

# Aplicação de um Modelo de Desdobramento Graduado Generalizado – GGUM da Teoria da Resposta ao Item

SILVANA LIGIA VINCENZI BORTOLOTTI

Professora da Universidade Tecnológica Federal do Paraná  
silvanaligia@gmail.com

DALTON FRANCISCO DE ANDRADE

Professor Phd. da Universidade Federal de Santa Catarina  
dandrade@inf.ufsc.br

## Resumo

A avaliação como instrumento de medida ocupa um importante lugar nas organizações atentas às exigências crescentes do mundo atual. A Teoria da Resposta ao Item (TRI) constitui-se em um conjunto de modelos para a avaliação que representam a relação entre a probabilidade de um indivíduo apresentar uma certa resposta a um item, as características dos itens e seu(s) traço(s) latente(s), definido(s) com base no estudo em questão. A finalidade deste artigo consiste em apresentar o Modelo de Desdobramento Graduado Generalizado – GGUM, da TRI, para medidas de satisfação. Este modelo foi desenvolvido para analisar respostas binárias e graduadas baseadas numa relação de proximidade. Os resultados obtidos na aplicação do modelo GGUM, que serviu para avaliar a satisfação dos alunos dos cursos de tecnologia da Universidade Tecnológica Federal do Paraná, campus Medianeira, ilustram o grande potencial desse modelo.

**Palavras-chave:** Teoria da Resposta ao Item, modelos de desdobramentos, grau de satisfação.

## Resumen

La evaluación como instrumento de medición ocupa un importante lugar en las organizaciones atentas a las crecientes exigencias del mundo actual. La Teoría de la Respuesta al Item (TRI) consiste en un conjunto de modelos para la evaluación que representan la relación entre la probabilidad de que un individuo presente una cierta respuesta a un item, las características de los items y su(s) trazo(s) latentes(s) definido(s) a partir del estudio en cuestión. La finalidad de este artículo consiste en presentar el GGUM – Modelo de Desdoblamiento Graduado Generalizado de la TRI, para medidas de satisfacción. Este modelo fue desarrollado para analizar respuestas binarias y graduadas basadas en una relación de proximidad. Los resultados obtenidos al aplicar el modelo GGUM (que sirvió para evaluar la satisfacción de los alumnos de los cursos de tecnología de la Universidad Tecnológica Federal de Paraná – campus Medianeira) ilustran su gran potencial.

**Palabras-clave:** Teoría de la Respuesta al Item, modelos de desdoblamiento, grado de satisfacción.

**Abstract**

Evaluation as a measurement instrument plays an important role in the organizations focused on the growing demands of the current world. Item Response Theory (IRT) provides models for evaluation. IRT is constituted of a set of models that represent the relationship between an individual's probability to give a certain response to an item, the characteristics of the items and their latent trait(s), as defined by the study. The purpose of this article is to introduce the GGUM – Generalized Graded Unfolding Model of IRT for satisfaction measurement. This model analyzes binary and graded responses based on a proximity relationship. The adequacy of the GGUM will be shown through an application to evaluate the satisfaction of undergraduate Technology students at the Federal Technological University of Paraná, at the Medianeira campus, Paraná, regarding the course they attend. The results obtained in the application illustrate the great potential of this model.

**Key-words:** Item Response Theory, unfolding models, satisfaction measurement.

## 1 INTRODUÇÃO

Em muitos estudos, algumas variáveis de interesse não podem ser medidas diretamente, tais como: proficiência em determinado conteúdo na avaliação educacional, grau de satisfação do consumidor em relação a determinado produto, nível de depressão. Essas variáveis são denominadas de variáveis latentes; são referidas por traço latente e devem ser inferidas com base na observação de variáveis secundárias que estejam relacionadas a ela. Geralmente se utilizam questionários ou testes com um conjunto de itens cujas respostas podem ou não obedecer a uma escala graduada, que são empregadas para desenvolver estimativas dessas variáveis, permitindo, assim, estimar traços latentes de indivíduos por meio de um conjunto de itens.

O problema consiste em como obter uma modelagem estatística para estimação do traço latente. Para essa finalidade, foram desenvolvidos vários modelos. A primeira iniciativa nesse sentido foi com a Teoria Clássica de Medidas (Lord, Novick, 1968), utilizando o total de pontos no teste. Embora a Teoria Clássica tenha sido muito útil, Hambleton e Swaminatan (1985) citam várias limitações que foram resolvidas por uma outra estrutura de teoria de medida, conhecida como Teoria de Resposta ao Item (Lord, 1980).

O objetivo deste trabalho consiste em apresentar o modelo de desdobramento da Teoria da Resposta ao Item, conhecido como Modelo de Desdobramento Graduado Generalizado – GGUM, bem como uma aplicação deste modelo para estudar a satisfação dos alunos quanto ao curso que freqüentam. A aplicação foi desenvolvida com alunos dos cursos de Tecnologia em Alimentos – modalidade Carnes e Laticínios, Tecnologia Ambiental, Tecnologia em Informática e Tecnologia em Eletromecânica da Universidade Tecnológica Federal do Paraná (UTFPR) – câmpus de Medianeira. A UTFPR é a primeira a receber essa denominação no Brasil e, por isso, tem uma história um pouco diferente das outras universidades. A instituição não foi criada e, sim, transformada a partir do Centro Federal de Educação Tecnológica do Paraná (Cefet-PR). Atualmente, a UTFPR tem como principais focos: a graduação, a pós-graduação e a extensão. Com ampla abrangência no Paraná, a UTFPR tem nove campi no Estado. O câmpus de Medianeira oferece o ensino médio, cursos técnicos e de tecnologia.

## 2 TEORIA DA RESPOSTA AO ITEM

A Teoria de Resposta ao Item fornece modelos para os traços latentes, propondo formas de representar a relação entre a probabilidade de um indivíduo dar uma certa resposta a um item, as características dos itens e seus traços latentes, na área de conhecimento a ser avaliada ou verificada, os quais não podem ser observados diretamente (Andrade, Tavares, Valle, 2000). Os diversos modelos existentes de respostas ao item se distinguem na forma matemática da função característica do item e/ou no número de parâmetros especificados no modelo, e detalhes podem ser vistos em Masters, Wright (apud Van Der Linden, Hambleton, 1997) e em Andrade, Tavares e Valle (2000). Todos os modelos podem conter um ou mais parâmetros relacionados aos itens e ao indivíduo e são tidos como modelos cumulativos (seção 2.1). Além desses modelos de natureza cumulativa, destacam-se também os modelos de desdobramentos (seção 2.2).

### 2.1 Modelos Cumulativos

Os modelos cumulativos da TRI são modelos em que a probabilidade de um indivíduo dar uma resposta correta ao item aumenta com o aumento do seu traço latente, isto é, níveis maiores de traço latente conduzem a valores mais altos de probabilidade de resposta correta. Dentre os modelos dicotômicos cumulativos, destacam-se o modelo logístico de 1 parâmetro (referido por alguns autores como modelo de Rasch), modelo logístico de 2 parâmetros e modelo logístico de 3 parâmetros. Dentre os modelos não dicotômicos, ressaltam-se o modelo de resposta nominal (Bock, 1972), o modelo de resposta gradual (Samejima, 1969) e o modelo de crédito parcial generalizado formulado por Muraki (1992). Para ilustrar um modelo cumulativo consideremos o modelo de Muraki cuja probabilidade de escolha da  $k$ -ésima categoria do item  $i$ , pelo indivíduo  $j$ , é dada por:

$$P_{i,k}(\theta_j) = \frac{\exp \left[ \sum_{u=0}^k Da_i(\theta_j - b_{i,u}) \right]}{\sum_{u=0}^{m_i} \exp \left[ \sum_{v=0}^k Da_i(\theta_j - b_{i,v}) \right]}, \quad i=1, 2, \dots, I; j=1, 2, \dots, n; k=0, 1, 2, \dots, m_i;$$

onde:

