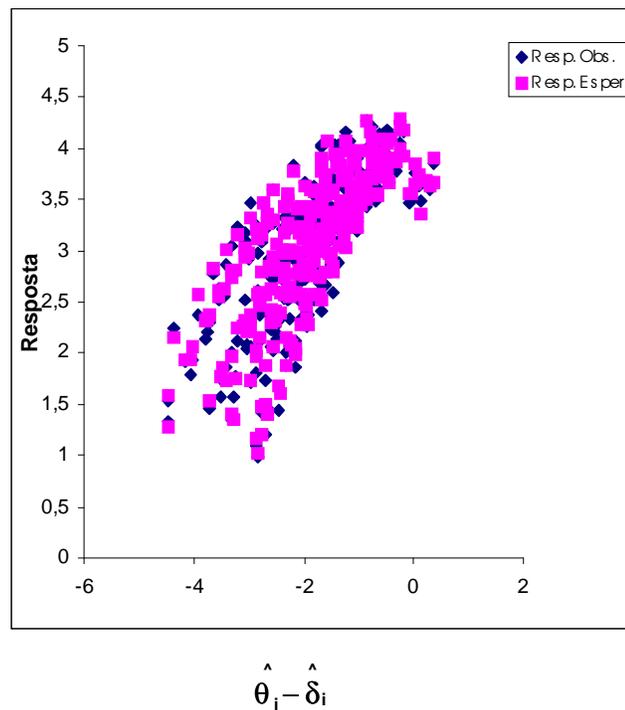


Figura 4.4: Médias observadas e esperadas das respostas dos itens em função das médias $\hat{\theta}_j - \hat{\delta}_i$.

O GGUM localiza tanto os parâmetros dos itens como os parâmetros dos alunos num *continuum* θ . Nota-se na Figura 4.5 que quando $|\hat{\theta}_j - \hat{\delta}_i|$ torna-se menor, o grau de satisfação aumenta na resposta. Dessa forma, o ajuste do GGUM parece ser razoável.

No ajuste do modelo, computou-se também o *infit* e *outfit* e a correlação entre a respostas observadas e esperadas para cada item (veja Tabela 4.5). Para o *infit* e *outfit*, o software GGUM2000 solicita que se especifiquem critérios, então foram utilizados os valores recomendados pelo software.

O *infit* variou de 0,96 a 1,01 e segundo Adams e Kho (1993) é aceitável de 0,77 a 1,30, valores acima de 1,30 indicam itens que possuem baixa discriminação e itens com *infit* abaixo de 0,77 são itens que provêm de informação redundante, logo os valores de *infit* são aceitáveis, não sugerindo itens com falta de ajuste.

As correlações entre as respostas observadas e esperadas, dadas na Tabela 4.5, revelam uma forte correlação entre elas para todos os itens. Nenhum item apresentou correlação baixa. A média global das correlações foi de 0,99.

Para uma ilustração, as Figuras 4.5, 4.6 e 4.7 apresentam as médias dos valores observados e esperados para os itens 14, 18, 34 respectivamente.

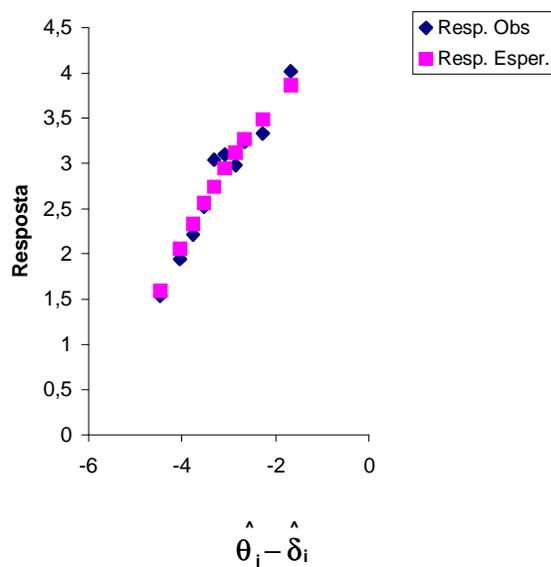


Figura 4.5: Médias observadas e esperadas das respostas aos itens em função

das médias de $\hat{\theta}_j - \hat{\delta}_i$ do item 14.

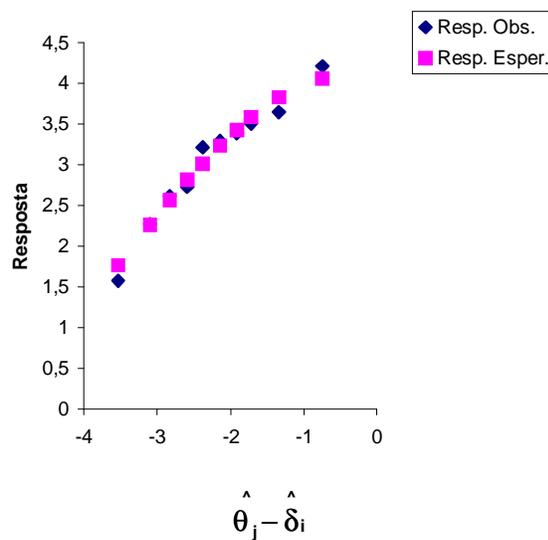


Figura 4.6: Médias observadas e esperadas das respostas aos itens em função das médias de $\hat{\theta}_j - \hat{\delta}_i$ do item 18.

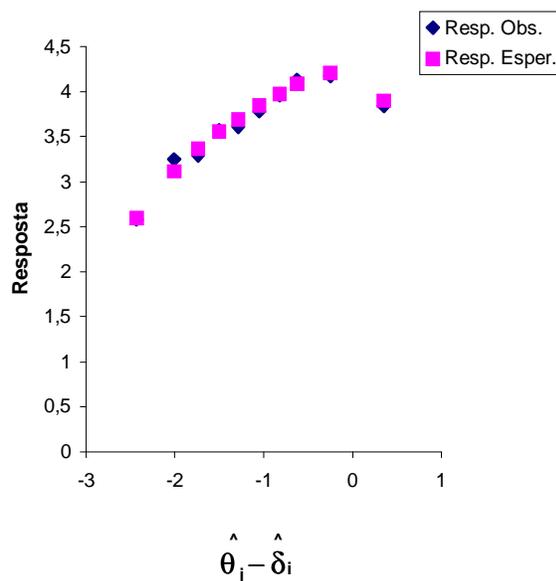


Figura 4.7: Médias observadas e esperadas das respostas aos itens em função das médias de $\hat{\theta}_j - \hat{\delta}_i$ do item 34.

Observam-se nas Figuras 4.5, 4.6 e 4.7 que os itens 14, 18 e 34 mostraram um excelente ajuste, como o revelado através de suas correlações ($r_{17} = 0,97$; $r_{18} = 0,98$; $r_{34} = 0,99$), não evidenciando nenhum *misfit*. Logo, os itens mostraram-se consistentes com as propriedades não monótonas do GGUM.

4.7 Interpretação dos parâmetros dos itens

Como já foi descrito na revisão de literatura, α_i é o parâmetro de discriminação do item i , δ_i parâmetro de locação do item i , θ_j parâmetro de locação do aluno j , neste caso, satisfação do aluno, e τ_{ik} parâmetro de locação dos limiares da categoria de respostas subjetivas num *continuum* de atitude relativa a locação do item i , e quando $\theta_j - \delta_i$ se aproxima de zero o indivíduo é mais propenso a estar mais satisfeito com o item.

De posse dos valores dos parâmetros podem-se fazer algumas considerações. Por exemplo:

Item 4: “Qualidade dos Planos de Ensino apresentados pelos professores”

As características e os resultados obtidos do item 4 são as seguintes:

Z_i = resposta observável de um item que pode assumir os valores 0, 1, 2, 3, 4, ou 5 conforme a concordância em um nível de categoria do item 4.

$$\delta_i = 1,083$$

$$\alpha_i = 1,279$$

$$\tau_{ik} = (-4,1; -3,1; -2,6; -1,4; 1,2)$$

De acordo com o modelo de desdobramento, a probabilidade de um aluno com parâmetro de locação θ_j escolher uma categoria z dentre as D categorias do item i é dada por:

$$P(Z = z | \theta_j) = \frac{\exp\left[\alpha_i\left(z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}\right)\right] + \exp\left[\alpha_i\left((M - z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}\right)\right]}{\sum_{v=0}^D \left[\exp\left[\alpha_i\left(v(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik}\right)\right] + \exp\left[\alpha_i\left((M - v)(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik}\right)\right] \right]}$$

A Figura 4.8 ilustra a função de probabilidade das categoria de resposta do item 4 em função de $\hat{\theta}_j - \hat{\delta}_i$.

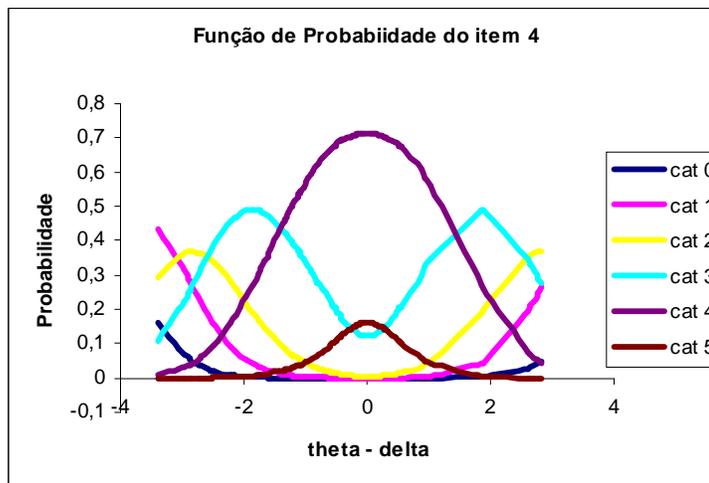


Figura 4.8 : Função de Probabilidade das categoria de resposta observável do item 4 em função de $\hat{\theta}_j - \hat{\delta}_i$.

Na Figura 4.8 observa-se que na função de probabilidade da categoria 4, quando $\hat{\theta}_j - \hat{\delta}_i$ se aproxima de 0 a função de probabilidade é maior em relação as outras categorias de respostas.

Para uma satisfação (θ) de 1,1544, $\hat{\theta}_j - \hat{\delta}_i = 0,0714$ (perto de zero) têm-se :

$$P (Z = 0 | 1,1544) = 4,47 \cdot 10^{-7}$$

$$P (Z = 1 | 1,1544) = 9,18 \cdot 10^{-5}$$

$$P (Z = 2 | 1,1544) = 0,004554$$

$$P (Z = 3 | 1,1544) = 0,121878$$

$$P (Z = 4 | 1,1544) = 0,712073$$

$$P (Z = 5 | 1,1544) = 0,161403$$

Desta forma, se forem reunidos alunos com uma satisfação igual a 1,1544, 71% responderiam que estão satisfeitos com as medidas adotadas para aprimorar a avaliação dos alunos nas disciplinas do curso, 16,1% estão totalmente satisfeitos com este item, 12,2% estão pouco satisfeitos, 0,4% estão pouco insatisfeitos, $4,74 \cdot 10^{-7}\%$ estão totalmente insatisfeitos, $9,18 \cdot 10^{-5}\%$ estão insatisfeitos.

