



Universidade de Brasília
Departamento de Estatística

JUNIO DE LIMA FREITAS

**APLICAÇÃO DE MODELOS DE TEORIA DA RESPOSTA AO ITEM NO
ESTUDO DAS RESPOSTAS PARA A PROVA DE MATEMÁTICA DO
VESTIBULAR DA UnB**

Brasília
Outubro de 2021

JUNIO DE LIMA FREITAS

**APLICAÇÃO DE MODELOS DE TEORIA DA RESPOSTA AO ITEM NO
ESTUDO DAS RESPOSTAS PARA A PROVA DE MATEMÁTICA DO
VESTIBULAR DA UnB**

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Orientador: Prof. Antônio Eduardo Gomes

**Brasília
Outubro de 2021**

RESUMO

O presente trabalho tem como objetivo avaliar um método alternativo de pontuação para as questões do tipo A (certo ou errado) e C (múltipla escolha de 4 alternativas), somente para a prova de matemática do vestibular da UnB elaborada pelo CEBRASPE. Esse novo método de pontuação consiste em aplicar o Modelo de Resposta Gradual (MRG), da Teoria da Resposta ao Item (TRI), ao considerar como uma alternativa de resposta a “não resposta”. Esta análise começa com um estudo descritivo das respostas dos indivíduos que realizaram a prova, seguida por algumas comparações dos escores calculados. Primeiro será comparado o Método Convencional e o Modelo de Resposta Gradual (MRG), e posteriormente haverá comparação entre este e o Modelo dicotômico de 3 parâmetros também da TRI, tomando a não resposta como erro no de 3 parâmetros. Os resultados mostraram uma alta proximidade entre as notas estimadas pelo modelo de resposta gradual e as notas obtidas pelo método convencional. Dessa forma, pôde-se concluir que o método proposto é eficiente em avaliar os indivíduos com a adição das vantagens que a Teoria da Resposta ao Item oferece.

Palavras Chave :Teoria da Resposta ao Item, Modelo de Resposta Gradual, Modelo dicotômico de 3 parâmetros, Vestibular, CEBRASPE, Matemática, Não Resposta, proficiência.

ABSTRACT

The present work aims to propose an alternative method of points for questions of type A (right or wrong) and C (multiple choice of 4 alternatives), only for the math test for the UnB entrance exam prepared by CEBRASPE. This new scoring method consists of applying the Scoring Gradual Response Model (GRM), from Item Response Theory (IRT), when considering as an alternative answer to “no answer”. This analysis begins with a descriptive analysis of the individuals’ responses who performed the test, followed by some comparisons of the calculated scores. First, the Conventional Method and the Gradual Response Model (GRM) will be compared, and then the comparison between this and the 3 parameters will be made as well of the IRT, taking the non-response as an error in the 3 parameters. The results showed high proximity between the grades estimated by the gradual response model and the grades obtained by the conventional method. Thus, it was possible to conclude that the proposed method is efficient in evaluating individuals with the addition of the advantages that the Item Response Theory offers.

Keywords :Item Response Theory, Gradual Response Model, Dichotomous Model of 3 parameters, Admission test, CEBRASPE, Mathematics, Non-Response, proficiency.

Lista de Figuras

1	Exemplo de uma representação gráfica do MRG.	11
2	Exemplo de uma CCI com $a_i = 1.3$, $b_i = 1.2$ e $c_i = 0.2$	11
3	Curva Característica do Item e sua respectiva Curva de Informação do Item	12
4	Informações sobre o vestibular 2014/2 (Edital N ^o 1 - Abertura, CEBRASPE)	15
5	Frequência de acerto para a prova de matemática.	16
6	Frequência de acerto para a prova de matemática (Tipo A).	17
7	Frequência de acerto para a prova de matemática (Tipo C).	17
8	Frequência da pontuação total obtida na prova de matemática	18
9	Frequência da pontuação obtida na prova de matemática (Tipo A)	19
10	Frequência da pontuação obtida na prova de matemática (Tipo C).	19
11	Curvas características dos itens 1 a 4 (Tipo A)	23
12	Curvas características dos itens 9, 12 e 29 (Tipo C)	23
13	Função de informação dos itens (Questões do tipo A).	24
14	Função de informação dos itens (Questões do tipo C).	25
15	Função de Informação do Teste (FIT)	25
16	Curva característica do item 6 (Tipo A)	28
17	Curva característica dos itens 9, 12 e 29 (Tipo C)	28
18	Função de Informação do Teste (FIT)	29
19	Histograma das Notas Convencionais Padronizadas e das Notas pelo MRG.	31
20	Diagrama de Dispersão entre as Notas Convencionais Padronizadas e as Notas pelo MRG.	31
21	Histogramas das proficiências estimadas pelo modelo dicotômico e pelo MRG.	32
22	Gráfico de dispersão do escore total versus a diferença nas proficiências estimadas pelos 2 modelos.	32
23	Diagrama de Dispersão entre as proficiências estimadas pelo modelo di- cotômico e pelo MRG.	33

Lista de Tabelas

1	Tabela dos quartis para cada tipo de questão.	20
2	Frequência das respostas por item e parâmetros dos itens.	21
3	Análise descritiva para os parâmetros dos itens e proficiências.	22
4	Tabela dos quartis para os parâmetros dos itens e proficiências.	26
5	Parâmetros dos itens para o modelo MRG	27
6	Tabela dos quartis para as notas calculadas pelos modelos propostos. . . .	30

Sumário

1 Introdução	5
2 Objetivos	6
2.1 Objetivo Geral	6
2.2 Objetivos Específicos.	6
3 Teoria de Resposta ao Item	7
3.1 Breve Histórico	7
3.2 Modelos da Teoria de Resposta ao Item	7
3.2.1 Modelos Para Itens Dicotômicos	8
3.2.1.1 Modelo logístico unidimensional de 1 parâmetro	8
3.2.1.2 Modelo logístico unidimensional de 2 parâmetros	8
3.2.1.3 Modelo logístico unidimensional de 3 parâmetros	8
3.2.2 Modelo de Resposta Gradual	9
3.3 Curva Característica do Item	11
3.4 Função de Informação do Item	12
3.5 Função de Informação do Teste	13
4 Metodologia	14
4.1 Materiais	15
5 Resultados	16
5.1 Análise Descritiva	16
5.2 Análise dos Parâmetros dos Itens e da Curva característica dos Itens	20
5.2.1 Análise para o Modelo Dicotômico de 3 parâmetros	20
5.3 Análise para o Modelo de Resposta Gradual	26
5.4 Comparação entre as notas das provas corrigidas pelo Método Convencional, pelo Modelo Dicotômico de 3 parâmetros e pelo MRG	30
6 Conclusão	34
Referências	35

1 Introdução

Os modelos estatísticos de Teoria de Resposta ao Item (TRI) apresentam um leque extenso de formas que frequentemente são utilizadas em programas de avaliações, tais como o Sistema de Avaliação da Educação Básica (SAEB), Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos (CEBRASPE) e entre outros inúmeros testes e provas. A TRI é uma forma mais versátil de medir a habilidade dos indivíduos, quando comparada aos métodos convencionais, pois, além de contabilizar o número total de acertos, tem como foco principal a avaliação dos itens.