$m_i + 1$  é o número de categorias de resposta do item  $i$  (este número pode variar para cada item);

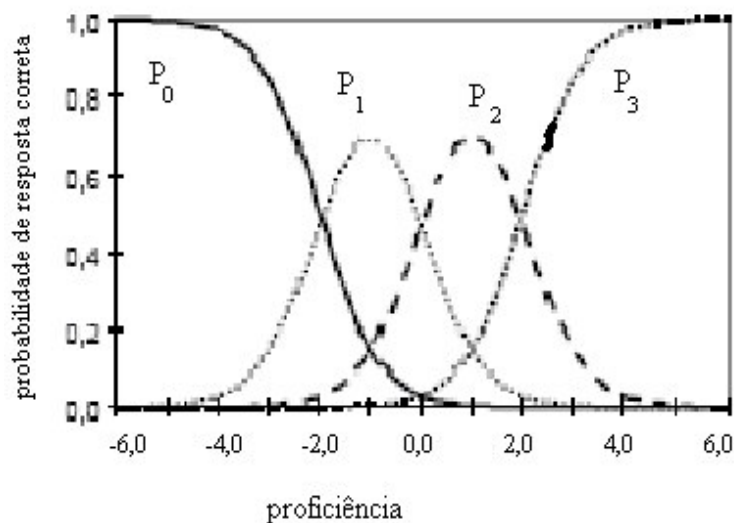
$a_i$  é o parâmetro de discriminação do item  $i$ ;

$D$  é um fator de escala constante, igual a 1 no modelo logístico;

$b_{i,k}$  é um ponto do *continuum* latente onde as curvas de  $P_{i,k-1}(\theta)$  e  $P_{i,k}(\theta)$  se interceptam.

Na Figura 1 temos a representação gráfica deste modelo para um item com 4 categorias de resposta medindo, por exemplo, proficiência em matemática, com os seguintes parâmetros:  $a = 1,0$ ,  $b_1 = -2,0$ ,  $b_2 = 0$  e  $b_3 = 2,0$ . Estes valores estão representados na escala (0,1) que será descrita mais adiante.

**Figura 1 – Representação gráfica de um item segundo o modelo de resposta gradual**



Fonte: Andrade, Tavares, Valle, 2000.

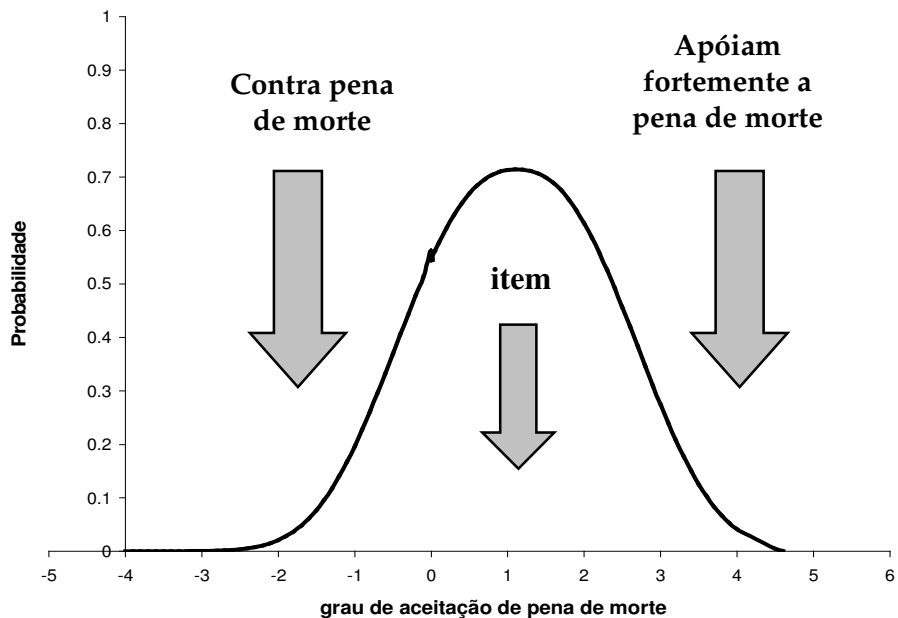
Segundo a figura, indivíduos com proficiência até -2,0 têm maior probabilidade de responder a categoria 0 (representada por  $P_0$ ), indivíduos com proficiência entre -2,0 e 0,0 têm mais chance de responder a categoria 1 ( $P_1$ ), enquanto indivíduos com proficiência entre 0,0 e 2,0 a maior probabilidade é que eles respondam a categoria 2 ( $P_2$ ). Enfim, indivíduos com proficiência acima de 2,0 têm maior probabilidade de responder à categoria 3 ( $P_3$ ) (que neste exemplo representa o acerto total).

De acordo com esse modelo de natureza cumulativa, observa-se que indivíduos com maior proficiência em matemática têm maior probabilidade de alcançar o acerto total representado pela categoria 3. Contudo, existem situações de respostas dicotômicas e não dicotômicas em que nem sempre a probabilidade de responder as categorias de resposta mais alta aumenta quando aumenta o traço latente indivíduo.

Por exemplo, considere o seguinte item extraído de um questionário aplicado por Roberts e Laughlin (1996), utilizado para medir o construto ou traço latente “aceitação de pena de morte”: **“A pena de morte é errada, porém é necessária em nossa civilização imperfeita”**, com as seguintes categorias de repostas: fortemente discordo ( $P_0$ ), discordo ( $P_1$ ), concordo ( $P_2$ ) e fortemente concordo ( $P_3$ ).

Neste item, as pessoas que têm sentimentos fortes contra pena de morte, ou baixa aceitação de pena de morte, escolheriam a categoria de resposta “fortemente discordo”, porque não concordariam com parte do item “porém é necessária”. Pessoas que têm sentimentos de nível médio tenderiam a concordar com este item, isto é, escolheriam a categoria “concordo”; entretanto, pessoas cujo grau de aceitação de pena de morte é alto optariam pela categoria de resposta “fortemente discordo”, pois não concordariam com parte do item “pena de morte é errada”. Nota-se que neste item níveis altos do construto “aceitação de pena de morte” não implicam categorias de respostas mais altas, como ocorre com os modelos cumulativos. Neste caso, o modelo cumulativo não seria adequado para a estimação do traço latente. A representação gráfica mais apropriada para a categoria de resposta “fortemente concordo”, em um modelo cumulativo, indicada na Figura 1 por  $P_3$ , seria a apresentada na Figura 2.

Figura 2 – Representação gráfica da probabilidade relacionada com a categoria de resposta “fortemente concordo”



Modelos com esta característica são denominados de modelos de desdobramentos, os quais serão apresentados a seguir.

## 2.2 Modelos de Desdobramentos

Os modelos de desdobramentos são modelos de probabilidade cujos parâmetros dos indivíduos e dos itens são colocados numa mesma escala, sendo: o parâmetro do indivíduo alocado de acordo com sua opinião, e os parâmetros dos itens, localizados nessa mesma escala, alocados segundo seu conteúdo. Esses modelos se distinguem dos modelos cumulativos por serem modelos de proximidade, em que categorias de resposta mais altas são mais prováveis (indicativo de níveis mais forte de concordância) quando a distância entre os parâmetros do indivíduo e o de posição do item na escala diminui (Coombs, 1964), ou seja, a probabilidade de um indivíduo dar uma resposta a um item está relacionada à distância entre os parâmetros do indivíduo e o de posição do item na escala, e não em função do parâmetro do indivíduo, como nos modelos cumulativos. O uso de

modelos de desdobramentos para uma medida de atitude se dá ao acreditar que um indivíduo concorda com uma categoria de resposta de um item, uma vez que o sentimento transmitido pela resposta ao item combina com sua própria opinião.

Foram desenvolvidos vários modelos de desdobramentos, tais como: o modelo Cosseno Hiperbólico (Andrich, Luo, 1993), o modelo Parella (Hojtink, 1990) e o Modelo de Desdobramento Graduado Generalizado, conhecido por GGUM, desenvolvido por Roberts, Donoghue e Laughlin (2000).