Observa-se que para um $\theta = -2,3014$, isto é, para $\hat{\theta}_j - \hat{\delta}_i$ extremo como, isto é $\hat{\theta}_j - \hat{\delta}_i = -3,844$ têm-se:

$$P (Z = 0 | -2,3014) = 0,162688$$

$$P (Z = 1|-2,3014) = 0,432024$$

$$P (Z = 2|-2,3014) = 0,291581$$

$$P (Z = 3|-2,3014) = 0,105425$$

$$P (Z = 4|-2,3014) = 0,008257$$

$$P (Z = 5|-2,3014) = 2,52 \cdot 10^{-5}$$

Neste caso, alunos que têm $\theta(\text{satisfação}) = -2,3014$ a probabilidade é maior para a categoria 1, 43,2% responderiam que estão insatisfeitos com o item.

Ou, para $\theta = 3,8999$, assim $\hat{\theta}_j - \hat{\delta}_i = 2,8169$ têm-se as seguintes probabilidades:

$$P (Z = 0|3,8999) = 0,048099$$

$$P (Z = 1|3,8999) = 0,263944$$

$$P (Z = 2|3,8999) = 0,368119$$

$$P (Z = 3|3,8999) = 0,275041$$

$$P (Z = 4|3,8999) = 0,044513$$

$$P (Z = 5|3,8999) = 0,000285$$

Já neste caso, alunos com satisfação igual a 3,8999, a maior probabilidade ocorre para a categoria 2, isto é, a probabilidade de alunos que responderiam que estão pouco insatisfeitos com qualidade dos planos de ensino apresentados pelos professores que apresentam este parâmetro é de 36,8% e 27,5% dos alunos responderiam que estão pouco satisfeitos.

Item 23: “Medidas adotadas para aprimorar a avaliação dos alunos nas disciplinas do curso”

As características e os resultados obtidos do item 23 são as seguintes:

Z_i = resposta observável de um item que pode assumir os valores 0, 1, 2, 3, 4, ou 5 conforme a concordância em um nível de categoria do item 23.

$$\delta_i = 1,318$$

$$\alpha_i = 1,709$$

$$\tau_{ik} = (-3,1; -2,6; -2,1; -1,0; 1,5)$$

A probabilidade de um aluno com parâmetro de locação θ_j escolher uma categoria z dentre as D categorias do item i é dada por:

$$P(Z = z | \theta_j) = \frac{\exp\left[\alpha_i \left(z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik} \right)\right] + \exp\left[\alpha_i \left((M - z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik} \right)\right]}{\sum_{v=0}^D \left[\exp\left[\alpha_i \left(v(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik} \right)\right] + \exp\left[\alpha_i \left((M - v)(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik} \right)\right] \right]}$$

A Figura 4.9 ilustra a função de probabilidade das categoria de resposta do item 23 em função de $\hat{\theta}_j - \hat{\delta}_i$.

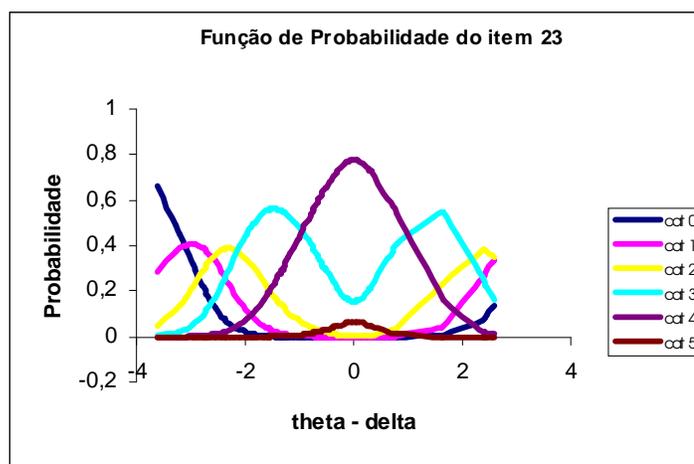


Figura 4.9 : Função de Probabilidade das categoria de resposta observável do item 23 em função de $\hat{\theta}_j - \hat{\delta}_i$.

Para $\theta(\text{satisfação}) = 1,274$, $\hat{\theta}_j - \hat{\delta}_i = -0,044$ (perto de zero) têm-se :

$$P(Z = 0 | 1,274) = 2,41 \cdot 10^{-7}$$

$$P(Z = 1 | 1,274) = 4,83 \cdot 10^{-5}$$

$$P(Z = 2 | 1,274) = 0,004085$$

$$P(Z = 3 | 1,274) = 0,154912$$

$$P(Z = 4 | 1,274) = 0,776803$$

$$P(Z = 5 | 1,274) = 0,064151$$

Assim, se forem reunidos alunos com satisfação igual a 1,274, 77,6% responderiam que estão satisfeitos com as medidas adotadas para aprimorar a avaliação dos alunos nas disciplinas do curso, 6,4% estão totalmente satisfeitos com este item, 15,5% estão pouco satisfeitos, 0,4% estão pouco insatisfeitos, $2,41 \cdot 10^{-7}\%$ estão totalmente insatisfeitos, $4,83 \cdot 10^{-5}\%$ estão insatisfeitos.

Observa-se que $\theta = -2,3014$, ou seja, para $\hat{\theta}_j - \hat{\delta}_i = -3,6194$ (extremo) têm-se:

$$P(Z = 0 | -2,3014) = 0,664378$$

$$P(Z = 1 | -2,3014) = 0,281533$$

$$P(Z = 2 | -2,3014) = 0,050073$$

$$P(Z = 3 | -2,3014) = 0,003975$$

$$P(Z = 4 | -2,3014) = 4,15 \cdot 10^{-5}$$

$$P(Z = 5 | -2,3014) = 7,11 \cdot 10^{-9}$$

Neste caso, alunos que têm θ (satisfação) = -2,3014 a probabilidade é maior para a categoria 0, 66% de alunos responderiam que estão totalmente insatisfeitos com item.

Ou para $\theta = 3,8999$, $\hat{\theta}_j - \hat{\delta}_i = 2,5819$ têm-se as seguintes probabilidades:

$$P(Z = 0 | 3,8999) = 0,135017$$

$$P(Z = 1 | 3,8999) = 0,336934$$

$$P(Z = 2 | 3,8999) = 0,352906$$

$$P(Z = 3 | 3,8999) = 0,164987$$

$$P(Z = 4 | 3,8999) = 0,010145$$

$$P(Z = 5 | 3,8999) = 1,03 \cdot 10^{-5}$$

Já neste caso, alunos com $\theta = 3,8999$, a maior probabilidade ocorre para a categoria 2, isto é, a probabilidade de alunos que responderiam que estão pouco insatisfeitos com as medidas adotadas para aprimorar a avaliação dos alunos nas disciplinas do curso que apresenta este parâmetro é de 35% e 34% dos alunos responderiam que estão insatisfeitos.

Item 28: “Mecanismos de atendimento e orientação acadêmica dos alunos de graduação que realiza ”

As características e os resultados obtidos do item 28 são as seguintes:

Z_i = resposta observável de um item que pode assumir os valores 0, 1, 2, 3, 4, ou 5 conforme a concordância em um nível de categoria do item 28.

$$\delta_i = 1,943$$

$$\alpha_i = 1,195$$

$$\tau_{ik} = (-3,5; -3,2; -2,8; -1,8; 0,2)$$

A probabilidade de um aluno com parâmetro de locação θ_j escolher uma categoria z dentre as D categorias do item i é dada por:

$$P(Z = z | \theta_j) = \frac{\exp\left[\alpha_i \left(z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}\right)\right] + \exp\left[\alpha_i \left((M-z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}\right)\right]}{\sum_{v=0}^D \left[\exp\left[\alpha_i \left(v(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik}\right)\right] + \exp\left[\alpha_i \left((M-v)(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik}\right)\right] \right]}$$

A Figura 4.10 ilustra a função de probabilidade das categoria de resposta do item 28 em função de $\hat{\theta}_j - \hat{\delta}_i$.

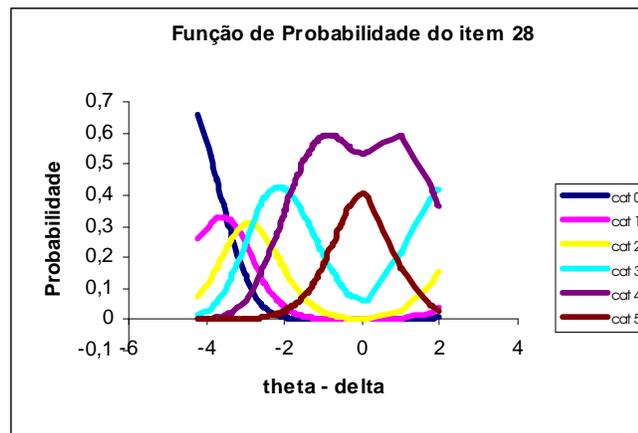


Figura 4.10: Função de Probabilidade das categoria de resposta observável do item 28 em função de $\hat{\theta}_j - \hat{\delta}_i$.

Portanto, para $\theta(\text{satisfação}) = 1,9493$, $\hat{\theta}_j - \hat{\delta}_i = -0,0063$ (perto de zero) têm-se :

$$P(Z = 0 | 1,9493) = 7,48 \cdot 10^{-7}$$

$$P(Z = 1 | 1,9493) = 4,7 \cdot 10^{-5}$$

$$P(Z = 2 | 1,9493) = 0,002086$$

$$P(Z = 3 | 1,9493) = 0,059577$$

$$P(Z = 4 | 1,9493) = 0,533777$$

$$P(Z = 5 | 1,9493) = 0,404512$$

Assim sendo, se forem reunidos alunos com $\theta = 1,9493$, 53% responderiam que estão satisfeitos com os mecanismos de atendimento e orientação acadêmica dos alunos no cotidiano dos cursos, 40% estão totalmente satisfeitos com este item, 6% estão pouco satisfeitos, 0,2% estão pouco insatisfeitos, $7,48 \cdot 10^{-7}\%$ estão totalmente insatisfeitos, $4,7 \cdot 10^{-5}\%$ estão insatisfeitos.

Observa-se que $\theta = -2,3014$ ($\hat{\theta}_j - \hat{\delta}_i = -4,2444$ extremos) têm-se:

$$P(Z = 0 | -2,3014) = 0,655999$$

$$P(Z = 1 | -2,3014) = 0,258469$$

$$P(Z = 2 | -2,3014) = 0,071927$$

$$P(Z = 3 | -2,3014) = 0,012879$$

$$P(Z = 4 | -2,3014) = 0,000723$$

$$P(Z = 5 | -2,3014) = 3,46 \cdot 10^{-6}$$

Neste caso, alunos que têm θ (satisfação) = -2,3014 a probabilidade é maior para a categoria 0, 66% de alunos responderiam que estão totalmente insatisfeitos com qualificação dos alunos para a escrita de trabalhos científicos durante o curso de graduação.