Neste projeto estudaremos algumas estimações para exames com questões de certo ou errado e de múltipla escolha apenas da disciplina de matemática como, por exemplo, questões do tipo A e C das provas do vestibular da UnB realizado pelo CEBRASPE. Nas questões do tipo A deste tipo de exame, o respondente ganha 1 ponto quando acerta uma questão, perde 1 ponto quando erra e ganha zero pontos quando não responde. As questões do tipo C contêm quatro opções de respostas, designadas pelas letras A, B, C e D, das quais apenas uma constituirá a resposta correta. Nesse tipo de questão, o respondente ganha 2 pontos quando acerta uma questão, perde 0.667 pontos quando erra e ganha zero pontos quando não responde, mas iremos desconsiderar as penalizações para as questões do tipo A e C.

Embora existam diversas técnicas de medição de conhecimento, boa parte delas não são fáceis e diretas por diversos aspectos técnicos, como é o caso da Teoria de Resposta ao Item. Foi criada como alternativa a Teoria Clássica das Medidas (TCM), cujo papel principal é a proposição de modelos de medição para traços latentes como, por exemplo, a proficiência para uma matéria específica disposta na prova do Vestibular da UnB.

Propomos o cálculo da nota do respondente utilizando o modelo de resposta gradual da Teoria de Resposta ao Item (TRI) considerando três categorias: “Errar”, “Não responder” e “Acertar”. Este modelo considera que as categorias de resposta de uma questão podem ser ordenadas entre si, de tal forma que a categoria mais baixa contribua menos para o escore do respondente e a categoria mais alta contribua mais. Sendo assim, serão comparados os escores tradicionais com a estimação via modelo de resposta gradual.

2 Objetivos

2.1 Objetivo Geral

Avaliar um método alternativo de pontuação para as questões do tipo A e C do vestibular UnB aplicado pelo CEBRASPE. Para este intento, trabalhar-se-á em aplicar o Modelo de Resposta Gradual (MRG) com o propósito de melhorar a avaliação dos indivíduos, visto que a teoria da resposta ao item apresenta inúmeras vantagens em relação ao método convencionalmente utilizado pelo CEBRASPE.

2.2 Objetivos Específicos

- Verificar a existência de correlação entre as notas de matemática e quais as vantagens e desvantagens de se modelar as respostas destes tipos de testes através do modelo de resposta gradual da TRI;
- Comparar a estimação do Modelo de Resposta Gradual com os aplicados convencionalmente pelo CEBRASPE;
- Identificar possíveis ganhos de precisão e informação da pontuação obtida por uma nova técnica;
- Comparar a ordenação dos candidatos utilizando modelos de TRI e a nota convencional para a prova de matemática do vestibular da UnB.

3 Teoria de Resposta ao Item

3.1 Breve Histórico

Em muitos estudos a variável de interesse é de compreensão intuitiva para todos. Entretanto, na maioria dos casos, essa variável não é observável, ou seja, não pode ser medida diretamente e é comumente denominada traço latente. Por exemplo, conhecimento sobre algum assunto, nível de depressão, nível de ansiedade, entre outras.

A partir disso, quando se tem interesse em medir um traço latente, torna-se necessário criar uma escala de medida tal que essa variável assuma valores contidos nessa escala. Entretanto, definir essa escala de medida, o intervalo desta escala e sua interpretação em relação ao traço latente medido são tarefas bastante difíceis por vários motivos técnicos.

Uma das primeiras técnicas utilizadas nessa área foi a Teoria Clássica dos Testes (TCT), que utiliza o escore no teste como sua referência de medida. Já na Teoria de resposta ao Item, o foco principal é o item e não o teste como um todo. No entanto, a TCM apresenta várias limitações. O fato de os resultados encontrados dependerem do particular conjunto de itens (questões) que compõem o instrumento de medida faz com que as análises e interpretações estejam sempre associadas à prova como um todo, tornando-se inviável a comparação entre indivíduos que não foram submetidos às mesmas provas.

Com o intuito de corrigir essas limitações da TCM, surgiu a Teoria da Resposta ao Item (TRI), que propõe modelos de medição para traços latentes com algumas vantagens em relação à técnica tradicionalmente utilizada. A medida de proficiência de um aluno não depende dos itens apresentados a ele, e os parâmetros de discriminação e de dificuldade do item não dependem do grupo de respondentes. Ou seja, um item mede determinada habilidade independentemente de quem o está respondendo, e a proficiência de um aluno não depende dos itens que estão sendo apresentados a ele.

3.2 Modelos da Teoria de Resposta ao Item

Andrade et al. (2000) definem a TRI como o conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros dos itens e da habilidade ou habilidades do respondente, dependendo do modelo em questão. O conjunto de modelos foi desenvolvido de forma que, quanto maior a habilidade do indivíduo, maior a probabilidade de ele dar uma resposta

correta ao item em questão.

Tais modelos dependem essencialmente de três atributos: natureza do item, sendo esta classificada como dicotômica ou não dicotômica; número de populações envolvidas no modelo, discriminado em apenas uma ou mais de uma população; e, a quantidade de traços latentes que estão sendo medidos pelo modelo, classificados como unidimensional (apenas um traço latente medido) ou multidimensional (mais de um traço latente). Neste trabalho, serão abordados modelos unidimensionais, dicotômicos e não dicotômicos, relacionados somente a população das questões de matemática.

3.2.1 Modelos Para Itens Dicotômicos

Em muitos casos, os modelos mais utilizados da TRI são os modelos logísticos para itens dicotômicos. Existem três tipos de modelos nesse contexto, os quais se diferenciam pela quantidade de parâmetros que é utilizada para descrever o item.

3.2.1.1 Modelo logístico unidimensional de 1 parâmetro

O modelo logístico unidimensional de 1 parâmetro (ML1), também conhecido como modelo de Rasch, é dado por

$$P(U_{i,j} = 1|\theta_j) = \frac{1}{1 + e^{-(\theta_j - b_i)}}, i = 1, 2, \dots, I; j = 1, 2, \dots, n; \quad (3.2.1)$$

em que b_i é o parâmetro de dificuldade do item.