### 3 MODELO DE DESDOBRAMENTO GRADUADO GENERALIZADO – GGUM

O GGUM é um modelo de Teoria de Resposta ao Item unidimensional, desenvolvido para analisar tanto respostas binárias quanto graduadas baseadas numa relação de proximidade (Coombs, 1964, apud Roberts, Wedell, Laughlin, 1998). Ele tem como aplicação típica situações de medidas em que os entrevistados são solicitados a indicarem seu nível de concordância em relação a um conjunto de itens que se situam numa escala bipolar, ou seja, que variam o seu conteúdo, do negativo para o positivo, passando pelo neutro.

Esse modelo foi desenvolvido a partir de quatro proposições básicas sobre o processo de resposta. A primeira salienta que quando um indivíduo é solicitado a expressar a sua opinião de aceitação em uma declaração de atitude tende a concordar com o item à medida que ele se localiza próximo de sua posição pessoal em uma escala do traço latente.

Por exemplo, se  $\delta_i$  denotar a localização do item  $i$  nesta escala e  $\theta_j$  denotar a localização do indivíduo  $j$  na mesma escala, então o indivíduo tende mais a concordar com o item à proporção que a distância entre  $\theta_j$  e  $\delta_i$  se aproxima de zero; isso é uma característica fundamental de um processo de ponto ideal (Coombs, 1964). Para exemplificar, vamos considerar três itens de um estudo sobre atitude em relação ao grau de aceitação do aborto, e seus respectivos parâmetros de localização  $\delta_i$ , representados em uma escala do traço latente com média 0 e desvio padrão 1 (detalhes em Roberts, Donoghue, Laughlin, 2000):



- a) “Aborto é a destruição de uma vida pela conveniência de outro”:  $\delta_1 = -3,2$ .
- b) “Há algumas situações claras onde aborto deveria ser legalizado, mas não deveria ser permitido em todas as situações”:  $\delta_2 = -0,6$ .
- c) “Uma mulher deveria deter o direito para escolher um aborto baseado nas próprias circunstâncias de vida dela”:  $\delta_3 = 2,5$ .

Então um indivíduo com  $\theta = -3,0$  tem alta probabilidade de concordar com o item a) e discordar dos itens b) e c), pois,  $|\theta_j - \delta_1| = |-3,0 - (-3,2)| = 0,2$  para o item a),  $|\theta_j - \delta_2| = |-3,0 - (-0,6)| = 2,4$  para o item b) e  $|\theta_j - \delta_3| = |-3,0 - (2,5)| = 5,5$  para o item c), a distância do parâmetro do item em relação ao parâmetro do indivíduo é menor em relação ao item a) e maior para os outros dois itens. Um indivíduo com  $\theta = 3,0$  irá discordar dos itens a) e b) e, com alta probabilidade, irá concordar com o item c), em razão da proximidade com esse item ( $\delta_3 = 2,5$ ).

A segunda proposição do modelo destaca que um indivíduo pode responder a uma determinada categoria de resposta por dois motivos distintos. Suponha que um indivíduo é solicitado a responder um questionário que pretende avaliar seu traço latente, como, por exemplo, o seu grau de satisfação. Esse questionário é formado por um conjunto de itens que, 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. Segundo Roberts, Lin e Laughlin (2001) um indivíduo pode discordar de um item porque seu conteúdo é tão negativo, se o seu parâmetro  $\delta$  é muito menor do que o parâmetro  $\theta$  do indivíduo, ou tão positivo, se o seu parâmetro  $\delta$  é muito maior do que o parâmetro  $\theta$  do indivíduo. Assim, se o conteúdo do item é tão negativo, o indivíduo “discorda acima do item”. Contudo, se o conteúdo do item for tão positivo, o indivíduo “discorda abaixo do item”. Em ambos os casos, essas categorias de respostas são denominadas de categorias de respostas subjetivas que o indivíduo pode usar. Na escala do traço latente, “acima” seria à “direita” do item, e “abaixo”, “à esquerda”.

A terceira proposição é que as respostas subjetivas dadas aos itens seguem um modelo de resposta ao item cumulativo (Andrich, Luo, 1993); mais especificamente, o GGUM utiliza o modelo de Muraki (1992) em razão da sua generalidade. O modelo de Crédito Parcial Generalizado proposto por esse autor, quando aplicado às respostas subjetivas, é dado por:

$$P[Y_i = y | \theta_j] = \frac{\exp \left[ \alpha_i \left( y(\theta_j - \delta_i) - \sum_{k=0}^y \tau_{ik} \right) \right]}{\sum_{v=0}^M \exp \left[ \alpha_i \left( v(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik} \right) \right]}, \quad (3.1)$$

com a seguinte restrição:

$$\sum_{k=0}^M \tau_{ik} = 0, \quad (3.2)$$

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;

$\theta_j$  = parâmetro de locação do indivíduo  $j$  na escala do traço latente;

$\delta_i$  = parâmetro de locação do item  $i$  na escala do traço latente;

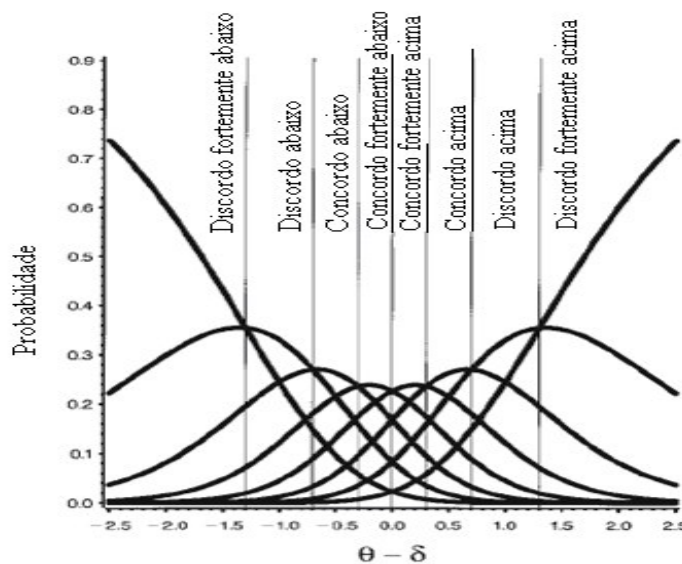
$\alpha_i$  = parâmetro de discriminação do item  $i$ ;

$\tau_{ik}$  = parâmetro de locação do limiar de categoria de resposta subjetiva  $k$  na escala do traço latente relativa à posição do item  $i$ ;

$M$  é o número das categorias de respostas subjetivas menos 1.

A Figura 3 ilustra o modelo dado em (3.1), apresentando as funções de probabilidade de um item com quatro categorias de respostas observáveis: “discordo fortemente”, “discordo”, “concordo”, “concordo fortemente”.

Figura 3 – Função de probabilidade de um item com quatro categorias de resposta em razão 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: Roberts, Donoghue e Laughlin, 2000 (modificada).

Observa-se na abscissa da Figura 3 uma escala do traço latente dada pela diferença entre o parâmetro do indivíduo e a posição do parâmetro do item ( $\theta - \delta$ ). O eixo das ordenadas representa as probabilidades de uma resposta do indivíduo cair em uma das oito categorias de respostas subjetivas. Há oito categorias de respostas subjetivas e oito funções de probabilidades associadas às respostas, pois um indivíduo pode responder quaisquer das 4 categorias de resposta observável, porque sua posição na escala do traço latente está ora abaixo, ora acima da localização 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 na escala. Logo, esses limiares dividem a escala em 8 intervalos, na qual 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  $\alpha$  (Roberts, Donoghue, Laughlin, 2000).