Para $\theta = 3,874$, ($\hat{\theta}_j - \hat{\delta}_i = 1,931$) têm-se as seguintes probabilidades:

$$P(Z = 0 | 3,874) = 0,005285$$

$$P(Z = 1 | 3,874) = 0,033049$$

$$P(Z = 2 | 3,874) = 0,145969$$

$$P(Z = 3 | 3,874) = 0,414814$$

$$P(Z = 4 | 3,874) = 0,370216$$

$$P(Z = 5 | 3,874) = 0,030667$$

Já neste caso, para um grau de satisfação 3,874 a maior probabilidade ocorre para a categoria 3, isto é, a probabilidade de alunos que responderiam que estão poucos satisfeitos com qualificação dos alunos para a escrita de trabalhos científicos durante o curso de graduação que apresenta este parâmetro é de 42% e 37% dos alunos responderiam que estão satisfeitos.

Item 15: “Qualificação dos alunos para a escrita de trabalhos científicos durante o curso de graduação”

As características e os resultados obtidos do item 15 são as seguintes:

Z_i = resposta observável de um item que pode assumir os valores 0, 1, 2, 3, 4, ou 5 conforme a concordância em um nível de categoria do item 15.

$$\delta_i = 3,130$$

$$\alpha_i = 1,352$$

$$\tau_{ik} = (-4,6; -4,3; -4,0; -2,6; 0,3)$$

A probabilidade de um aluno com parâmetro de locação θ_j escolher uma categoria z dentre as D categorias do item i é dada por:

$$P(Z = z | \theta_j) = \frac{\exp\left[\alpha_i \left(z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}\right)\right] + \exp\left[\alpha_i \left((M-z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}\right)\right]}{\sum_{v=0}^D \left[\exp\left[\alpha_i \left(v(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik}\right)\right] + \exp\left[\alpha_i \left((M-v)(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik}\right)\right] \right]}$$

A Figura 4.11 ilustra a função de probabilidade das categoria de resposta do item 15 em função de $\hat{\theta}_j - \hat{\delta}_i$.

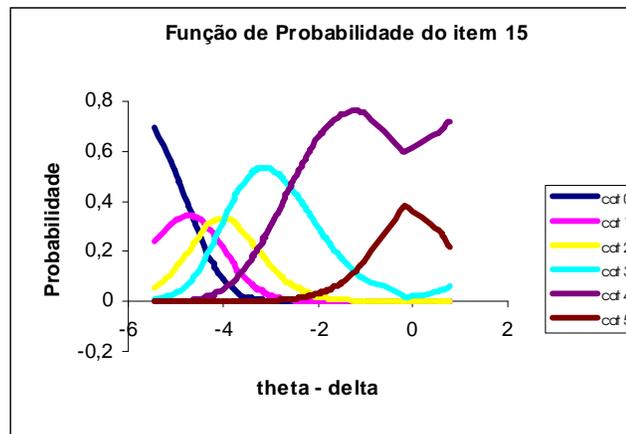


Figura 4.11: Função de Probabilidade das categoria de resposta observável do item 15 em função de $\hat{\theta}_j - \hat{\delta}_i$.

Nesse item se considerarmos $\theta(\text{satisfação}) = 2,9667$, $(\hat{\theta}_j - \hat{\delta}_i = -0,1633$ perto de zero) têm-se as seguintes probabilidades:

$$P(Z = 0 | 2,9667) = 7,91 \cdot 10^{-10}$$

$$P(Z = 1 | 2,9667) = 3,56 \cdot 10^{-7}$$

$$P(Z = 2 | 2,9667) = 0,000106$$

$$P(Z = 3 | 2,9667) = 0,019921$$

$$P(Z = 4 | 2,9667) = 0,598253$$

$$P(Z = 5 | 2,9667) = 0,38172$$

Deste modo, se forem reunidos alunos com satisfação 2,9667, 59,8% responderiam que estão satisfeitos com Qualificação dos alunos para a escrita de

trabalhos científicos durante o curso de graduação, 38% estão totalmente satisfeitos com este item.

Nota-se que para uma satisfação $\theta = -1,9632$, ($\hat{\theta}_j - \hat{\delta}_i = -5,0932$) a probabilidade para cada categoria de resposta é:

$$P(Z = 0 | -1,9632) = 0,559291$$

$$P(Z = 1 | -1,9632) = 0,306774$$

$$P(Z = 2 | -1,9632) = 0,109319$$

$$P(Z = 3 | -1,9632) = 0,023814$$

$$P(Z = 4 | -1,9632) = 0,000801$$

$$P(Z = 5 | -1,9632) = 5,48 \cdot 10^{-7}$$

Neste caso, alunos que têm $\theta(\text{satisfação}) = -2,3014$ a probabilidade é maior para a categoria 0, ou seja, 56% de alunos responderiam que estão totalmente insatisfeitos com a qualificação dos alunos para a escrita de trabalhos científicos durante o curso de graduação e 31% estão insatisfeitos.

Apresentaram-se 4 itens, isto é, 4, 23, 28, 15 de parâmetros 1,083, 1,318, 1,943 e 3,130 respectivamente, notou-se que a probabilidade de concordância do item era maior quando o parâmetro θ (satisfação) estava perto do item satisfazendo a primeira proposição básica do GGUM.

A Figura 4.12, em seguida, ilustra a distribuição de freqüência dos parâmetros θ (satisfação) obtidos dos alunos e os parâmetros dos itens colocados num mesmo *continuum* θ .

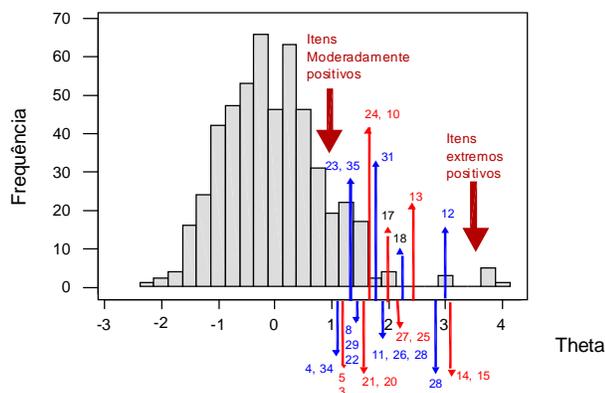


Figura 4.12: Distribuição dos parâmetros dos alunos e dos itens

Visualiza-se na Figura 4.12 que todos os itens tiveram seus conteúdos positivos e a maioria dos alunos tiveram sua opinião se concentrando de -1 a 1 .

Considerando que no modelo GGUM os níveis altos de concordância ocorre quando a posição aluno e a posição do item colocados no mesmo *continuum* latente, estão pertos. E, como, os 26 itens restantes, calibrados segundo este modelo foram colocados no *continuum* latente de acordo com seu conteúdo de moderadamente positivo a extremo positivo, não apresentando nenhum item com conteúdo negativo e os alunos respondentes tiveram sua opinião colocada neste mesmo *continuum* numa posição mais negativa em relação à posição dos itens, evidenciando uma certa distância em relação à posição do individuo e do item no *continuum* latente, então tal distribuição indica que os itens avaliados não foram, na maioria aceitos pelos alunos, isto é, não houve satisfação total da maior parte dos alunos em relação aos itens investigados.

Uma vez que os itens são pressupostos do indicativo do grau de satisfação dos alunos com relação ao curso, pode-se deduzir e verificar através das funções de probabilidade que os alunos de uma forma geral não estão satisfeitos com todos os itens.

5 CONCLUSÃO E RECOMENDAÇÕES

No Capítulo anterior verificou-se aplicação do modelo GGUM a dados reais. Observou-se que os itens que ficaram retidos na escala final mostraram-se consistentes com as propriedades do GGUM, que tem o processo de resposta baseado em proximidade, pois apresentaram uma função do valor esperado e observado em função de $\hat{\theta}_j - \hat{\delta}_i$, com características não monotônica de pico simples.

Observou-se que o modelo se ajustou razoável e adequadamente bem aos dados, pois não apresentou nenhum *misfit*, a correlação entre as respostas esperadas e observadas foram altas para cada item em questão, os itens apresentaram um *infit* e *outfit* próximos de 1 que é o esperado de acordo com a proposição encontrada na literatura.

A maioria dos modelos da TRI, especialmente modelos logísticos de 1, 2 e 3 parâmetros, necessitam de um número mínimo de itens, no caso 30, para se obter um modelo com um bom ajuste. Entretanto, verificou-se que no modelo de desdobramento pode-se obter um ótimo ajuste com um número menor de itens, com 20 ou 26 itens.

Dos 35 itens do questionário somente 26 itens ficaram retidos na escala final. Os itens 1, 2, 6, 7, 9, 19, 32 e 33 foram eliminados, pois apresentaram uma comunalidade menor que 0,3 e o item 30 foi eliminado antes das análises pelo fato de que alguns alunos deixaram de responder a este item. E dez alunos também foram retirados da análise por apresentarem um valor elevado no desvio padrão do parâmetro θ_j .

O modelo de desdobramento GGUM da Teoria da Resposta ao Item é um instrumento importante e pode contribuir expressivamente em muitas áreas não só na medida de satisfação, mas também em outras medidas de atitude.

O GGUM se ajusta adequadamente e corretamente aos dados, as estimativas dos itens são invariantes aos indivíduos e as estimativas dos indivíduos são invariantes aos itens, essas características sem dúvida facilitarão o uso de GGUM na criação de banco de itens. Sabe-se que existem outros fatores envolvidos na geração de itens para medidas de atitudes, porém o GGUM oferece uma grande contribuição, principalmente porque coloca os itens e os indivíduos num mesmo

continuum, os itens são colocados de acordo com seu conteúdo e os indivíduos são colocados de acordo com sua opinião ambos variando de negativo, neutro a positivo.

Neste trabalho foi apresentado o conceito básico do modelo de desdobramento GGUM da Teoria da Resposta ao Item, sendo, também, precursor na aplicação de modelo de desdobramento da TRI aqui no Brasil, uma vez que atualmente, quase todas as aplicações envolvem modelo de 1, 2 ou 3 parâmetros.

Dos resultados obtidos na escala final através do modelo GGUM na pesquisa em nível geral, obtiveram-se as seguintes conclusões:

- Na média quanto aos docentes, pode-se destacar que têm demonstrado conhecimento nas matérias que lecionam, tem qualidade nos planos de ensino, entretanto há necessidade de aperfeiçoar a metodologia no desenvolvimento das aulas na questão ensino – aprendizagem das disciplinas dos cursos, apresentar de forma organizada os conteúdos nas aulas; aperfeiçoar as medidas para a melhoria da qualidade do ensino; procurar medidas eficazes para superar as dificuldades dos alunos com deficiências nas disciplinas; aprimorar a avaliação dos alunos e melhorar a dinâmica das aulas para manter a atenção dos alunos. Quanto ao curso buscar procedimentos para a qualificação dos conteúdos desenvolvidos nas disciplinas. Aperfeiçoar os mecanismos de atendimento e orientação acadêmica no cotidiano dos cursos. Melhorar o comprometimento dos docentes com a qualificação do curso.
- Na média a pesquisa revelou que os alunos quando concluem o curso saem com um nível satisfatório, contudo eles devem ter mais seriedade acadêmica, que sentem necessidade de complementação de sua formação acadêmica, havendo, também necessidade de aprimorar a qualificação quanto à escrita de trabalhos científicos durante o curso, assim como melhorar a capacidade de leitura de textos científicos.
- Ainda quanto aos alunos eles estão pouco satisfeitos com o nível deles quando ingressam ao curso.