3.2.1.2 Modelo logístico unidimensional de 2 parâmetros

O modelo logístico unidimensional de 2 parâmetros (ML2), ao qual, além do parâmetro de dificuldade (b_i), é acrescentado o parâmetro de discriminação do item (a_i), é dado por

$$P(U_{i,j} = 1|\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, i = 1, 2, \dots, I; j = 1, 2, \dots, n. \quad (3.2.2)$$

3.2.1.3 Modelo logístico unidimensional de 3 parâmetros

O modelo logístico unidimensional de 3 parâmetros (ML3), ao qual, além dos parâmetros b_i e a_i , é introduzido o parâmetro de acerto ao acaso (c_i), é dado por

$$P(U_{i,j} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, i = 1, 2, \dots, I; j = 1, 2, \dots, n. \quad (3.2.3)$$

em que

- $U_{i,j}$: assume 1 quando o j -ésimo indivíduo fornece resposta correta para o i -ésimo item, ou assume 0 (zero) quando o j -ésimo indivíduo fornece resposta incorreta para o i -ésimo item;
- θ_j : representa a habilidade do j -ésimo respondente;
- $P(U_{i,j} = 1|\theta_j)$: probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i ;
- b_i : é o parâmetro de dificuldade do i -ésimo item ($-\infty < b_i < \infty$);
- a_i : é o parâmetro de discriminação, ou de inclinação do i -ésimo item ($a_i > 0$);
- c_i : é o parâmetro de acerto ao acaso do i -ésimo item ($0 < c_i < 1$).

3.2.2 Modelo de Resposta Gradual

O MRG é uma generalização do modelo logístico de 2 parâmetros, assumindo que as categorias de resposta de um item podem ser ordenadas entre si, de tal forma que, quanto mais alta a alternativa escolhida pelo respondente, maior deve ser a intensidade do traço latente. Por exemplo, como uma escala de Likert, a qual consiste em uma série de cinco proposições, das quais o inquirido deve selecionar uma, podendo estas ser: discorda totalmente, discorda, sem opinião, concorda, concorda totalmente. Sobre isso, Baker (1992) comenta que o modelo de resposta gradual tem uma importante contribuição à TRI uma vez que ele é uma extensão do modelo binário e desta forma abre precedentes à análise de escalas do tipo Likert. Como resultado, a resposta gradual pode ser modelada usando uma análise muito parecida com o caso dicotômico.

Suponha que os escores das categorias de um item i são arranjados em ordem crescente e denotados por $k = 0, 1, \dots, m_i$, onde $(m_i + 1)$ é o número de categorias do i -ésimo item. A probabilidade de um indivíduo j escolher uma categoria particular ou outra mais alta do item i pode ser dada como uma extensão do modelo logístico de 2 parâmetros por

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_{i,k})}}, \quad (3.2.4)$$

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

em que

- $P_{i,k}^+(\theta_j)$: é a probabilidade de um indivíduo j escolher a categoria k ou outras mais

altas do item i ;

- $b_{i,k}$: é o parâmetro de dificuldade da k -ésima categoria do item i ;
- θ_j : representa a habilidade do j -ésimo indivíduo;
- a_i : parâmetro de discriminação do item i , com valor proporcional à inclinação da curva característica do item, no ponto.

O parâmetro de discriminação a_i varia a cada item, mas é constante para o mesmo item. Essa restrição de igual inclinação em cada categoria tem a finalidade de evitar probabilidades negativas.

Para o parâmetro de dificuldade $b_{i,k}$, por definição, tem-se, sem perda de generalidade, que

$$b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,m_i} \quad (3.2.5)$$

ou seja, de acordo com a literatura, deve-se ter necessariamente uma ordenação entre o nível de dificuldade das categorias de determinado item i de acordo com a classificação de seus escores.

A probabilidade de um respondente j obter um escore k no item i é dada pela expressão

$$P_{i,k}(\theta_j) = P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j) \quad (3.2.6)$$

de modo que $P_{i,0}^+(\theta_j) = 1$ e $P_{i,m_i+1}^+(\theta_j) = 0$.

Portanto

$$P_{i,0}(\theta_j) = P_{i,0}^+(\theta_j) - P_{i,1}^+(\theta_j) = 1 - P_{i,1}^+(\theta_j)$$

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

e então temos que

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

,

Note que em um item com $(m_i + 1)$ categorias, m_i valores de dificuldade necessitam ser estimados, além do parâmetro de discriminação do item. Assim, para cada item, o número de parâmetros a ser estimado será dado pelo seu número de categorias de resposta.

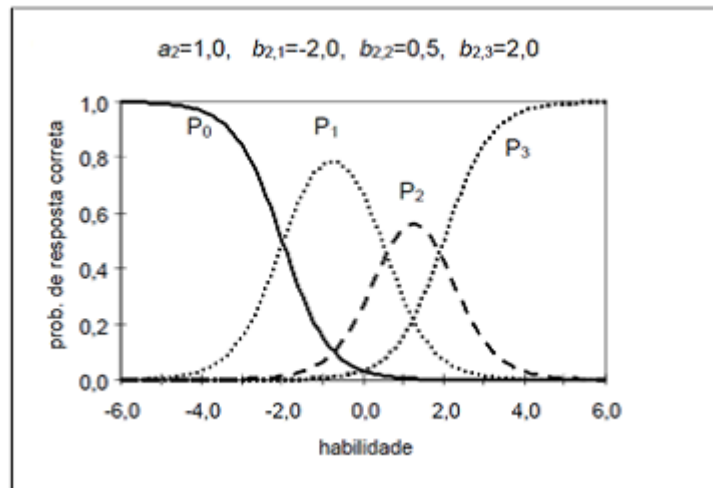


Figura 1: Exemplo de uma representação gráfica do MRG.

Fonte: Andrade, Tavares e valle (2000).

3.3 Curva Característica do Item

Os modelos mais utilizados dentro da TRI atualmente são paramétricos e assumem forma funcional específica para a curva característica do item (CCI), normalmente logística. A CCI evidencia o progresso da probabilidade de acerto do item de acordo com a proficiência do respondente.

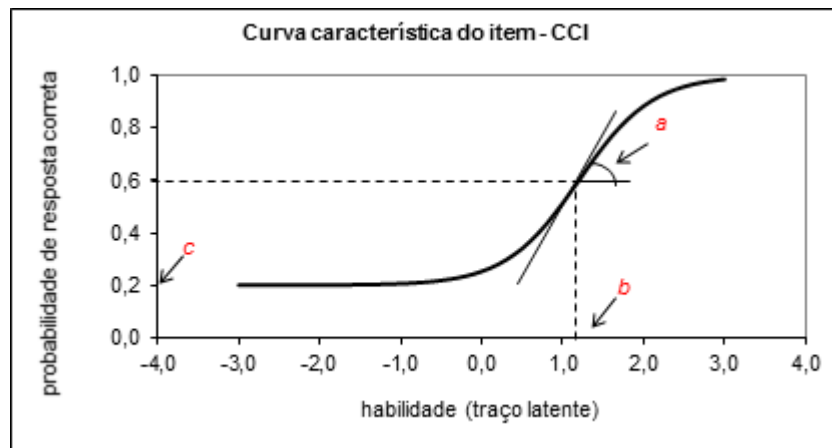


Figura 2: Exemplo de uma CCI com $a_i = 1.3$, $b_i = 1.2$ e $c_i = 0.2$

Fonte: Andrade et al. (2000)

Analisando a figura 2, percebe-se que o valor de a_i é diretamente proporcional à inclinação da curva característica do item (CCI), que fornece $P(U_{ij} = 1|\theta_j)$ como função de θ_j ; e quanto maior o valor de a_i , maior a inclinação da curva. Ou seja, maior a diferença entre as probabilidades de resposta positiva de dois indivíduos, um com proficiência acima de b_i e outro com proficiência abaixo de b_i . Com isso, quanto maior o valor de a_i , maior

é o poder de discriminação do item entre indivíduos com proficiências acima e abaixo do parâmetro de dificuldade do item.