Como o modelo de Muraki (1992) define um modelo de resposta ao item para níveis de respostas subjetivas, e cada categoria de resposta observável está associada com duas possibilidades de resposta subjetiva (isto é, uma abaixo do item e outra acima), e as duas categorias de respostas subjetivas correspondentes a uma dada categoria de resposta observável são mutuamente exclusivas, então a probabilidade de um indivíduo responder uma categoria observável particular é simplesmente a soma das probabilidades associadas com as duas respostas subjetivas correspondentes:

$$P(Z_i = z | \theta_j) = P(Y_i = z | \theta_j) + P(Y_i = (M - z) | \theta_j), \quad (3.3)$$

onde:

$Z_i$  = uma resposta observável à declaração de atitude  $i$ ;  
 $z = 0, 1, 2, 3, \dots, H$ ;  $z = 0$  corresponde ao nível de discordância mais forte e  $z = H$  corresponde ao nível de concordância mais forte;  
 $H$  = número de categorias de respostas observáveis menos 1.  $M = 2H + 1$ .

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

$$\tau_{i(H+1)} = 0, \text{ e} \quad (3.4)$$

$$\tau_{iz} = -\tau_{i(M-z+1)}, \text{ para } z \neq 0. \quad (3.5)$$

Essa proposição implica que os indivíduos estão propensos a concordar com um item localizado tanto em  $-p$  unidades ou  $+p$  unidades da posição do indivíduo na escala do traço latente. Essa proposição conduz à identidade:

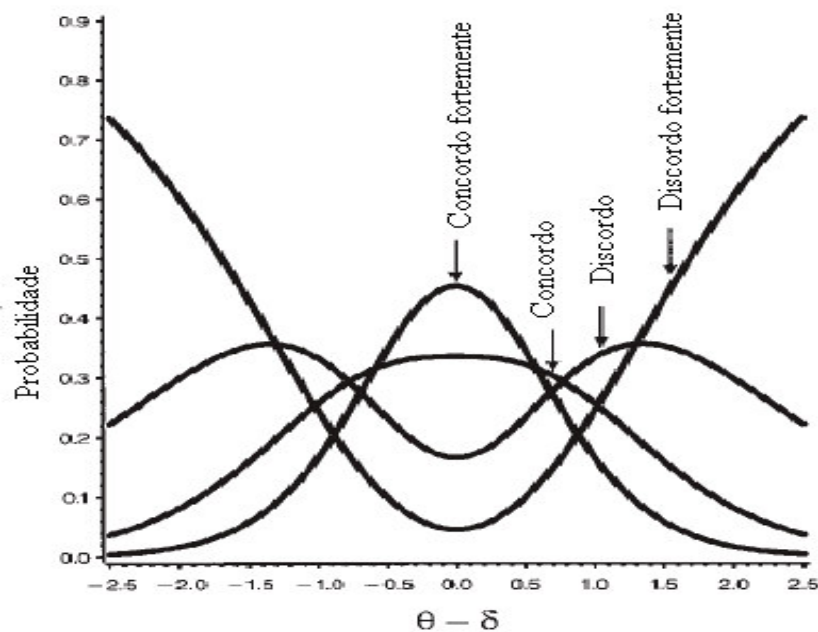
$$\sum_{k=0}^z \tau_{ik} = \sum_{k=0}^{M-z} \tau_{ik}. \quad (3.6)$$

Incorporando esta identidade em (3.3) resulta na função de probabilidade do GGUM (Roberts, Wedell, Laughlin, 1998):

$$P(Z=z|\theta_j) = \frac{\exp\left[\alpha_i\left(z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}\right)\right] + \exp\left[\alpha_i\left((M-z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}\right)\right]}{\sum_{v=0}^H \left[ \exp\left[\alpha_i\left(v(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik}\right)\right] + \exp\left[\alpha_i\left((M-v)(\theta_j - \delta_i) - \sum_{k=0}^v \tau_{ik}\right)\right] \right]} \quad (3.7)$$

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 4 apresenta as funções de probabilidade de respostas observáveis das categorias de respostas para o mesmo item referenciado na Figura 3. Cada uma dessas funções de probabilidade é a soma de duas funções de probabilidades das respostas subjetivas correspondentes, mostradas na Figura 3.

**Figura 4 – Função de probabilidade de um item com quatro categorias de respostas observáveis em razão de  $\theta - \delta$**

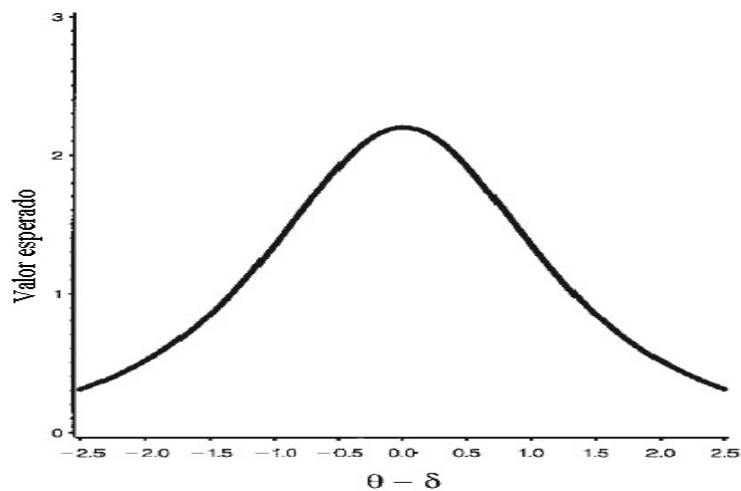


Fonte: Roberts, Donoghue e Laughlin, 2000 (modificada).

Observe que as funções de probabilidades das respostas observáveis não se interceptam em  $\tau_{i1}, \tau_{i2}, \tau_{i3}, \dots, \tau_{iH}$ . Assim, os parâmetros  $\tau_{ik}$  perdem sua simples interpretação ao nível de resposta observável. Da mesma maneira, o parâmetro  $\alpha_i$  indexa a discriminação para um nível de resposta subjetiva. Um outro ponto com respeito aos parâmetros  $\tau_{ik}$  é que eles não precisam ser sucessivamente ordenados na escala do traço latente. Esses valores simplesmente indicam onde as funções de probabilidades de categoria de respostas subjetivas sucessivas se interceptam. 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, Donoghue, Laughlin, 2000).

O GGUM é um modelo de desdobramento de processo de resposta. Isso pode ser visto calculando-se o valor esperado de uma resposta observável para vários valores de  $\theta - \delta$ , usando a função probabilidade mostrada em (3.7). A Figura 5 apresenta o valor esperado de uma resposta observável para o item representado na Figura 3 com quatro categorias de resposta. As categorias são codificadas com os valores inteiros de 0 a 3 onde cada número corresponde às respostas “discordo fortemente”, “discordo”, “concordo”, “concordo fortemente”, respectivamente.

**Figura 5 – Valor esperado de uma resposta observável de um item com quatro categorias de resposta em função de  $\theta - \delta$**



Fonte: Roberts, Donoghue e Laughlin, 2000 (modificada).

A Figura 5 mostra que o item apresenta níveis maiores de concordância à medida que a distância entre o indivíduo e o item na escala diminui (Roberts, Donoghue, Laughlin, 2000). A função do valor esperado do GGUM tem natureza de pico simples, e valores esperados mais altos ocorrem quando  $|\theta - \delta|$  se aproxima de zero.

Uma aplicação no Brasil deste modelo foi feito por Samartini (2006) para modelar a importância de atributos de serviços ou produtos.

Na seção seguinte será apresentada a estimação dos parâmetros, bem como o pacote computacional GGUM2004<sup>1</sup>.

#### 4 ESTIMAÇÃO E O PROGRAMA GGUM2004

Os parâmetros dos itens do GGUM são usualmente estimados pelo método da máxima verossimilhança marginal (MVM) (Bock, Lieberman, 1970; Muraki, 1992, *apud* Roberts, Donoghue e Laughlin, 2000). O algoritmo proposto faz um paralelo com o procedimento de Muraki (1992) baseado no algoritmo EM (Esperança e Maximização) e *scoring* de Fischer. A estimação dos parâmetros dos itens é comumente referenciada como calibração. No processo de estimação a suposição de unidimensionalidade é primordial, uma vez que sua violação pode invalidar as estimativas dos parâmetros do modelo. Essa suposição é verificada por meio da Análise Fatorial, utilizando, por exemplo, o método dos componentes principais (Escofier, Pagès, 1992).