Para desenvolver e aperfeiçoar a aplicação do modelo de desdobramento em medidas de atitudes recomenda-se a realização de outras pesquisas, dentre as quais e relacionadas a este estudo assinalam-se:

- Ampliar o número de alunos, em nível geral, estendendo a outras unidades do CEFET – PR.
- Comparar os resultados do modelo de desdobramento da TRI com um outros modelos da TRI.
- Medir graus de satisfação dos alunos especificamente dentro de cada curso oferecido pelo CEFET – PR.
- Comparar graus de satisfação dos alunos entre os cursos oferecidos pelo CEFET – PR.
- Aplicar o modelo de desdobramento da TRI para medir satisfação dos alunos quanto ao ambiente e relações humanas, à pesquisa, comunicação e informação.
- Aplicar o modelo de desdobramento para medir satisfação dos docentes e técnicos administrativos quanto à pesquisa, ambiente, condições de trabalho, organização e objetivos institucionais.

REFERÊNCIAS

ADAMS, R. J & KHOO, S. T. **Quest the interactive test analyses system**. Hawthorn, Victoria: ACER, 1993.

ALEXANDRE, J.W.C.; ANDRADE, D.F; VASCONCELOS, AP.; ARAUJO, A M S. **Aplicação da TRI na Gestão de Qualidade: Proposta de um modelo probabilístico**. UFC, 2001.

ANDRADE, D. F. **Comparando o Desempenho de Grupos (Populações) de Respondentes Através da Teoria da Resposta ao Item**. Tese apresentada ao Departamento de Estatística e Matemática Aplicada da UFC para o Concurso de Professor Titular. Fortaleza, 1999.

ANDRADE, D. F. e KLEIN, R. Métodos estatísticos para avaliação educacional: teoria de resposta ao item. **Boletim da ABE**, 43,21-28, 1999.

ANDRADE, D. F. e TAVARES, H. R. e VALLE, R.C. **Teoria de Resposta ao Item: conceitos e aplicações**. ABE — Associação Brasileira de Estatística, 4º SINAPE, 2000.

ANDRADE, D. F e VALLE, R.C. **Introdução a Teoria de Resposta ao Item: Estudos em Avaliação Educacional**. São Paulo: Fundação Carlos Chagas, 18, p. 13-32, 1998.

ANDRICH, D. A rating formulation for ordered response categories. **Psychometrika**, **43**, p. 561-573, 1978.

ANDRICH, D. The application of an unfolding model of the PIRT type to the measurement of attitude. **Applied Psychological Measurement**, **12**, p. 33-51, 1988.

ANDRICH, D. A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. **British Journal of Mathematical and Statistical Psychology**, **49**, 347-365, 1996.

ANDRICH, D & LUO, G. A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. **Applied Psychological Measurement**, **17**, p. 253-276, 1993.

ASSUNÇÃO, C. N. B. **Estimação dos parâmetros de modelos de Teoria de Resposta ao item e Aplicações**. Dissertação de Mestrado, UFMG, 1999.

BAYLEY, Scott. Measuring customer satisfaction". **Evaluation Journal of Australasia**, v. 1, (new series), n. 1, march, 2001.

BIRNBAUM, A., **Some latent Trait Models and Their Use in Inferring an Examinee's Ability**. In F. M. Lord and M. R. Novick. Statistical theories of Mental Test Score. Reading, M.A: Addison-Wesley, 1968.

BOCK, R. D. & AITKIN, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. **Psychometrika**, **46**, p. 443-459, 1981.

BOCK, R. D. & LIEBERMAN, M. Fitting a response model for n dichotomously scored items. **Psychometrika**, **35**, p. 179-197, 1970.

BOCK, R. D. & MISLEVY, R. J. Adaptive EAP Estimation of ability in a microcomputer environment. **Applied Psychological Measurement**, **6**, p. 431-444, 1982.

CLIFF, N, COLLINS, L. M., ZATKIN, J., GALLIPEAU, D., & McCORMICK, D. J. An ordinal scaling method for questionnaire and other ordinal data. **Applied Psychological Measurement**, **12**, p. 83-97, 1988.

COOMBS, C. H. **A theory of data**. New York: Wiley, 1964.

COSTA, M. B. F. **Técnica derivada da Teoria de Resposta ao Item (TRI) aplicada ao setor de serviços**. Dissertação de Mestrado, UFPR, Curitiba, 2001.

DeROOS, Yosikazu & MEARES, Paula Allen. "Application of Rasch analysis: exploring differences in depression between african - american and white children". **Journal of Social Service Research**, v. 23, n 3/4, p. 93-107, 1998.

DESARBO, W. S., & HOFFMAN, D. L. Constructing MDS joint spaces from binary choice data: A multidimensional unfolding threshold model for marketing research. **Journal of Marketing Research**, 24, p. 40-54, 1987.

DONOGHUE, J.R. **Establishing two important properties of two IRT-based models for unfolding data**. Manuscript submitted for publication. 1999.

GRANGER, Carl V. & DEUTSCH, Anne. "Rasch analysis of the functional independence measure (FIMTM) mastery test". **Arch Phys med Rehabil**, January, v.79, p.52-57, 1998.

HAMBLETON, R. K. & SWAMINATAN, H. **Item Response Theory Principles and Applications**. Boston: Kluwer Academic Publishers, 1985.

HAMBLETON, R. K. & SWAMINATAN, H.. and ROGERS, H. J. **Fundamentals of Item Response Theory**. Newbury Park: Sage Publications, 1991.

HENARD, D. H. Item Response Theory. **Reading and Understanding more multivariate Statistics**, Washington, Dc., American Psychological Association, p. 67-97, 2000.

KLEIN R. & FONTANIVE N.S. S.Avaliação em Larga Escala: uma Proposta inovadora. Disponível em <[http:// www. Est.ufmg.Br/broav/avalesec.html](http://www.Est.ufmg.Br/broav/avalesec.html)>. Acesso em 04 dez. 2001.

LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**, 140, p. 5-53, 1932.

LINACRE, J. M. & WRIGHT, B. D. **A user's guide to BIGSTEPS (Version 2.4)**. Chicago: MESA PRESS, 1994.

LINDEN, W.J. Van Der and HAMBLETION, R. K. **Handbook of Modern Item Response Theory**. New York: Springer-Verlag, 1997.

LORD, F. M. **Applications of Item Response Theory to Practical Testing Problems**. Hillsdale NJ: Erlbaum, 1980.

LORD, F. M. and NOVICK, M. R. **Statistical Theories of Mental Test Score**. Reading: Addison-Wesley, 1968.

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

MATOS, G. S. **Teoria de Resposta ao Item: Uma Proposta de Modelo Multivariado**. Pernambuco DE/UFPE, Tese de Mestrado, UFC2001.

MISLEVY, R. J. Recent developments in the factor analysis of categorical variables. **Journal of Educational Statistics**, **11**, p. 3 -31, 1986.

MISLEVY, R. J. & BOCK, R. D. **PC BILQG 3: Item Analysis and Test Scoring with Binary Logistic Models**. Chicago: Scientific Software, Inc., 1990.

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

NOJOSA, Ronald T. **Modelos Multidimensionais para a Teoria de Resposta ao Item**. Pernambuco, UFPE, Tese de Mestrado, 2001.

NOVICK, M. R. The axioms and principal results of classical test theory. **Journal of Mathematical Psychology**, **3**, p.1-18, 1966.

RAO C. R. **Statistical inference and its applications** (2nd ed.). New York: Wiley, 1973.

RASCH, G. **Probabilistic Models for some Intelligence and Attainment Test.** Copenhagen: Danish Institute for Educational Research, 1960.

RECKASE, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. **Journal of Educational Statistics**, 4, p. 207-230, 1979.

ROBERTS, J. S. **Item response theory approaches to attitude measurement.** (Doctoral dissertation, University of South Carolina, Columbia), Dissertation Abstracts International, 56, 7089B, 1995.

ROBERTS, J. S. **GGUM2000: Software to estimate parameters of the generalized graded unfolding model.** Under review, 2000.

ROBERTS, J. S. **Equating parameter estimates from the generalized graded unfolding model.** Paper presented at the annual meeting of the American Educational Research Association, April 14, Seattle, Washington, 2001a.

ROBERTS, J. S. GGUM2000: Software to Estimation of parameters in the generalized graded unfolding model. **Applied Psychological Measurement**, 25, 38, 2001b.

ROBERTS, J. S., DONOGHUE, J.R. & LAUGHLIN, J. E. **A generalized item response model for unfolding responses from a graded scale.** Paper presented at the 61st annual meeting of the Psychometric Society, June 28, Banff, Alberta, Canada, 1996.

ROBERTS, J. S., DONOGHUE, J.R. & LAUGHLIN, J. E. The generalized graded unfolding model: A general parametric item response model for unfolding graded responses (RR-98-32). Princeton, NJ: **Educational Testing Service**, 1998.

ROBERTS, J. S., DONOGHUE, J.R. & LAUGHLIN, J. E. **Estimability of parameters in the generalized graded unfolding model.** Presented at the annual meeting of the American Educational Research Association, April 22, Montreal, Canada, 1999a.

ROBERTS, J. S., DONOGHUE, J.R. & LAUGHLIN, J. E. **Estimating parameters in the generalized graded unfolding model: Sensitivity to the prior distribution assumption and the number of quadrature points used.** Paper Presented at the Annual Meeting of the National Council on Measurement in Education, April 22, Montreal, PQ, Canada, 1999b.

ROBERTS, J. S., DONOGHUE, J.R. & LAUGHLIN, J. E. A general model for unfolding Unidimensional polytomous responses using item response theory. **Applied Psychological Measurement, 24**, p. 3-32, 2000.

ROBERTS, J. S. & LAUGHLIN, J. E. A. Unidimensional item response model for unfolding responses from a graded disagree-agree response scale. **Applied Psychological Measurement, 20**, p. 231-255, 1996a.

ROBERTS, J. S. & LAUGHLIN, J. E. The graded unfolding model: A Unidimensional item response model for unfolding graded responses (Research Rep. RR-96-16). Princeton N J: **Educational Testing Service**, 1996b.

ROBERTS, J. S., LAUGHLIN, J. E. & WEDELL, D. H. Validity issues in the Likert and Thurstone Approaches to Attitude Measurement. **Educational and Psychological Measurement, 59**, p. 211-233, 1999.

ROBERTS, J. S., LIN, Y. & LAUGHLIN, J. E. Computerized adaptive testing with the generalized graded unfolding model. **Applied Psychological Measurement, 25**, p. 177-196, 2001.

ROBERTS, J. S., WEDELL, D. H., LAUGHLIN, J. E. **Heightened Sensitivity of Likert Attitude Scales to Restriction of Sample Range.** Paper presented at the 1998 American Educational Research Association Annual Meeting, April 17, San Diego, California, 1998.

ROMANO, C. A. **O Desafio de uma nova proposta para a graduação na Educação Profissional: O caso do CEFET – PR.** Florianópolis, 151 f. Dissertação

de mestrado em Engenharia de Produção – Programa de Pós – Graduação, UFSC, 2000.

SAMEJIMA, F. A. Estimation of latent ability using a response pattern of graded scores. **Psychometric Monograph**, **17**, 1969.

SOARES, J. F.; MARTINS, M. I.; ASSUNÇÃO, C. N. B.. Heterogeneidade acadêmica dos alunos admitidos na UFMG e PUC-MG. **Estudos em Avaliação Educacional**. São Paulo: Fundação Carlos Chagas, **17**, 61-72, 1998.