3.4 Função de Informação do Item

A função de informação do item é um poderoso instrumento para análise de itens, possibilitando o conhecimento não só de quanto de informação um item acumula num determinado valor de θ , mas também em que valor de θ o item possui maior quantidade de informação.

Ao analisarmos as funções de informação dos testes e, em seguida, dos itens, podemos deduzir quando elas expressam sua capacidade máxima de informação respectivamente, em relação aos modelos logísticos de um, dois e três parâmetros. Nos modelos de um e dois parâmetros os itens atingem sua capacidade máxima de informação quando θ for igual à dificuldade dos itens ($\theta = b$) e os valores da função de informação são simétricos para cada valor de θ acima ou abaixo do valor de $\theta = b$. No modelo de três parâmetros a quantidade de informação será sempre menor que nos dois outros modelos devido à influência do parâmetro probabilidade de acerto ao acaso c . A função de informação do item é dada por

$$I_i(\theta) = \frac{[p'(\theta)]^2}{p(\theta)[1 - p(\theta)]}, \quad (3.4.1)$$

em que

- $I_i(\theta)$: informação fornecida pelo item i no nível de habilidade θ ;
- $U_i \sim \text{Bernoulli}(p(\theta))$;
- $p(\theta) = P(U_i = 1|\theta)$.

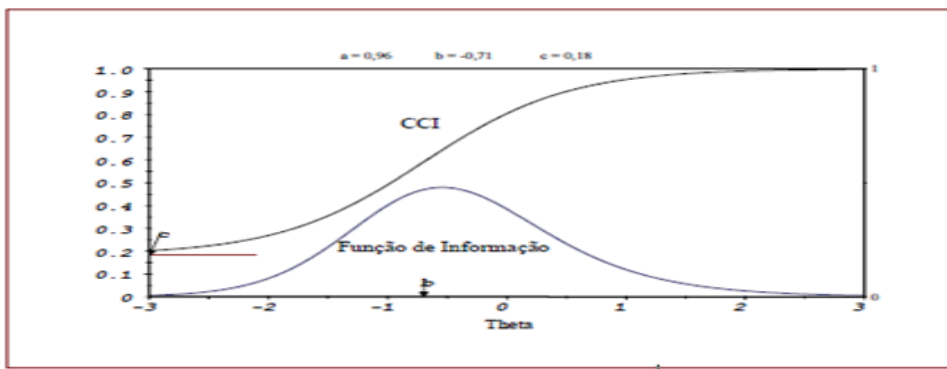


Figura 3: Curva Característica do Item e sua respectiva Curva de Informação do Item

Fonte: Andriola (2009)

De acordo com a Figura 2, o gráfico mostra uma maior quantidade de informação quando θ está entre -1.5 e 0.5.

3.5 Função de Informação do Teste

A Função de Informação do Teste é fornecida simplesmente pela soma das informações fornecidas por cada item que compõe o referido teste. Assim, pode ser calculada por

$$I(\theta) = \sum_{i=1}^I I_i(\theta). \quad (3.5.1)$$

A função de informação do teste ainda pode ser calculada pelo erro padrão de estimação, que é dado por

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}}. \quad (3.5.2)$$

Com isso, pode-se verificar que todas essas medidas de informação dependem do valor de θ .

4 Metodologia

A Teoria de Resposta ao Item surge como uma alternativa ao método clássico de estimação. Sua modelagem utiliza dados provenientes de itens de questionários. Segundo Demars (2010), os modelos da Teoria da Resposta ao Item demonstram que existe uma relação entre as habilidades medidas pelo instrumento e uma resposta a um determinado item. As respostas dos itens podem ser classificadas como dicotômicas (duas categorias), tais como certo ou errado, sim ou não, concordar ou discordar; ou podem ser classificadas como politômicas, que visam analisar itens que apresentam mais de dois escores possíveis (categorias), por exemplo, opções de respostas na escala de Likert de algum determinado traço latente.

Como citado na introdução, para o cálculo da nota do respondente para o vestibular da UnB utilizando o modelo de resposta gradual da Teoria de Resposta ao Item (TRI), serão consideradas três categorias: “Errar”, “Não responder” e “Acertar”. Sendo assim, será elaborado um sistema de pontuação, com pontuação diferente para cada tipo de questão, mas considerando a categoria adicional de “não resposta” como uma categoria intermediária, visto que ela não é tão desvantajosa para o indivíduo como a categoria “Errar”, mas também não é tão vantajosa para o indivíduo como a categoria “Acertar”.

Neste sentido, podemos comparar os escores tradicionais com a estimação via modelo de resposta gradual, analisando, por exemplo, se há correlação entre as notas e quais as vantagens e desvantagens de se modelar as respostas destes tipos de testes através do modelo de resposta gradual da TRI.

A diferença entre o cálculo das proficiências dos indivíduos pelo modelo da TRI e o cálculo da nota final pelo método convencional está no fato de que, neste último, a pontuação adquirida pelas respostas das questões é sempre a mesma, independentemente de se o respondente acertou uma questão fácil ou difícil, ou também, se errou uma questão fácil ou difícil. Já o modelo da teoria de resposta ao item funciona de forma distinta, pois há diferença na pontuação final do indivíduo caso ele acerte uma questão fácil ou uma questão difícil, ganhando menos pontos ou mais pontos, respectivamente.

No caso em estudo, o modelo de resposta gradual pode ter, ainda, uma diferença de pontuação da questão entre as três categorias de respostas. Por exemplo, em determinada questão, dependendo dos parâmetros de dificuldade de suas respectivas categorias, o indivíduo pode ganhar uma alta pontuação se acertar a questão e, caso ele erre ou não responda, pode não haver uma diferença tão grande entre a pontuação obtida ao se marcar uma dessas últimas categorias. Ou seja, não será padrão, pois no modelo de resposta gradual não é feita a soma das pontuações, mas sim a estimação das probabilidades baseadas no modelo apresentando na seção 3.2.2.

Em geral será dada continuidade ao estudo estimando a proficiência utilizando o modelo logístico de 3 parâmetros considerando “não resposta” como erro, além do modelo alvo desse trabalho, o Modelo de Resposta Gradual, considerando “não resposta” como uma categoria intermediária entre acerto e erro do item. Serão feitas estimações com esses modelos para que sejam comparadas com as notas obtidas pelo método convencional. Também será feita análise das curvas características do item e do teste para verificar cada parâmetro estudado nos modelos citados. Para a análise serão usados os pacotes “lrm” e “mirt” do software R, que podem ser utilizados para a estimação.