Para a estimação dos parâmetros dos indivíduos utilizam-se as estimativas dos parâmetros dos itens, como se fossem seus verdadeiros valores, em conjunto com as respostas observadas. As estimativas dos parâmetros dos indivíduos são obtidas pelo método bayesiano de estimação EAP. Esses métodos estão implementados no pacote computacional GGUM2004. Para a estimação dos parâmetros dos itens assume-se uma distribuição de probabilidade *a priori* para  $\theta$  dos respondentes. Esse programa utiliza como *priori* a distribuição normal padrão com média 0 e desvio padrão 1, isto é,  $N(0; 1)$ , e denomina de ciclos internos a execução dos passos de maximização, enquanto o estágio da esperança é denominado de ciclos externos.

---

<sup>1</sup> Disponível para download em: <[www.psychology.gatech.unfolding/FreeSoftware.html](http://www.psychology.gatech.unfolding/FreeSoftware.html)>.

As equações de verossimilhança envolvem integrações que são solucionadas pela utilização de pontos de quadratura discretos. Existem outros programas computacionais desenvolvidos para ajuste de modelos de desdobramentos, como, por exemplo, RATEFOLD para o Modelo de Cosseno Hiperbólico e Modelo Parella e o MUDFOLD para Desdobramento Unidimensional Múltiplo.

No GGUM2004 é o usuário que especifica o número de pontos de quadratura. Esse número pode chegar no máximo até 50 pontos de quadratura e são igualmente espaçados num intervalo de  $-4,0$  a  $+4,0$ . O Programa apresenta algumas limitações: ele suporta, no máximo, 2000 indivíduos respondentes, e 100 itens com, no máximo, 10 categorias de resposta por item.

Na seção seguinte será apresentada uma aplicação do modelo GGUM, destacando como foi feita a coleta de dados, o desenvolvimento do banco de itens, a verificação da dimensionalidade, a seleção dos itens, as estimativas dos parâmetros dos itens, as estimativas dos parâmetros dos indivíduos, o ajuste do modelo e a interpretação dos parâmetros dos itens.

## 5 APLICAÇÃO

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

### 5.1 Coleta de Dados

Para a coleta de dados, utilizou-se uma forma adaptada de um questionário de avaliação de ensino da Unioeste – Cascavel – PR<sup>2</sup>, obedecendo aos objetivos propostos. Ele foi aplicado a 525 alunos dos cursos de Tecnologia do Cefet – MD – PR, contendo 35 itens, com seis categorias de resposta cada um: totalmente insatisfeito ( $P_0$ ), insatisfeito ( $P_1$ ), pouco insatisfeito ( $P_2$ ), pouco satisfeito ( $P_3$ ), satisfeito ( $P_4$ ), totalmente satisfeito ( $P_5$ ).

---

<sup>2</sup> Disponível em: <[www.md.cefetpr.br/silvana](http://www.md.cefetpr.br/silvana)>.



## 5.2 Desenvolvimento do Banco de Itens

### 5.2.1 Dimensionalidade

A Tabela 1 mostra o resultado da extração dos autovalores da análise fatorial pelo método dos componentes principais. O item 30 foi eliminado antes de ser feita a análise, pois alguns alunos não responderam a este item, e o programa utilizado não permite trabalhar com respostas faltantes.

**Tabela 1 – Extração dos autovalores por meio dos componentes principais**

Fator	Autovalor	Variância Total	Autovalor Cumul.	Cumul. %
1	11,295	33,222	11,295	33,217
2	2,060	6,058	13,355	39,279

Segundo Reckase (1979), os resultados podem indicar uma unidimensionalidade se o primeiro fator da conta for de pelo menos 20% da variância total. Como se observa na Tabela 1, o primeiro fator explica 33% da variância, satisfazendo este critério; logo pode-se 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 2, observa-se o resultado das cargas fatoriais dos dois primeiros componentes e a comunalidade de cada item. A análise fatorial foi realizada com o método dos componentes principais extraindo-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 comuns com outros itens. Nota-se que nesta análise com todos os itens (menos o item 30), sete itens apresentam comunalidade menor que 0,3, por isso foram removidos da análise. Os dados foram analisados usando o programa computacional SPSS® versão 10.

**Tabela 2 – Matriz dos dois primeiros fatores e comunalidade dos itens**

<b>Item</b>	<b>Fator 1</b>	<b>Fator 2</b>	<b>Comunalidade</b>
1	-0,529	-0,059	<b>0,284</b>
2	-0,518	-0,134	<b>0,289</b>
3	-0,612	-0,395	0,531
4	-0,577	-0,406	0,499
5	-0,546	-0,340	0,413
6	-0,517	-0,215	0,314
7	-0,475	0,085	<b>0,233</b>
8	-0,616	-0,281	0,457
9	-0,475	0,071	<b>0,230</b>
10	-0,592	-0,146	0,371
11	-0,588	0,022	0,346
12	-0,508	0,324	0,363
13	-0,558	0,363	0,444
14	-0,608	0,390	0,522
15	-0,650	0,420	0,600
16	-0,522	0,386	0,422
17	-0,609	0,190	0,407
18	-0,548	0,192	0,337
19	-0,442	0,087	<b>0,203</b>
20	-0,664	-0,041	0,442
21	-0,671	-0,036	0,452
22	-0,700	-0,122	0,505
23	-0,681	-0,086	0,471
24	-0,727	0,044	0,531
25	-0,592	0,195	0,388
26	-0,505	0,293	0,341
27	-0,601	0,344	0,479
28	-0,660	0,046	0,437
29	-0,689	-0,139	0,493
31	-0,558	-0,047	0,314
32	-0,240	-0,219	<b>0,105</b>
33	-0,361	-0,369	<b>0,266</b>
34	-0,514	-0,354	0,390
35	-0,657	-0,212	0,476

### 5.2.2 Seleção de itens para a escala final

Foram realizadas várias calibrações, utilizando a máxima verossimilhança marginal com uma distribuição a priori  $N(0;1)$  para os  $\theta_j$ . Foram usados 50 pontos de quadratura igualmente espaçados entre  $-4,0$  e  $+4,0$ .

Para o ajuste dos itens foram utilizadas as estatísticas denominadas *infit* (média dos resíduos quadráticos ponderados pela variância do resíduo para o indivíduo) e *outfit* (média dos resíduos quadráticos não ponderados) (Linacre, Wright, 1994; Masters, Wright, 1997; Wright, Masters, 1982). O valor esperado dessas estatísticas é 1. Valores maiores que 1,3 (Karabatsos, 2000; Smith, Schumacker, Bush, 1998) ou 1,5 (Engelhard, 1992; Lunz, Wright, Linacre, 1990) são utilizados para eliminar itens com ajuste ruim. Nas calibrações realizadas não foi observado nenhum item com *infit* ou *outfit* superior a 1,3.

Utilizou-se também o *log* da razão do valor observado sobre o previsto denominado de estatística  $G^2$  para cada item e para o modelo como um todo. Essa estatística avalia o ajuste do modelo para as frequências observadas e esperadas para cada grupo de respondentes que é especificado pelo usuário no GGUM2004. Para ambas as estatísticas considerou-se a hipótese nula: os dados se ajustam ao modelo, ou seja, o valor previsto e o observado são próximos.