SPEARMAN, C. The proof and measurement of association between two things. **American Journal of Psychology**, **15**, p. 72-101, 1904.

TAVARES, H. R. **Teoria da Resposta ao Item para dados longitudinais**. Tese de doutorado. São Paulo: IME/USP, 2001.

THURSTONE, L. L. Attitudes can be measured. **The American Journal of Sociology**, **26**, 249-269, 1928.

VALLE, Raquel da Cunha. **Teoria de Resposta ao Item**. São Paulo. Dissertação de Mestrado. IME/USP, 1999.

VAN DER LINDEN, w. J. and HAMBLETON, R. K. **Handbook of Modern Item Response Theory**. New York: Springer- Verlag, 1997.

VAN SCHUUR, W. H. **Structure in political beliefs: A new model for stochastic unfolding with application to European party activists**. Amsterdam: CT Press, 1984.

VAN SCHUUR, W. H. Nonparametric Unidimensional unfolding for multicategory data. In J. R. Freeman (Ed.), **Political Analysis**, **4**, p. 41-74, Ann Arbor MI: University of Michigan Press, 1993.

VAN SCHUUR, W. H., & KIERS, H. A. L. Why factor analysis is often the incorrect model for analyzing bipolar concepts, and what model can be used instead. **Applied Psychological Measurement**, **18**, p. 97-110, 1994.

VERHELST, N. D., & VERSTRALEN, H. H. F. M. A stochastic unfolding model derived from the partial credit model. **Kwantitative Methoden**, **42**, p. 73-92, 1993.

VIANNA, H.. M. Testes em Educação. São Paulo: IBRASA, 1987.

XIMÉNEZ, M. C. e REVUELTA, J. Aplicación de un modelo politômico de TRI para la obtención de medidas conmensurables en una escala de ajuste persona-organización. **Psicológica**, **20**, 135-150, 1999.

ZIMOWSKI, M. F., MURAKI, E. MISLEVY, R. J. & BOCK, R. D. **BILOG-MG: Multiple-Group IRT Analysis and Test Maintenance for Binary Items**. Chicago: Scientific Software, Inc, 1996.

APÊNDICES

APÊNDICE A – Questionário

Questionário nº _____ (Não preencher esta linha)

Caro Aluno (a):

O presente questionário se destina a uma pesquisa para elaboração de uma dissertação de mestrado sobre a satisfação dos alunos do CEFET com relação aos cursos de Tecnologia nas áreas oferecidas pelo CEFET - Paraná. Solicito a você obséquio de responder com a maior honestidade possível. A sua contribuição será de grande importância no sentido de oferecer subsídios para a melhoria desses cursos.

A partir de sua experiência pessoal no CEFET- PR, indique o grau da satisfação ou insatisfação que você sente em

relação a cada um dos indicadores ou questões propostas a seguir:

Onde: 1 Totalmente Insatisfeito; 2 Insatisfeito; 3 pouco insatisfeito; 4 pouco satisfeito; 5 satisfeito; 6 totalmente satisfeito

	1	2	3	4	5	6
1. Qualidade do curso de graduação que realiza;						
2. Estrutura curricular (de disciplinas) do curso;						
3. Organização na exposição de conteúdos pelos docentes;						
4. Qualidade dos Planos de Ensino apresentados pelos professores;						
5. Metodologia para o desenvolvimento do ensino-aprendizagem nas aulas;						
6. Formas de avaliação utilizadas nas disciplinas para “medir” os níveis de aprendizagem dos alunos;						
7. Alternativas oferecidas aos alunos para a complementação de sua formação global;						
8. Criatividade demonstrada pelos docentes no desempenho das atividades de ensino;						
9. Notas obtidas nas disciplinas em relação à aprendizagem alcançada;						
10. Seriedade acadêmica manifestada pelos docentes do curso;						
11. Nível de formação atingido pelos alunos que concluem o curso;						
12. Seriedade acadêmica dos alunos do curso;						

13. Nível de formação dos alunos quando ingressam no curso;						
14. Capacidade manifestada pelos alunos para a leitura de textos científicos durante o curso de graduação;						
15. Qualificação dos alunos para a escrita de trabalhos científicos durante o curso de graduação;						
16. Qualificação manifestada pelos alunos para a elaboração de monografia e/ou trabalho de conclusão de curso;						
17. Oportunidade de iniciação dos alunos na pesquisa no curso que realiza;						
18. Oportunidade de treinamento e inserção no mercado de trabalho oferecido pelo curso;						
19. Número de alunos que concluem o curso a cada ano;						
20. Medidas adotadas para a melhoria da qualidade do ensino no curso;						
21. Medidas adotadas para superar as dificuldades dos alunos com deficiências nas disciplinas;						
22. Medidas adotadas para aprimorar a metodologia das aulas nas disciplinas do curso;						
23. Medidas adotadas para aprimorar a avaliação dos alunos nas disciplinas do curso;						
24. Procedimentos adotados pelo curso para a qualificação dos conteúdos desenvolvidos nas disciplinas;						
25. Condições dos alunos para a dedicação ao curso de graduação;						
26. Tempo dedicado ao estudo das disciplinas que cursa;						
27. Iniciativa dos alunos para a complementação de sua formação acadêmica;						
28. Mecanismos de atendimento e orientação acadêmica dos alunos no cotidiano dos cursos;						
29. Comprometimento efetivo dos docentes com a qualificação do curso de graduação que realiza;						
30. Conhecimento da situação dos alunos que já concluíram o curso no mercado de trabalho;						
31. Satisfação em relação ao curso que está realizando;						
32. Limpeza e estado de conservação da sala de aula;						
33. Pontualidade e assiduidade dos docentes nas aulas;						
34. Conhecimento demonstrado pelos docentes nas matérias que lecionam;						
35. Dinâmica das aulas para manter a atenção dos alunos;						

Marque com "X" item correspondente ou responda a situação atual

36. Sexo: () 1. Feminino () 2. Masculino

37. Trabalha: () 1. Sim () 2. Não

38. Idade_____

39. Onde mora_____

49. Curso que faz no CEFET: Tecnologia em

APÊNDICE B - Alguns resultados dos parâmetros dos itens do modelo GGUM segundo o software GGUM2000

ITEM PARAMETER FILE

***** TODAY'S DATE: 11/15/2003 TODAY'S TIME: 15:37:59 *****

ITEM# 3 INITIAL= .89530945 DELTA= 1.23000354 DSTD= .09586627 ALPHA= 1.32254164 ASTD= .13346414 CORR= -.21411520
THRESHOLD# 1 INITIAL= .00000000 TAU= .00000000 TAUSTD= .00000000
THRESHOLD# 2 INITIAL= -3.33333368 TAU= -4.26144973 TAUSTD= .47177509
THRESHOLD# 3 INITIAL= -2.66666695 TAU= -2.81677339 TAUSTD= .18575249
THRESHOLD# 4 INITIAL= -2.00000021 TAU= -2.50951505 TAUSTD= .16701827
THRESHOLD# 5 INITIAL= -1.33333347 TAU= -1.25058106 TAUSTD= .12166455
THRESHOLD# 6 INITIAL= -.66666674 TAU= 1.15200825 TAUSTD= .24884251
ITEM# 4 INITIAL= .52986145 DELTA= 1.08258683 DSTD= .10056247 ALPHA= 1.27904681 ASTD= .12996630 CORR= -.20162025
THRESHOLD# 1 INITIAL= .00000000 TAU= .00000000 TAUSTD= .00000000
THRESHOLD# 2 INITIAL= -3.33333368 TAU= -4.14823343 TAUSTD= .59371114
THRESHOLD# 3 INITIAL= -2.66666695 TAU= -3.07719169 TAUSTD= .23950072
THRESHOLD# 4 INITIAL= -2.00000021 TAU= -2.58941693 TAUSTD= .18035892
THRESHOLD# 5 INITIAL= -1.33333347 TAU= -1.39257882 TAUSTD= .12767463
THRESHOLD# 6 INITIAL= -.66666674 TAU= 1.15424037 TAUSTD= .24046146
ITEM# 5 INITIAL= .69999695 DELTA= 1.16179466 DSTD= .11945558 ALPHA= 1.02390548 ASTD= .10816800 CORR= -.13374798
THRESHOLD# 1 INITIAL= .00000000 TAU= .00000000 TAUSTD= .00000000
THRESHOLD# 2 INITIAL= -3.33333368 TAU= -3.86191000 TAUSTD= .54069179
THRESHOLD# 3 INITIAL= -2.66666695 TAU= -3.32640602 TAUSTD= .32016863
THRESHOLD# 4 INITIAL= -2.00000021 TAU= -2.73906535 TAUSTD= .20632587
THRESHOLD# 5 INITIAL= -1.33333347 TAU= -1.28168949 TAUSTD= .15142860
THRESHOLD# 6 INITIAL= -.66666674 TAU= 1.53424068 TAUSTD= .34876386
ITEM# 8 INITIAL= 1.03034973 DELTA= 1.35213460 DSTD= .11942713 ALPHA= 1.22746109 ASTD= .11501895 CORR= -.12724795
THRESHOLD# 1 INITIAL= .00000000 TAU= .00000000 TAUSTD= .00000000
THRESHOLD# 2 INITIAL= -3.33333368 TAU= -3.76588212 TAUSTD= .37109744
THRESHOLD# 3 INITIAL= -2.66666695 TAU= -2.77397808 TAUSTD= .19496405
THRESHOLD# 4 INITIAL= -2.00000021 TAU= -2.61668489 TAUSTD= .17931886
THRESHOLD# 5 INITIAL= -1.33333347 TAU= -1.22604469 TAUSTD= .14503602
THRESHOLD# 6 INITIAL= -.66666674 TAU= 1.02027556 TAUSTD= .24977411
ITEM# 10 INITIAL= .52986145 DELTA= 1.63871409 DSTD= .15052054 ALPHA= .97791174 ASTD= .09177256 CORR= -.20186109
.00000000 TAU= .00000000 TAUSTD= .00000000
TAU= -4.35444464 TAUSTD= .56323218
-3.36803101 TAUSTD= .29622526
TAUSTD= .24536947
TAUSTD= .20518742
TAUSTD= .23117986