Para trabalhos futuros, tem-se também a possibilidade de ser estudada uma outra ordenação para as categorias do Modelo de Resposta Gradual, considerando as 3 seguintes categorias: “Não responder”, “Errar” e “Acertar”.

4.1 Materiais

O conjunto de dados fornecido para este trabalho é do vestibular da UnB do segundo semestre de 2014 disponibilizado pelo CEBRASPE. A aplicação das provas foi realizada em 2 dias, com 3 três tipos de cadernos (I,II,III) em cada dia. No total, foram 21.968 respondentes e 300 questões. Neste projeto serão analisadas somente 7232 respostas referentes as 33 questões de matemática, sendo 30 do tipo A (certo-errado) e 3 do tipo C (múltipla escolha com 4 alternativas) do caderno provas do tipo I para o segundo dia.

Segue abaixo um quadro com algumas descrições do vestibular UnB 2014/2.

Data	Prova	Disciplinas-Foco	Nº de itens	Duração
1º DIA 7/6/2014	Conhecimentos – Parte I	Língua Espanhola, Língua Francesa ou Língua Inglesa	30	300min
	Conhecimentos – Parte II	Língua Portuguesa e Literaturas de Língua Portuguesa, Geografia e História, Artes (Artes Cênicas, Artes Visuais e Música), Filosofia e Sociologia	120	
	Redação em Língua Portuguesa	–	–	
2º DIA 8/6/2014	Conhecimentos – Parte III	Biologia, Física, Química e Matemática	150	300min

Figura 4: Informações sobre o vestibular 2014/2 (Edital N^o1 - Abertura, CEBRASPE)

5 Resultados

5.1 Análise Descritiva

Analisando os dados disponibilizados pelo CEBRASPE, percebe-se que havia três tipos de provas diferentes. Cada tipo continha as mesmas questões e quantidades. Entretanto, elas se diferenciavam pela ordem em que as questões foram dispostas, não havendo nenhum padrão de ordenação, nem de forma geral e nem dentro de cada assunto/área abordada. Em outras palavras, a primeira questão da prova do tipo I não correspondia à primeira questão da prova do tipo II e/ou III. Assim, dentro das informações do vestibular de 2014/2, foi identificado que 7.347 indivíduos responderam a prova do tipo I, 7.307 responderam a prova do tipo II e 7.312 responderam a prova do tipo III. O nosso interesse é em relação à prova do tipo I, somente para as questões de matemática, que são no total 33 questões, sendo 30 questões do tipo A (certo-errado) e 3 questões do tipo C (multipla escolha de 4 alternativas, somente uma correta).

Em síntese, foram analisadas as 238.656 respostas de 7.232 indivíduos que responderam 33 questões de matemática do vestibular da UnB de 2014/2 elaborado pelo Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos (CEBRASPE).

Para essa análise foram feitos gráficos com as frequências para ambos tipos de questões. O gráfico da Figura 5 representa a quantidade de acerto dos respondentes para a prova de matemática do tipo I, considerando certo, errado e não resposta. Em geral, de acordo com o gráfico, entre os respondentes, há uma alta taxa de não resposta. Talvez isso se dê por existir uma grande insegurança em se chegar na resposta certa.

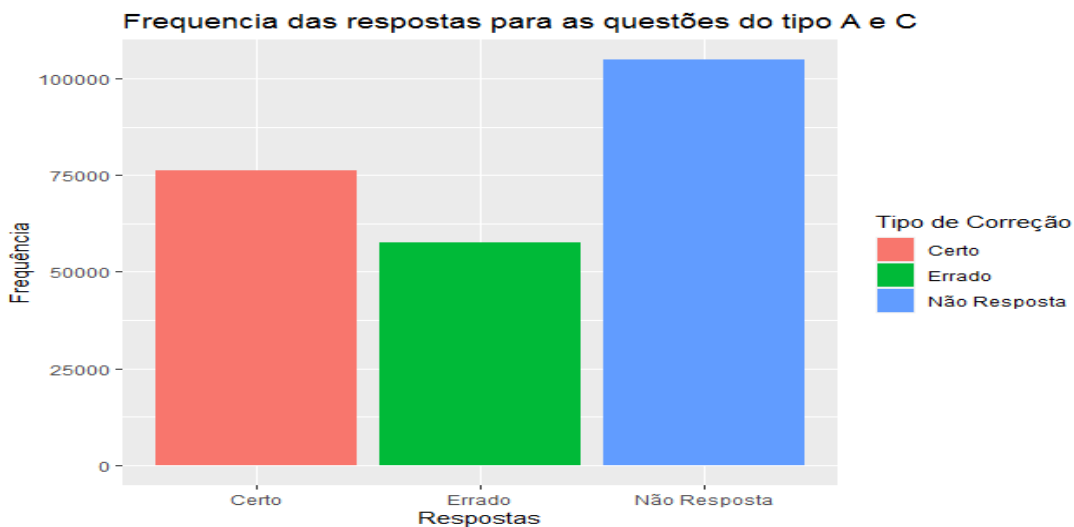


Figura 5: Frequência de acerto para a prova de matemática.

Os gráficos das Figuras 6 e 7 representam a frequência de acerto para a prova de matemática separada por tipo de questão, A e C respectivamente. Quando separadas por tipo de questão, percebe-se que as questões do tipo A têm um índice de acerto maior que as questões do tipo C, indicando que as questões do tipo C aparentam ter um maior nível de dificuldade.

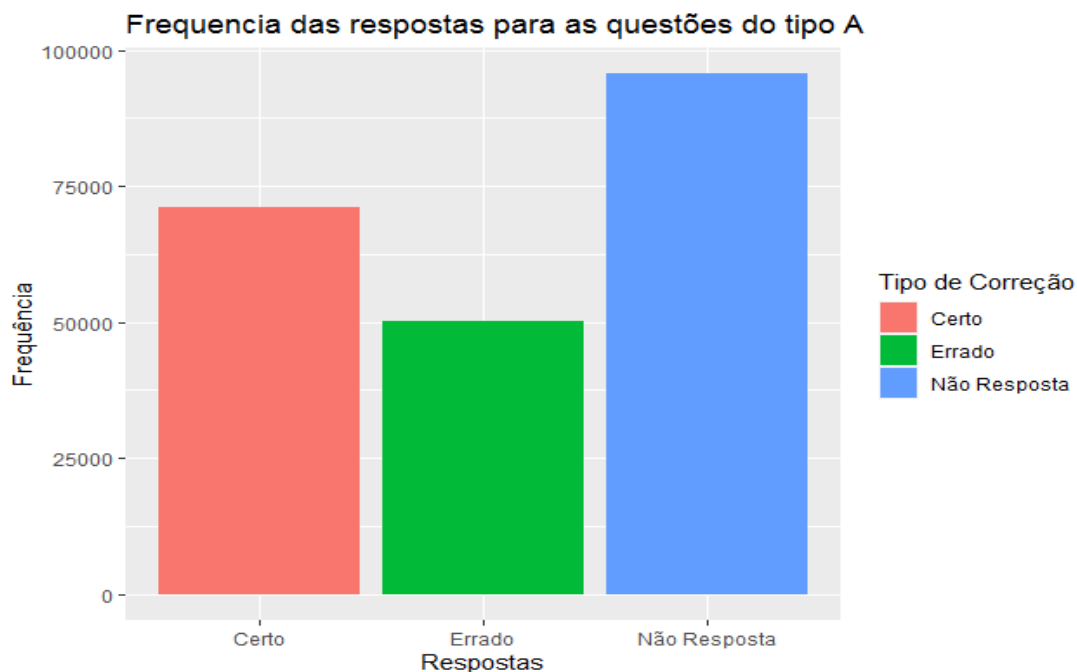


Figura 6: Frequência de acerto para a prova de matemática (Tipo A).