Na primeira calibração observou-se que 9 estimativas dos parâmetros  $\theta$  apresentaram o erro padrão da estimativa muito maior do que o erro padrão das demais estimativas, assim optou-se por eliminar os alunos que geraram essas estimativas; um item foi eliminado porque apresentou para a estatística  $G^2$  o p-valor  $< 0,0001$ , abaixo de 0,05. Os valores das estatísticas *infit* e *outfit*, para os itens, não foram significantes (p-valor  $> 0,01$ ) nas calibrações realizadas. Dos itens que permaneceram na escala final nenhum deles apresentou falta de ajuste, de acordo com as estatísticas utilizadas; portanto, esses 26 itens foram mantidos na calibração.

### 5.2.3 Estimativas dos parâmetros dos itens

A Tabela 3 apresenta as estimativas dos parâmetros dos itens. Os itens estão listados de acordo com a ordem crescente do parâmetro  $\delta_i$ . Os 26 itens que permaneceram na escala final tiveram conteúdo positivo estimado, variando de moderado a extremo; não foi verificado nenhum item de conteúdo negativo ou neutro. As estimativas dos parâmetros de discriminação  $\alpha_i$  apresentaram valores que variaram de 0,658 a 1,970.

Tabela 3 – Estimativas dos parâmetros dos 26 itens ( $\hat{\delta}_i, \hat{\alpha}_i, \hat{\tau}_{ik}$ )

Item	$\hat{\delta}_i$	$\hat{\alpha}_i$	$\hat{\tau}_{i1}$	$\hat{\tau}_{i2}$	$\hat{\tau}_{i3}$	$\hat{\tau}_{i4}$	$\hat{\tau}_{i5}$
34	1,148	0,956	-4,357	-2,866	-3,038	-2,103	0,544
4	1,171	1,267	-4,796	-3,184	-2,678	-1,496	1,112
5	1,249	1,071	-4,129	-3,462	-2,774	-1,372	1,379
6	1,255	0,816	-3,798	-2,540	-3,147	-1,795	1,762
23	1,283	1,732	-3,105	-2,577	-2,089	-0,933	1,442
3	1,310	1,318	-4,692	-2,868	-2,614	-1,341	1,105
22	1,318	1,951	-2,962	-2,560	-1,882	-0,883	0,871
21	1,372	1,469	-2,837	-2,441	-1,728	-0,710	1,000
35	1,385	1,246	-3,199	-2,708	-2,287	-1,138	0,778
29	1,419	1,523	-3,517	-2,528	-2,374	-1,448	0,570
20	1,455	1,175	-3,318	-2,199	-2,140	-0,816	1,229
8	1,502	1,183	-3,830	-2,957	-2,803	-1,349	0,925
31	1,610	0,776	-3,790	-2,835	-3,404	-2,476	-0,183
24	1,652	1,970	-4,095	-3,204	-2,525	-1,465	0,548
11	1,682	0,945	-4,146	-3,564	-3,012	-2,365	0,870
10	1,709	0,984	-4,390	-3,537	-3,173	-2,597	0,260
17	1,930	0,946	-3,548	-3,244	-2,732	-1,487	0,777
28	1,931	1,188	-3,450	-3,153	-2,806	-1,815	0,251
25	2,071	0,943	-4,402	-3,926	-2,858	-1,993	0,946
27	2,074	0,987	-4,508	-3,503	-2,907	-2,151	1,018
18	2,214	0,658	-4,188	-3,704	-2,935	-2,764	0,464
13	2,407	0,745	-4,355	-3,753	-3,141	-1,751	2,049
12	3,082	0,699	-6,341	-4,735	-4,350	-3,483	0,162
16	3,153	0,756	-5,874	-4,706	-4,275	-3,066	0,216
15	3,420	1,333	-4,957	-4,638	-4,279	-2,878	0,189
14	3,430	1,042	-5,852	-4,798	-4,133	-2,846	0,166

### 5.3 Estimativas do Parâmetro $\theta_j$

Foram utilizadas as estimativas dos parâmetros dos itens juntamente com as respostas observadas, para obter as estimativas dos parâmetros dos alunos. A média das estimativas foi -0,025 com um desvio padrão de 0,937. Os valores mínimo e máximo foram respectivamente -2,314 e 3,915. A distribuição dessas estimativas está apresentada na Figura 9 (seção 5.5).

### 5.4 Teste de Ajuste do Modelo

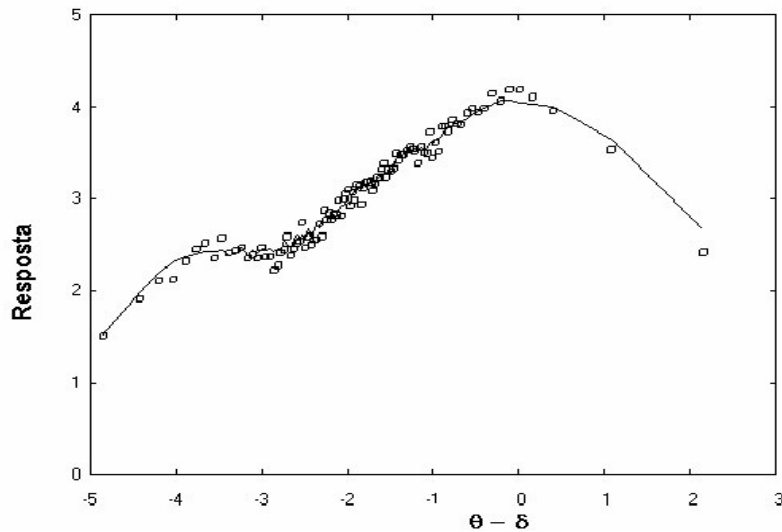
O teste de ajuste do modelo foi feito de forma heurística, utilizando as estatísticas citadas em 5.2.2 e também por meio de uma avaliação gráfica do ajuste global do GGUM, para verificar se os pontos das respostas observadas e esperadas obtidas pelo modelo coincidem; se ocorrer grande discrepância entre os pontos, significa que o modelo não se ajusta adequadamente aos dados. Este gráfico é feito colocando as respostas observadas e esperadas no eixo  $y$  e no eixo  $x$  a distância  $\hat{\theta}_j - \hat{\delta}_i$  (Hambleton, Swaminathan, 1985).

Para construir este gráfico calcularam-se  $\hat{\theta}_j - \hat{\delta}_i$  para todo par indivíduo/item ( $516 \times 26 = 13.416$  pares), e essas distâncias foram ordenadas e divididas em aproximadamente 100 grupos homogêneos (o número de grupos foi especificado no GGUM2004 e corresponde ao número de pontos do gráfico) de mesmo tamanho (o tamanho de cada grupo é estabelecido pelo *software*). Calculou-se, para cada par, a resposta observada do aluno, isto é,  $Z_{ij}$  (resposta observada do indivíduo  $j$  para o item  $i$ ) e o valor esperado  $E_{ij}$  (resposta esperada do indivíduo  $j$  para o item  $i$ ), dado por:

$$E_{ij} = \sum_{z=0}^H z(P(Z_i = z | \hat{\theta}_j))$$

dentro de cada grupo. Em seguida, calculou-se a média de resposta observada e esperada e o valor médio de  $\hat{\theta}_j - \hat{\delta}_i$ , de cada grupo. A Figura 6 mostra a relação entre a resposta observada média e a esperada média em função do valor médio de  $\hat{\theta}_j - \hat{\delta}_i$  para cada um dos 100 grupos. Os quadrados vazios no gráfico representam a resposta observada do item para cada grupo, e a linha sólida representa a resposta esperada do item.

Figura 6 – Resposta observada média *versus* resposta esperada média



O GGUM localiza tanto os parâmetros dos itens como os parâmetros dos alunos numa mesma escala do traço latente  $\theta$  (grau de satisfação). Nota-se, na Figura 6, que quando  $|\hat{\theta}_j - \hat{\delta}_i|$  se torna menor, o grau de satisfação aumenta na resposta; observa-se também que o ajuste do GGUM parece ser razoável, pois a discrepância entre o valor esperado médio e observado médio é mínima. A estatística *infit* variou de 0,96 a 1,01 e a *outfit* variou de 0,92 a 0,99; esses valores são aceitáveis, não sugerindo itens com falta de ajuste.