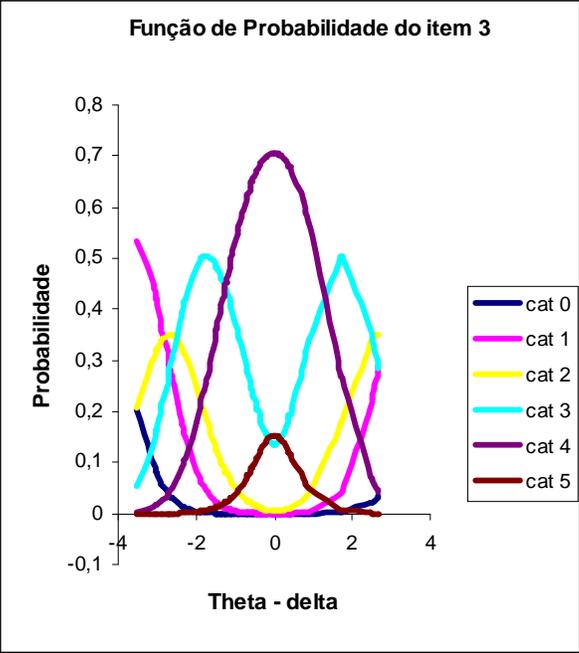
PERSON PARAMETER FILE

***** TODAY'S DATE: 11/15/2003 TODAY'S TIME: 15:37:59 *****

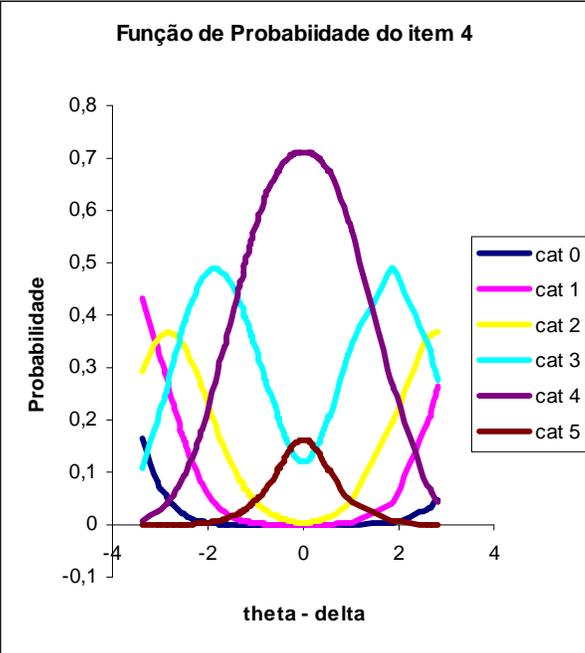
SUBJECT# 1058 THETA= -.82832734 TSTD= .16789192
SUBJECT# 1059 THETA= -.08919982 TSTD= .19545911
SUBJECT# 1060 THETA= -.47677143 TSTD= .17932064
SUBJECT# 1061 THETA= -.31658017 TSTD= .19281831
SUBJECT# 1062 THETA= -.34206097 TSTD= .18498895
SUBJECT# 1063 THETA= -.03975481 TSTD= .19723593
SUBJECT# 1066 THETA= .18567174 TSTD= .20606911
SUBJECT# 1067 THETA= -.54598854 TSTD= .17650491
SUBJECT# 1068 THETA= .09597974 TSTD= .20256478
SUBJECT# 1069 THETA= -.23069456 TSTD= .18945502
SUBJECT# 1070 THETA= -.59817956 TSTD= .17506419
SUBJECT# 1071 THETA= .20525521 TSTD= .20717088
SUBJECT# 1072 THETA= -.10900947 TSTD= .19422048
SUBJECT# 1073 THETA= 1.15444587 TSTD= .36355908
SUBJECT# 1074 THETA= 1.48853675 TSTD= .19306881
SUBJECT# 1075 THETA= 1.47695505 TSTD= .22860175
SUBJECT# 1076 THETA= .20038051 TSTD= .20631730
SUBJECT# 1077 THETA= .91338634 TSTD= .28398875
SUBJECT# 1079 THETA= 1.20239402 TSTD= .29175927
SUBJECT# 1080 THETA= 1.42016552 TSTD= .37962473
SUBJECT# 1081 THETA= 1.31345297 TSTD= .22242786
SUBJECT# 1082 THETA= -.25652117 TSTD= .18932539
SUBJECT# 1083 THETA= 1.29534301 TSTD= .38665484
SUBJECT# 1084 THETA= 1.36617271 TSTD= .61704683
SUBJECT# 1085 THETA= -.04767981 TSTD= .19655975
SUBJECT# 1086 THETA= -.35884168 TSTD= .18439826
SUBJECT# 1088 THETA= -.25487405 TSTD= .18799094
SUBJECT# 1089 THETA= .44987574 TSTD= .22644999
SUBJECT# 1090 THETA= .84484760 TSTD= .22863550
SUBJECT# 1091 THETA= -.47824864 TSTD= .17898344
SUBJECT# 1093 THETA= -.63182197 TSTD= .17354957
SUBJECT# 1094 THETA= .86515451 TSTD= .20274526
SUBJECT# 1095 THETA= .10449984 TSTD= .20228689
SUBJECT# 1096 THETA= -1.34134157 TSTD= .16431068
SUBJECT# 1097 THETA= .34864545 TSTD= .21471119
SUBJECT# 1098 THETA= -.77437934 TSTD= .16943354
SUBJECT# 1100 THETA= -.46419577 TSTD= .17985487
SUBJECT# 1101 THETA= .36201907 TSTD= .21000617
SUBJECT# 1102 THETA= -.10381459 TSTD= .19422459
SUBJECT# 1104 THETA= -1.27356383 TSTD= .16392011
SUBJECT# 1105 THETA= .29145502 TSTD= .21072900
SUBJECT# 1106 THETA= -.16277277 TSTD= .19217664
SUBJECT# 1107 THETA= .98445902 TSTD= .22164770
 THETA= .61629160 TSTD= .21361401
 = -.33657307 TSTD= .18470752
 .01442177 TSTD= .19800767
 TSTD= .19019429