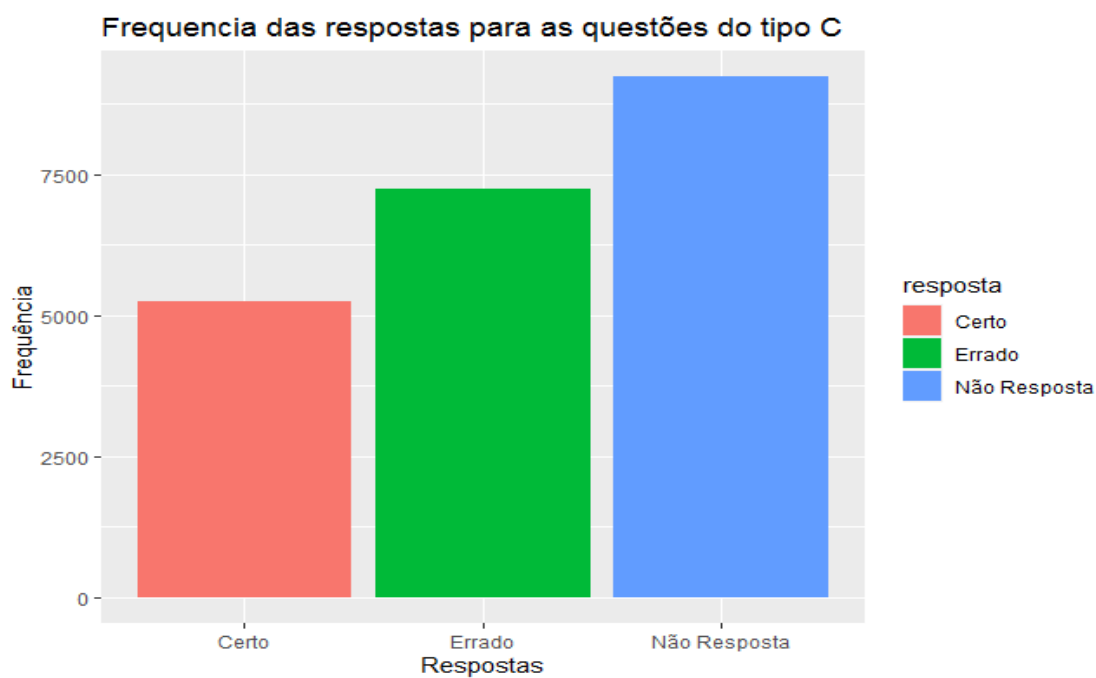


Figura 7: Frequência de acerto para a prova de matemática (Tipo C).

A fim de verificar a pontuação obtida por cada indivíduo em relação à prova de matemática, pode-se calcular o escore tradicional. O gráficos das Figuras 8-10 representam histogramas das notas obtidas por cada respondente. Na figura 8, temos histograma da pontuação total obtida na prova de matemática considerando os dois tipos de questões propostas nesse estudo.

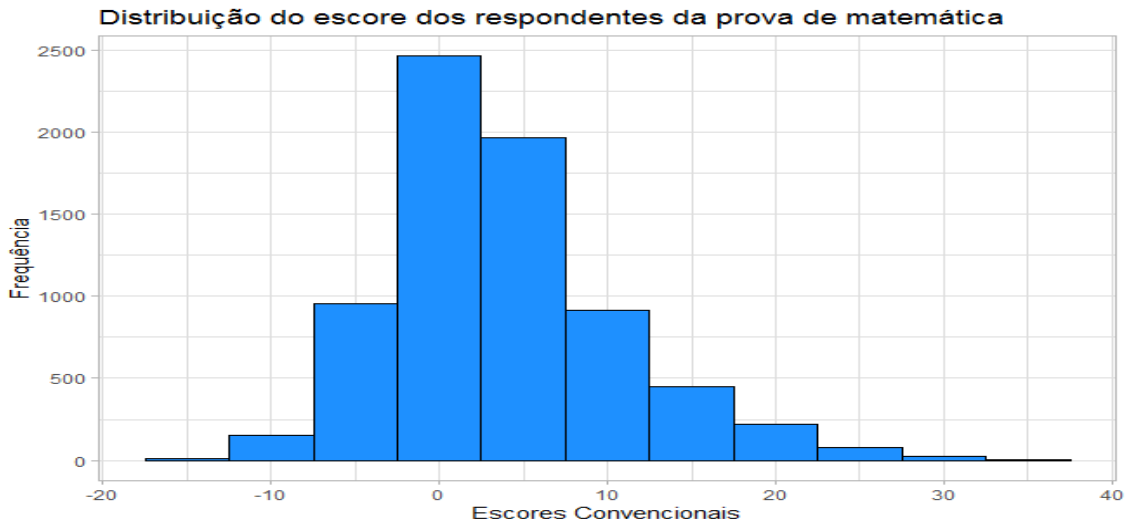


Figura 8: Frequencia da pontuação total obtida na prova de matemática

Os escores tradicionais para as questões do tipo A apresentaram uma concentração maior entre -10 e 10 (figura 9), indicando uma baixa pontuação para boa parte dos indivíduos. Para esse tipo de questão, vale lembrar que a pontuação é dividida da seguinte forma:

- +1 ponto caso o respondente responda corretamente;
- -1 ponto caso o respondente responda incorretamente;
- 0 pontos caso o responde não responda a questão.