### 5.5 Interpretação dos Parâmetros dos Itens

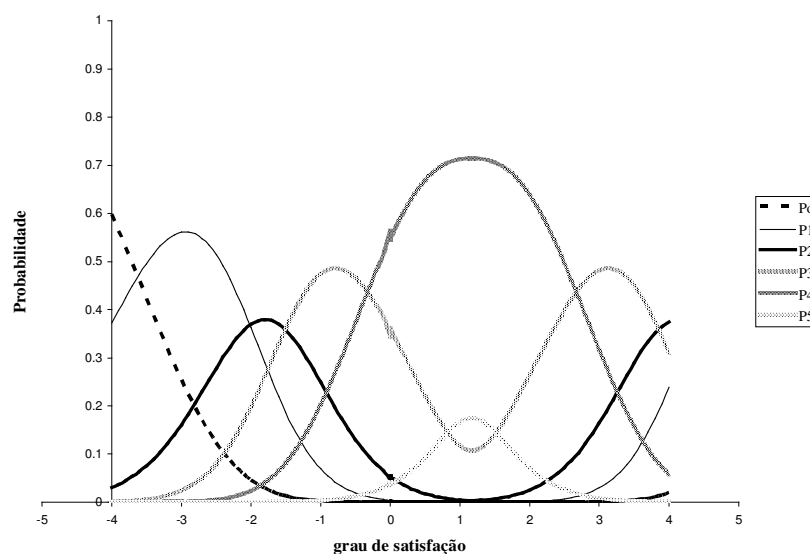
Conforme foi descrito na seção 3,  $\alpha_i$  é o parâmetro de discriminação do item  $i$ ;  $\delta_i$  o parâmetro de locação do item  $i$ ;  $\theta_j$  o parâmetro do aluno  $j$ , neste caso, a satisfação do aluno; e  $\tau_{ik}$  o parâmetro de locação dos limiares da categoria de respostas subjetivas na escala do traço latente relativa à posição do item  $i$ . Quando  $\theta_j - \delta_i$  se aproxima de zero, o indivíduo tende a estar mais satisfeito com o item. De posse dos valores dos parâmetros podem ser feitas algumas considerações. Por exemplo:

**Item 4:** As características e os resultados obtidos neste item são as seguintes:

$$\hat{\delta}_4 = 1.171; \hat{\alpha}_4 = 1.267; \hat{\tau}_{4k} = (-4.796; -3.184; -2.678; -1.496; 1.112).$$

A Figura 7 ilustra a função de probabilidade das categorias de resposta do item 4 em relação ao grau de satisfação.

**Figura 7 – Função de probabilidade das categorias de resposta observável do item 4 em relação ao grau de satisfação**



Na Figura 7, observa-se que na função de probabilidade da categoria 4, quando o grau de satisfação se aproxima de 1,171, o valor  $\theta - \delta_4$  se aproxima de 0, e neste ponto a função de probabilidade é maior em relação às outras categorias de respostas.

Para uma satisfação  $\theta = 1,2$ ,  $\theta - \delta_4 = 0,029$  (perto de zero) têm-se:

Categoria de resposta (z)	Probabilidade $P(Z = z   \theta_j)$
0	< 0,001
1	< 0,001
2	0,004
3	0,108
4	0,714
5	0,174

Dessa forma, se forem reunidos alunos com uma satisfação igual a 1,2, 71,4% responderiam que estão satisfeitos com o item 4, 17,4% totalmente satisfeitos com esse item, 10,8% pouco satisfeitos, 0,4% pouco insatisfeitos e aproximadamente nenhum dos alunos responderia estar totalmente insatisfeito ou insatisfeito.

Observa-se que para um  $\theta = -2,4$ , isto é, para  $\theta - \delta_4$  extremo igual a  $-3,571$  têm-se:

Categoria de resposta (z)	Probabilidade $P(Z = z   \theta)$
0	0,104
1	0,491
2	0,301
3	0,097
4	0,007
5	< 0,001

Neste caso, alunos que têm  $\theta$  (satisfação) =  $-2,4$  a probabilidade é maior para a categoria 1 – 49% responderiam que estão insatisfeitos com o item.

Para  $\theta = 4,0$ ,  $\theta - \delta_4 = 2,829$ , a probabilidade para cada categoria de resposta é:

Categoria de resposta (z)	Probabilidade $P(Z = z   \theta)$
0	0,020
1	0,240
2	0,374
3	0,310
4	0,060
5	< 0,001



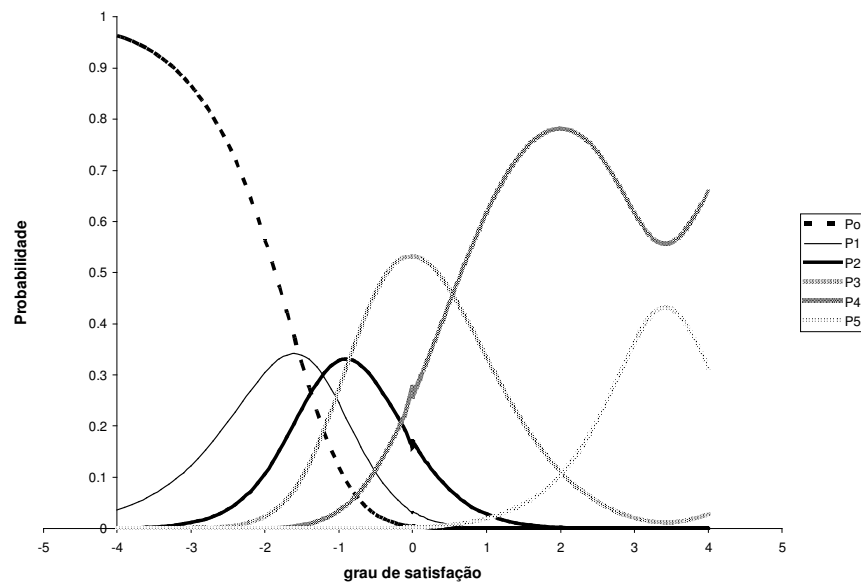
Já neste caso, alunos com satisfação igual a 4,0, a maior probabilidade ocorre para a categoria 2, isto é, a probabilidade de alunos que apresentam esse parâmetro e que responderiam que estão pouco insatisfeitos é de 37,4%. 31% dos alunos responderiam que estão pouco satisfeitos.

**Item 15** – As características e os resultados obtidos neste item são as seguintes:

$$\hat{\delta}_{15} = 3,420; \hat{\alpha}_{15} = 1,333; \hat{\tau}_{15k} = (-4,957; -4,638; -4,279; -2,878; 0,189).$$

A Figura 8 ilustra a função de probabilidade das categorias de resposta do item 15 dado o grau de satisfação.