APÊNDICE C – Gráficos das Funções de Probabilidades dos itens

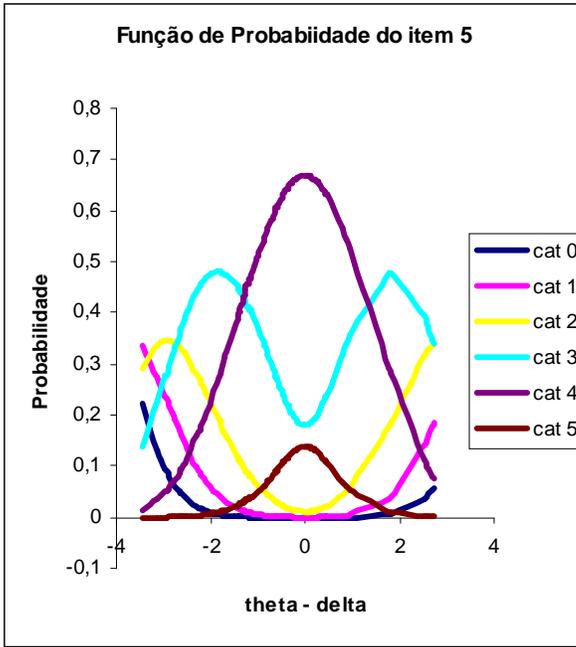
Item 3



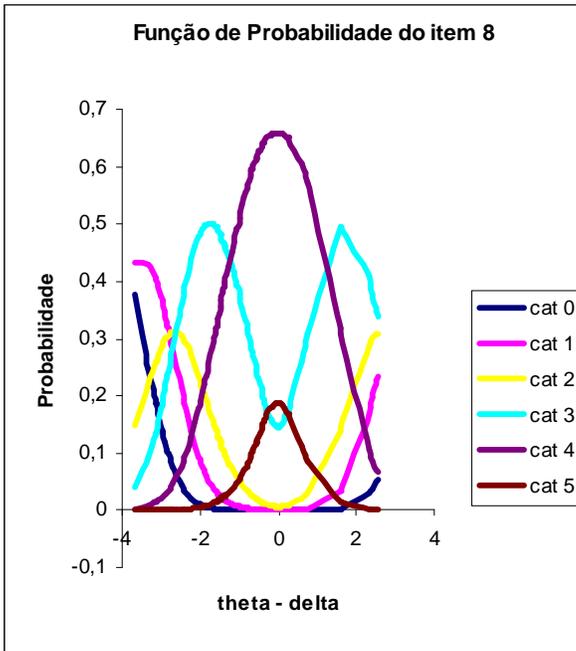
Item 4



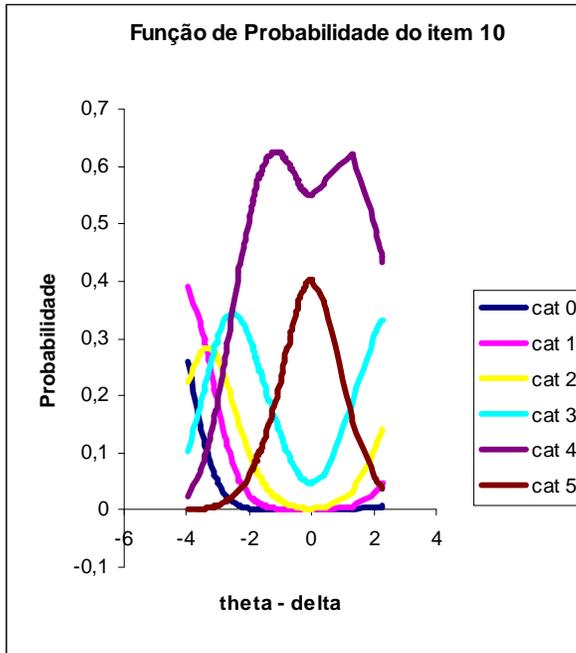
Item 5



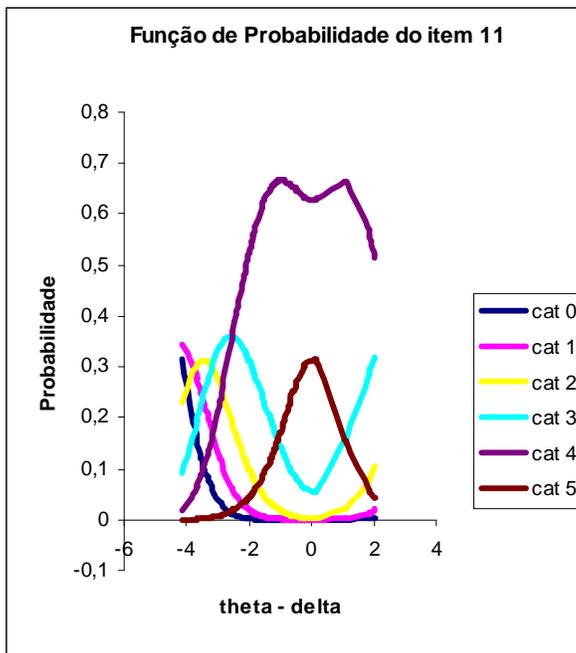
Item 8



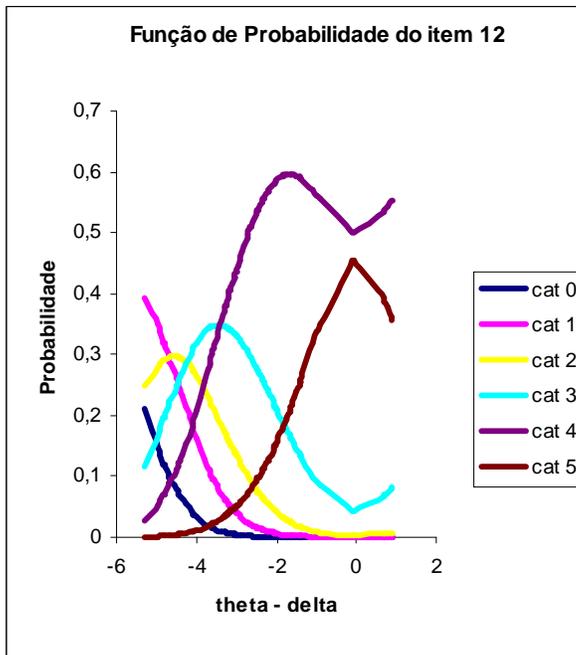
Item 10



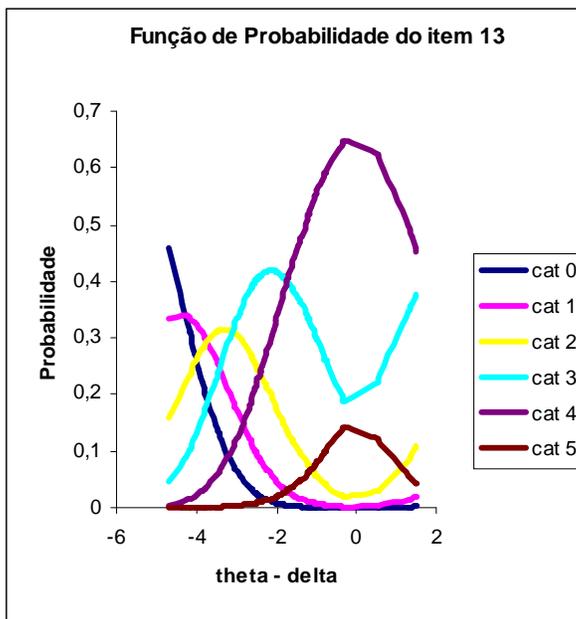
Item 11



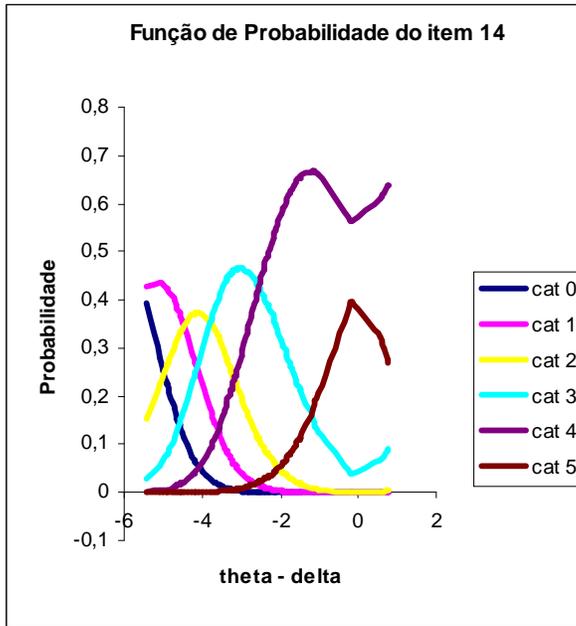
Item 12



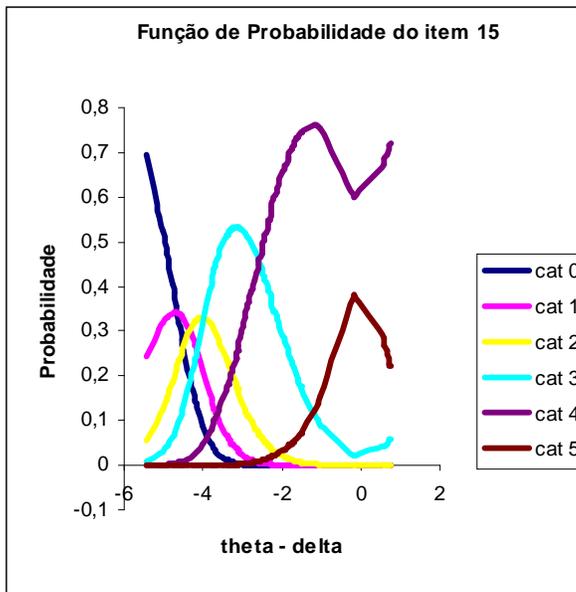
Item 13



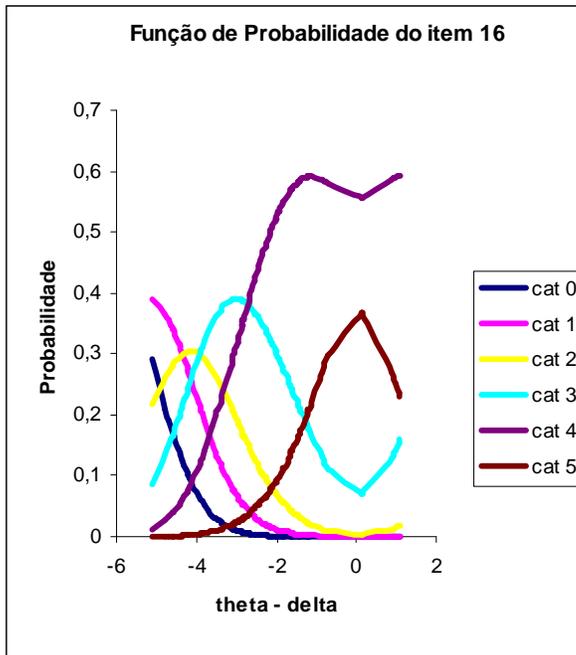
Item 14



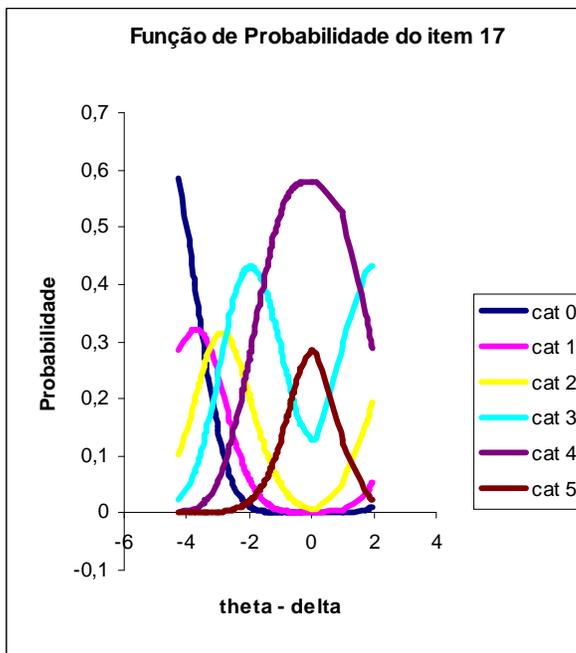
Item 15



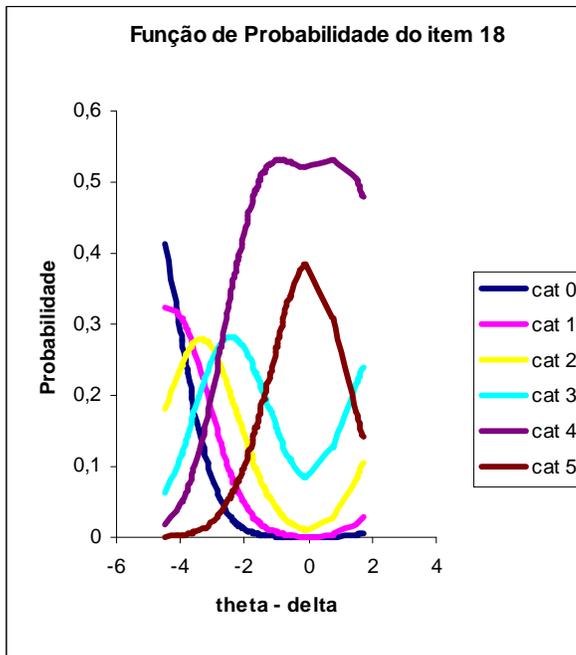
Item 16