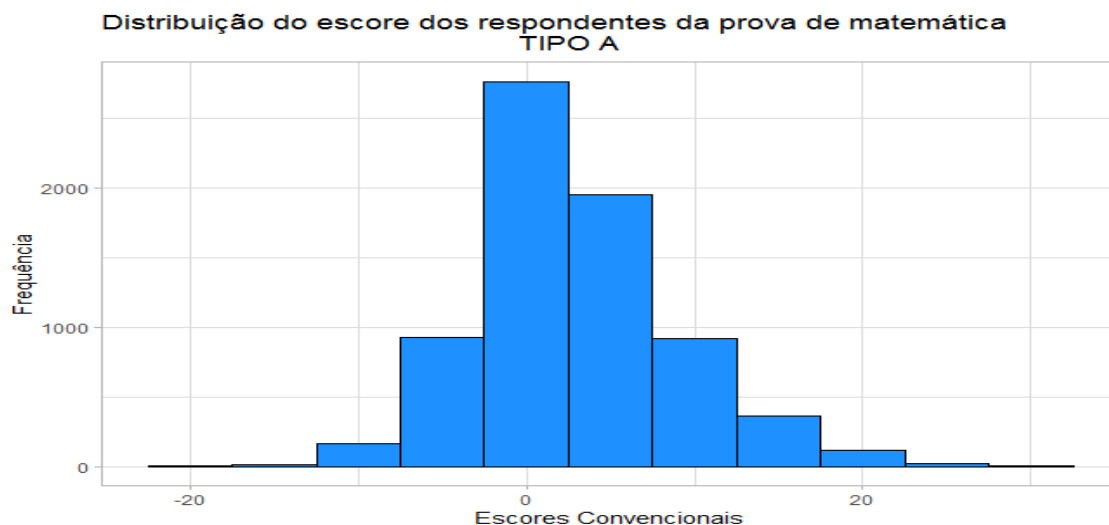


Figura 9: Frequência da pontuação obtida na prova de matemática (Tipo A)

Por outro lado, para as questões do tipo C (figura 10), verifica-se que as notas estão concentradas entre -5 e 5, também indicando uma baixa taxa de acerto. A pontuação para esse tipo de questão é dada de outra forma:

- +2 pontos caso o respondente responda corretamente;
- -0.667 ponto caso o respondente responda incorretamente;
- 0 pontos caso o respondente não responda a questão.

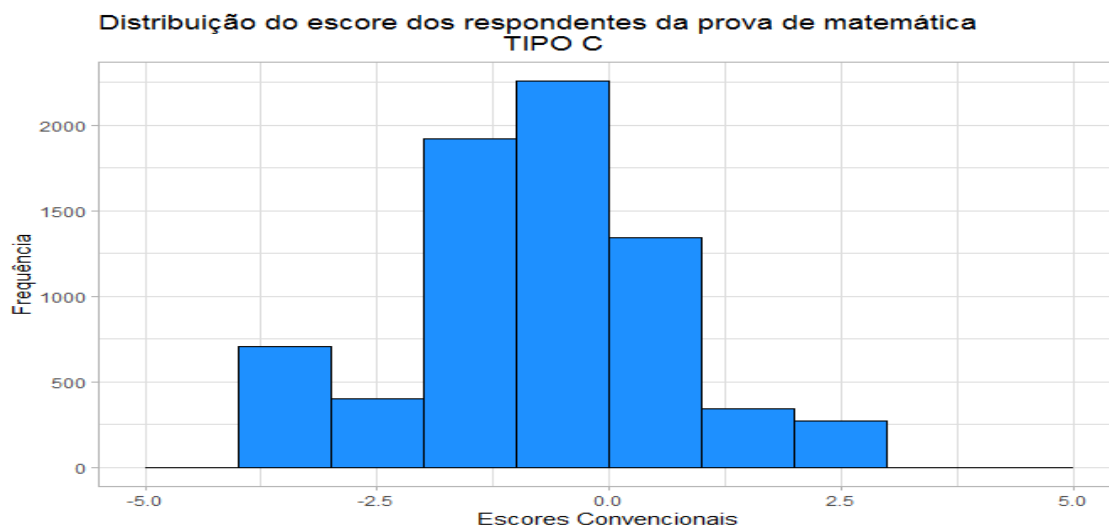


Figura 10: Frequência da pontuação obtida na prova de matemática (Tipo C).

A Tabela 1 representa uma descrição dos quartis para os escores obtidos para a nota geral (escores TIPO A+C). Temos um mínimo de -17 pontos e um máximo de 35 pontos, indicando que em geral havia questões fáceis e difíceis na prova de matemática. Porém, 50% dos respondentes obtiveram uma nota baixa.

	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Escores TIPO A	-18.00	-1.00	2.00	2.884	6.00	29.00
Escores TIPO C	-2.0010	-0.6670	0.6660	0.7806	2.00	6.00
Escores TIPO A + C	-17.001	-1.00	2.666	3.665	7.00	35.00

Tabela 1: Tabela dos quartis para cada tipo de questão.

5.2 Análise dos Parâmetros dos Itens e da Curva característica dos Itens

5.2.1 Análise para o Modelo Dicotômico de 3 parâmetros

Depois do tratamento do banco de dados, foi aplicado o modelo dicotômico de 3 parâmetros, a fim de estimar as respectivas proficiências dos indivíduos. A partir da aplicação dos modelos, foram estimados os parâmetros dos itens, as proficiências, as curvas características dos itens e por fim a curva de informação do teste.

Na Tabela 2 estão apresentados os valores dos parâmetros dos itens estimados e a frequência de respostas dos indivíduos por item, tanto para as questões do Tipo A quanto para o Tipo C. A partir desses resultados, podemos observar que, por exemplo, a porcentagem de acertos do item 1 (Tipo A) resultou em 58,85% de indivíduos que não responderam ou erraram essa questão e 41,15% acertaram. Analisando o índice de discriminação deste item, pode-se concluir que ele não discrimina muito bem os indivíduos e, ao se observar os valores dos parâmetros de dificuldade, também observamos valores baixos. Podemos concluir então que este seria um item relativamente fácil.

Observando agora o item 28 (Tipo A), verificamos que foi o item que os indivíduos mais acertaram. Há indícios de uma discriminação razoavelmente baixa dos indivíduos ($a_{28} = 0,65$), mas percebe-se que ele é um item muito fácil, pois tem baixo valor do parâmetro de dificuldade ($b_{28} = -0,485$), fazendo com que muitos indivíduos acertem a questão.

O item 15 (Tipo A) pode ser considerado um item difícil, pois observamos uma baixa taxa de acerto (15,16%), além de ter seu índice de dificuldade $b_{15} = 1,73$ relativamente alto.

Ainda na tabela 2, observamos os itens 9, 12 e 29 que, por sua vez, são os únicos itens de múltipla escolha (Tipo C). Dentre os três itens citados, o item 29 é considerado o mais difícil, por ter seu índice de dificuldade maior que todos os itens ($b_{29} = 2,49$), com uma taxa de acerto de apenas 17,13%. Apesar de ser um item relativamente difícil o item não discrimina muito bem os indivíduos ($a_{29} = 0,691$). O item 9 é parecido com o item 12, porém o item 12 discrimina melhor os indivíduos, com parâmetro $a_i = 1,702$. Ambos

também apresentam parâmetros de dificuldade bem parecidos e podem ser consideradas questões difíceis com taxas de acertos de apenas 31,55% e 23,70%, respectivamente.