**Figura 8 – Função de probabilidade das categorias de resposta observável do item 15 em relação ao grau de satisfação**



Nesse item para  $\theta$  (satisfação) = 3,4, ( $\theta - \delta_{14} = -0,02$  perto do zero) têm-se as seguintes probabilidades:

Categoria de resposta (z)	Probabilidade P(Z=z   $\theta$ )
0	< 0,001
1	< 0,001
2	< 0,001
3	0,012
4	0,560
5	0,432

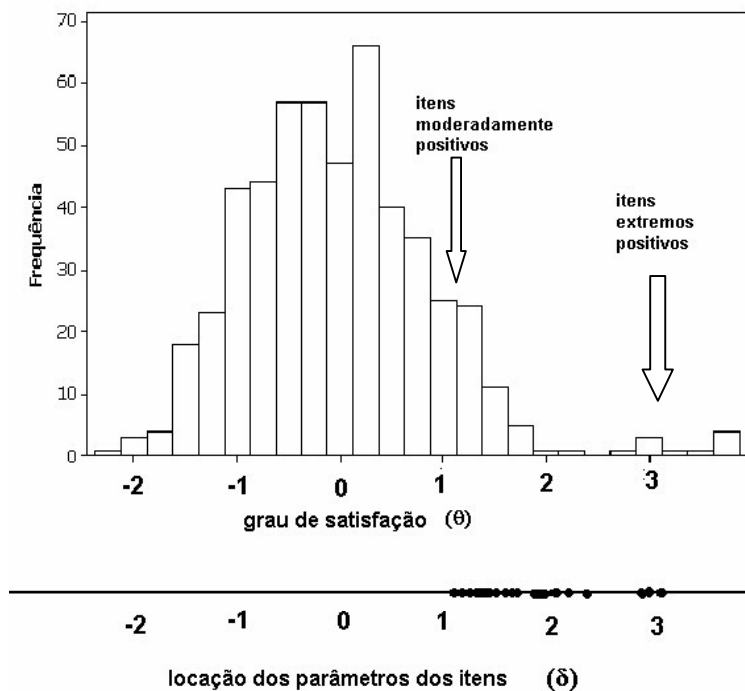
Desse modo, se forem reunidos alunos com grau de satisfação 3,4, 56% responderiam que estão satisfeitos com o item 15; 43,2% estão totalmente satisfeitos com este item. Nota-se que para uma satisfação  $\theta = -3,0$ , ( $\theta - \delta_{14} = -6,42$ ) a probabilidade para cada categoria de resposta é:

Categoria de resposta (z)	Probabilidade (P(Z=z   $\theta$ ))
0	0,865
1	0,123
2	0,011
3	< 0,001
4	< 0,001
5	< 0,001

Neste caso, alunos que têm  $\theta$  (grau de satisfação) = -3,0 a probabilidade é maior para as categorias 0 e 1, visto que 86,5% dos alunos responderiam que estão totalmente insatisfeitos, e 12,3% estão insatisfeitos.

A Figura 9, a seguir, ilustra a distribuição de frequência dos parâmetros  $\theta$  (grau de satisfação) obtidos dos alunos e os parâmetros dos itens colocados na mesma escala.

Figura 9 – Distribuição dos parâmetros dos alunos e dos itens



Visualiza-se na Figura 9 que todos os itens tiveram seus conteúdos positivos bem acima da maioria dos parâmetros dos alunos, indicando que a maioria discorda abaixo.

No modelo GGUM, os níveis altos de concordância ocorrem quando o parâmetro do aluno e a posição do item, representados na mesma escala latente, estão próximos. E, como os 26 itens restantes, calibrados segundo este modelo, foram colocados nesta escala 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 nesta mesma escala numa posição mais negativa em relação à posição dos itens, evidenciando certa distância em relação à posição do indivíduo e do item na escala latente, então tal distribuição indica que os itens avaliados não foram na maioria aceitos pelos alunos.

Uma vez que os itens são pressupostos do indicativo do grau de satisfação dos alunos com relação ao curso, pode-se deduzir e verificar, pelo modelo GGUM, que, de forma geral, não estão satisfeitos com o curso.

## CONCLUSÃO

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 o 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. No ajuste do modelo os itens apresentaram as estatísticas *infit* e *outfit* próximas de 1.

Dos 35 itens do questionário, 26 deles foram mantidos na escala final. Os itens 1, 2, 7, 9, 19, 32 e 33 foram eliminados, pois apresentaram uma comunalidade menor que 0,3; o item 30 foi retirado antes das calibrações, visto que alguns alunos deixaram de respondê-lo; e o item 26 também foi retirado, pois na primeira calibração teve o p-valor da estatística  $G^2$  baixo. Nove alunos também foram retirados da análise por apresentarem um valor elevado no erro padrão da estimativa dos parâmetros  $\theta_j$ .

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 e de avaliação.

O GGUM se ajusta adequada e corretamente aos dados; as estimativas dos itens são invariantes aos indivíduos; e as estimativas dos indivíduos são invariantes aos itens; essas características 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 modelo GGUM da TRI oferece grande contribuição, principalmente porque coloca os itens e os indivíduos numa mesma escala de medida. Os itens são colocados de acordo com seu conteúdo, e os indivíduos segundo seu traço latente, ambos variando de negativo, neutro a positivo.

## REFERÊNCIAS BIBLIOGRÁFICAS

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

ANDRADE, D. F.; VALLE, R. C. Introdução à teoria da resposta ao item: conceitos e aplicações. *Estudos em Avaliação Educacional*, n. 18, p.3-32, 1998.

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

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

BOCK, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, n. 37, p. 29-51, 1972.

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

BORTOLOTTI, S. L.V. *Aplicação de um modelo de desdobramento graduado generalizado da teoria da resposta ao item – TRI*. Florianópolis, 2003. Dissertação (mestr.) Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia de Produção. 107 p.

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

DEMARS, C. E. Type I error rates for the generalized graded unfolding model fit indices. *Applied Psychological Measurement*, n. 28, p. 48-71, 2004.

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

ENGELHARD JR., G. The measurement of writing ability with a many-faceted Rasch model. *Applied Measurement in Education*, n. 5, p. 171-191, 1992.

ESCOFIER, B.; PAGÈS, J. *Análisis Factoriales simples y múltiples*. Universidad del País Vasco, Editora Bilbao, 1992.

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

HOIJTINK, H. A Latent trait model for dichotomous choice data. *Psychometrika*, n. 55, p. 641-656, 1990.

- KARABATSOS, G. A Critique of Rasch residual fit statistics. *Journal of Applied Measurement*, n. 1, p. 152-176, 2000.
- LINACRE, J. M.; WRIGHT, B. D. *A User's guide to BIGSTEPS (Version 2.4)*. Chicago: Mesa Press, 1994.
- LORD, F. M. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale NJ: Erlbaum, 1980.
- LORD, F. M.; NOVICK, M. R. *Statistical Theories of Mental Test Score*. Reading: Addison-Wesley, 1968.
- LUNZ, M. E.; WRIGHT, B. D.; LINACRE, J. M. Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, n. 3, p. 331-345, 1990.
- MASTERS, G. N.; WRIGHT, B. D. The partial credit model. In: VAN DER LINDEN, W. J.; HAMBLETON, R. K. (eds.), *Handbook of modern item response theory*. New York: Springer, p. 101-121, 1997.
- MURAKI, E. A Generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement*, n.16, p. 159-176, 1992.
- RECKASE, M. D. Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics*, n. 4, p. 207-230, 1979.
- ROBERTS, J. S.; LAUGHLIN, J. E. A Unidimensional item response model for unfolding responses from a graded disagree-agree response scale. *Applied Psychological Measurement*, n. 20, p. 231-255, 1996.
- ROBERTS, J. S.; DONOGHUE, J. R.; LAUGHLIN, J. E. A General model for unfolding unidimensional polytomous responses using item response theory. *Applied Psychological Measurement*, v. 1, n. 24, p. 3-32, 2000.
- ROBERTS, J. S.; LIN, Y.; LAUGHLIN, J. E. Computerized adaptive testing with the generalized graded unfolding model. *Applied Psychological Measurement*, n. 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. San Diego, California, April 17, 1998.

ROBERTS, J. S. et al. *GGUM2004: a windows-based program to estimate parameters of the generalized graded unfolding model*. Manuscript preparation, 2004.

SAMARTINI, A. L. S. *Modelos com Variáveis Latentes Aplicados à Mensuração de Importância de Atributos*. São Paulo, 2006. Tese (dout.) Escola de Administração de Empresas de São Paulo.

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

SMITH, M.; SCHUMACKER, R. E.; BUSH, M. J. Using item mean squares to evaluate fit to the Rasch model. *Journal of Outcome Measurement*, n. 2, p.66-78, 1998.

SPSS – Statistical Package for the Social Sciences. Base 10.0 User's Guide. Chicago: SPSS, 1999.

WRIGHT, B. D.; MASTERS, G. N. *Rating scale analysis*. Chicago: Mesa Press, 1982.

Recebido em: abril 2007

Aprovado para publicação em: julho 2007