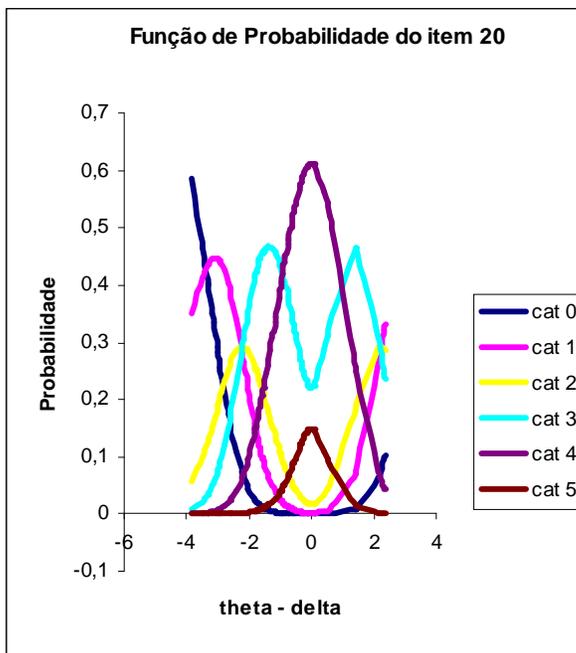
Item 17



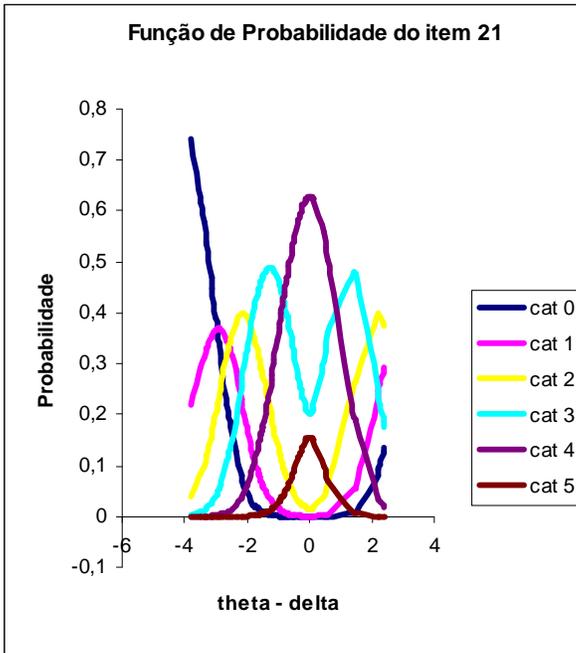
Item 18



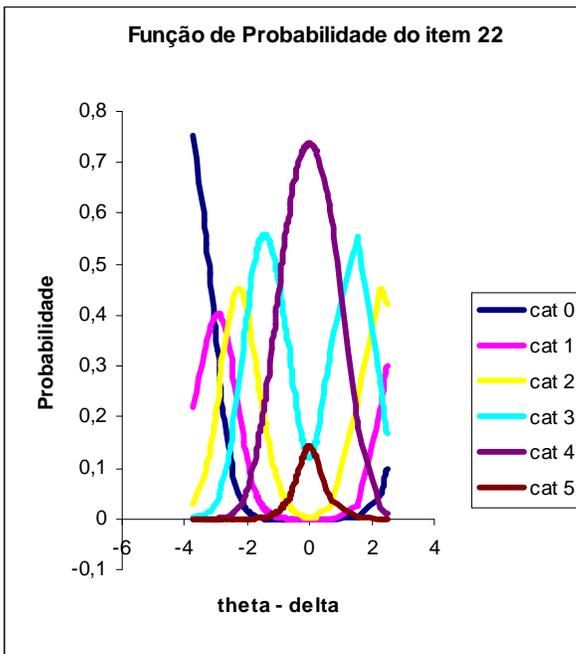
Item 20



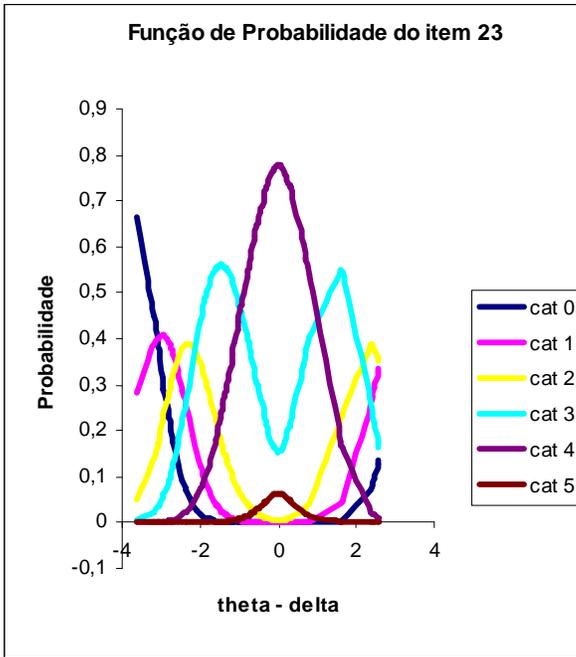
Item 21



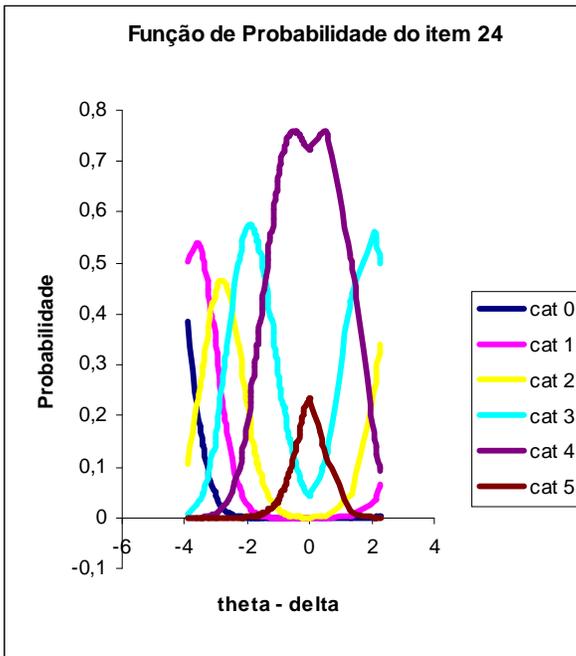
Item 22



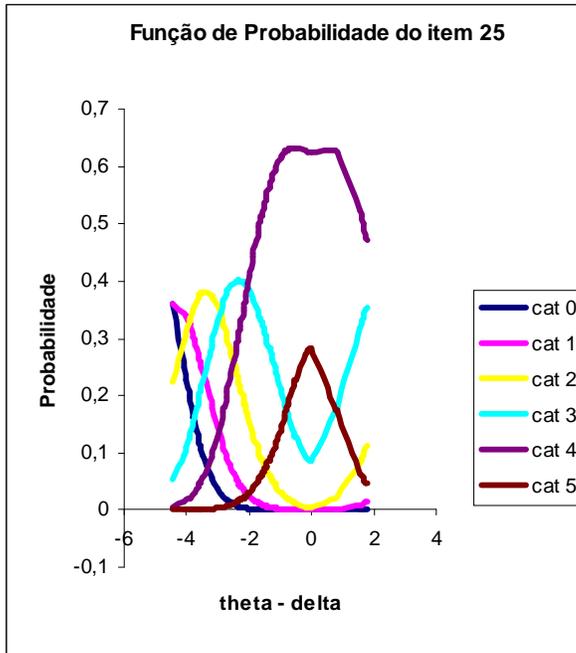
Item 23



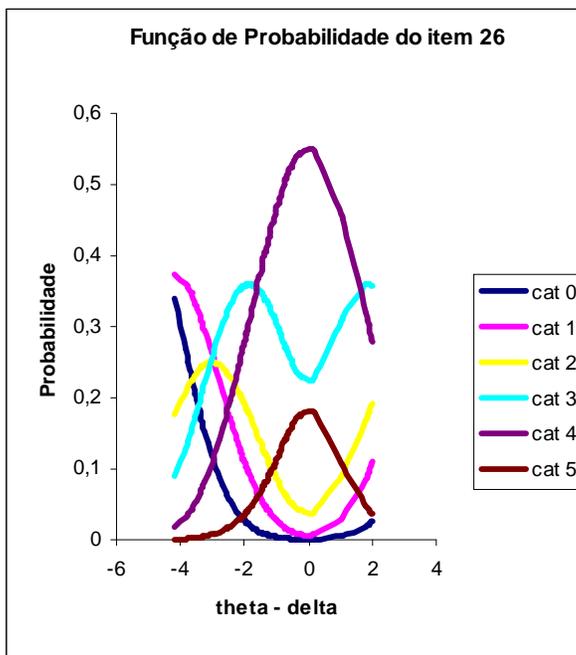
Item 24



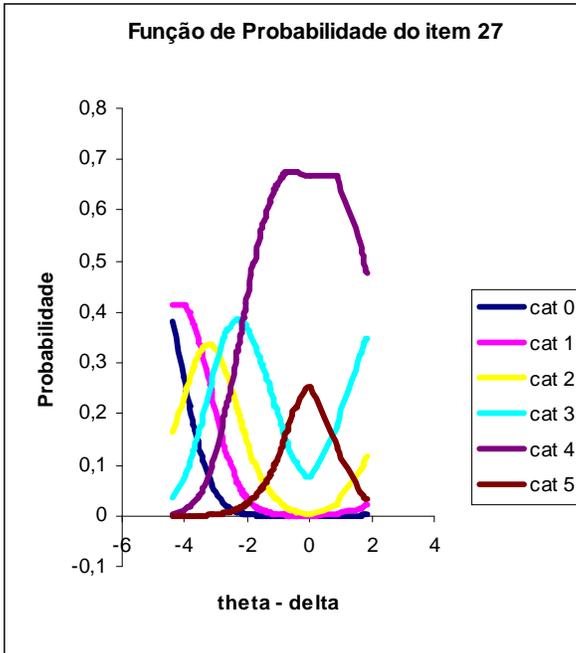
Item 25



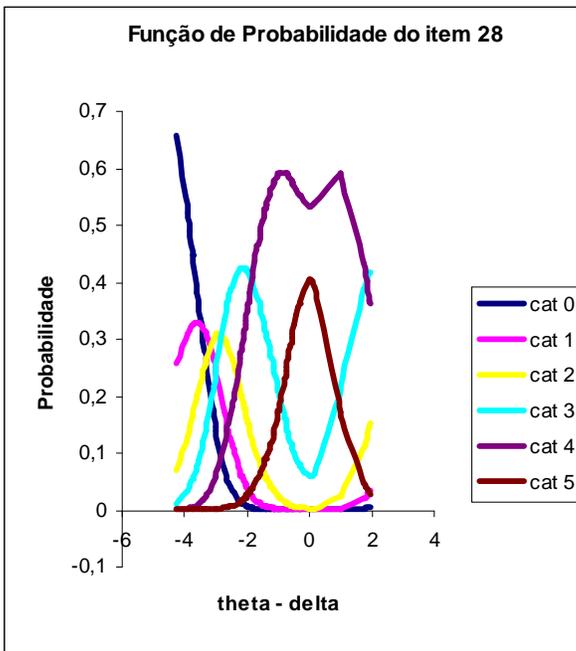
Item 26



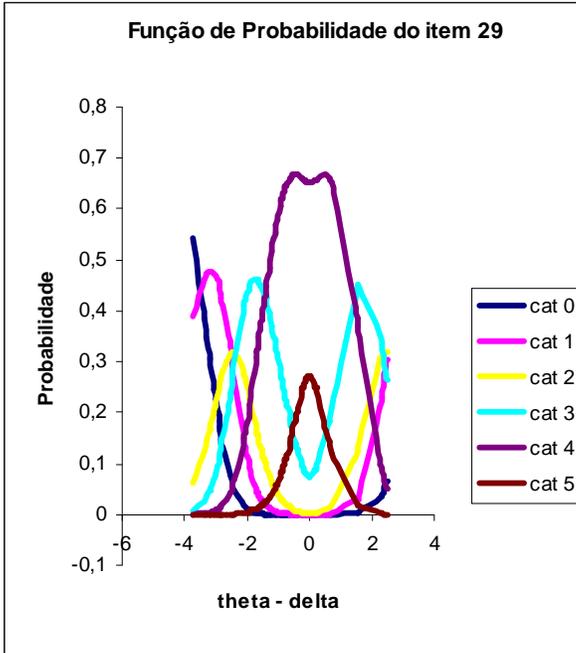
Item 27



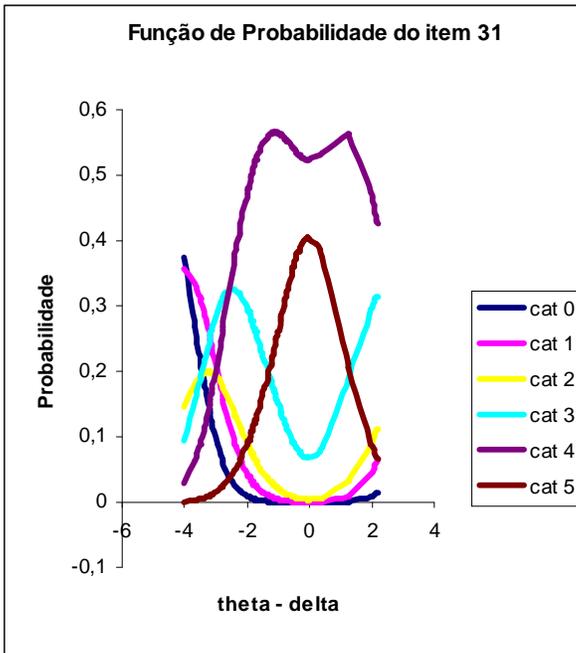
Item 28



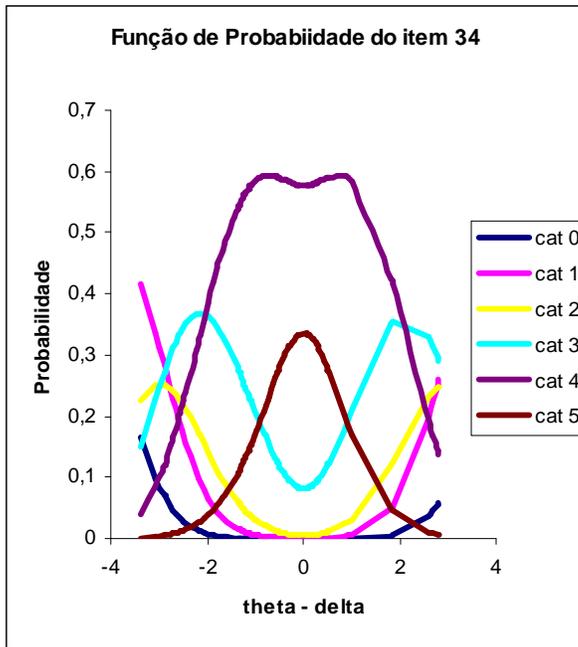
Item 29



Item 31



Item 34



Item 35