Itens	Categoria -1 e 0 (%)	Categoria 1 (%)	c_i	b_i	a_i
Item 1	58,85	41,15	0,000	0,490	0,864
Item 2	67,27	32,73	0,000	0,854	1,037
Item 3	64,93	35,07	0,000	0,613	1,405
Item 4	78,17	21,83	0,000	1,336	1,214
Item 5	70,23	29,77	0,000	1,105	0,915
Item 6	58,48	41,52	0,257	1,101	1,804
Item 7	61,91	38,09	0,084	0,650	1,766
Item 8	55,50	44,50	0,217	0,874	1,392
Item 9	68,45	31,55	0,100	1,201	1,237
Item 10	76,47	23,53	0,000	1,333	1,084
Item 11	68,56	31,44	0,000	0,858	1,159
Item 12	76,30	23,70	0,055	1,234	1,702
Item 13	69,83	30,17	0,000	0,910	1,181
Item 14	73,53	26,47	0,000	1,575	0,721
Item 15	84,40	15,60	0,000	1,713	1,244
Item 16	72,69	27,31	0,000	0,993	1,305
Item 17	61,48	38,52	0,000	0,541	1,101
Item 18	70,01	29,99	0,056	0,985	1,501
Item 19	80,09	19,91	0,000	1,951	0,804
Item 20	76,20	23,80	0,000	1,566	0,854
Item 21	65,22	34,78	0,000	0,789	0,959
Item 22	72,79	27,21	0,001	2,716	0,375
Item 23	67,05	32,95	0,000	1,509	0,499
Item 24	72,70	27,30	0,000	1,851	0,567
Item 25	72,33	27,67	0,000	2,132	0,474
Item 26	48,76	51,24	0,124	0,308	0,891
Item 27	49,43	50,57	0,144	0,420	0,893
Item 28	42,74	57,26	0,000	-0,485	0,652
Item 29	82,87	17,13	0,000	2,495	0,691
Item 30	59,40	40,60	0,000	0,679	0,614
Item 31	77,00	23,00	0,000	1,434	1,011
Item 32	71,83	28,17	0,000	1,415	0,741
Item 33	69,01	30,99	0,000	1,782	0,473

Tabela 2: Frequência das respostas por item e parâmetros dos itens.

Através da Tabela 3, é possível observar que, para os valores referentes ao parâmetro de discriminação, o valor mínimo resultou em 0,3752, indicando que houve itens com baixa discriminação, baseando-se no ponto de corte que a literatura indica, sendo este maior que 1. Por outro lado, podemos observar também que houve itens que discriminaram bem os indivíduos, visto que o valor máximo resultou em 1,8043. Apesar de termos itens com baixa discriminação, em média obtivemos itens com índices de discriminação em torno de 1,0040, indicando que, em média, o teste como um todo fará uma boa diferenciação entre os indivíduos baseado na escolha de responder ou não a questão.

Parâmetro	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Discriminação	0.3752	0.7213	0.9589	1.0040	1.2373	1.8043
Dificuldade	-0.4851	0.7895	1.1055	1.1796	1.5656	2.7163
Acerto ao acaso	0	0	0	0.03145	0.0009972	0.257
Proficiência	-1.9394	-0.6219	0.0335	0.0005	0.6317	3.2125

Tabela 3: Análise descritiva para os parâmetros dos itens e proficiências.

Os parâmetros de dificuldade estão entre $-0,4851$ e $2,7163$, indicando que há itens fáceis e difíceis. Já a proficiência está entre $-1,9394$ e $3,2125$, ou seja, há indivíduos com habilidades baixas, o que de fato acontece na maioria das provas que objetivam medir algum tipo de conhecimento, pois muitos indivíduos não respondem a questão em razão de não terem alto conhecimento do assunto abordado e da mesma forma há indivíduos com habilidade alta, indicando que também há indivíduos com um alto conhecimento do assunto.

Após analisar os valores da estimação dos parâmetros dos itens, foram gerados os gráficos das Curvas Características dos Itens (CCI), as quais representam a relação entre a probabilidade do indivíduo acertar o item dada a sua habilidade, $P(U_{ij} = 1|\theta_j)$, e os parâmetros do modelo. Os gráficos da CCI têm em seu eixo horizontal o valor do traço latente medido (habilidade) e o eixo vertical corresponde à probabilidade de o indivíduo ter sua resposta classificada como certa.

Abaixo temos a figura 11 que ilustra as CCI's dos itens 1 ao 4 (Tipo A). Nele podemos observar que em geral os indivíduos que concentram uma habilidade entre -1 e 2.5 têm maior probabilidade de acertar os itens citados. Também podemos observar pelo gráfico, que o item 1 (em preto), apresenta o ponto de inflexão mais à direita, indicando ser um item fácil com menor valor de b_i . Além disso, também possui menor inclinação, indicando ser o item que menos discrimina os indivíduos. O item 4 (em amarelo), de acordo com o gráfico, indica ser o item mais difícil e que mais discrimina os indivíduos dentre os 4 analisados, pois seu ponto de inflexão está mais à esquerda e sua curva apresenta a maior inclinação.

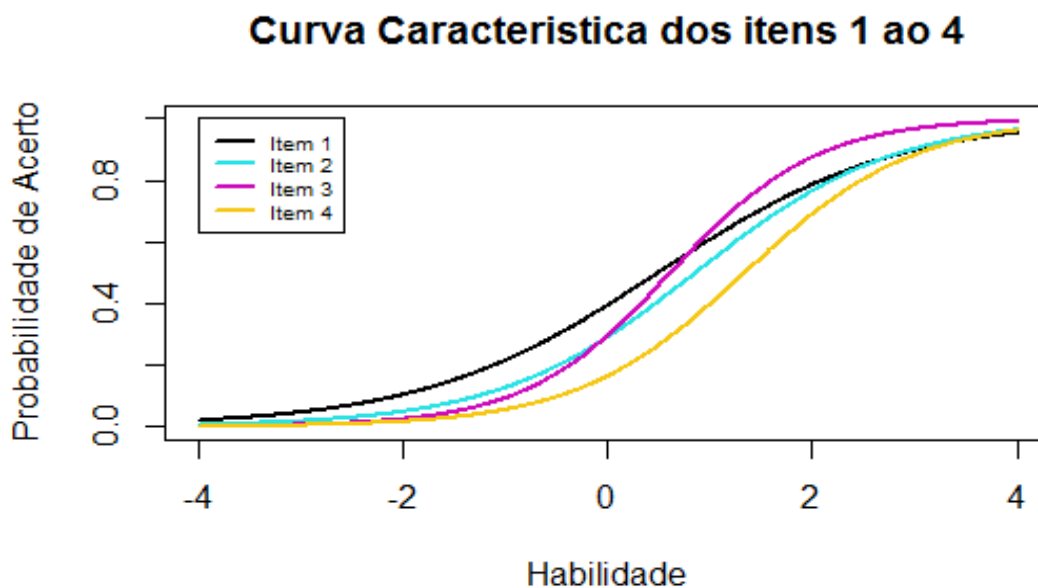


Figura 11: Curvas características dos itens 1 a 4 (Tipo A)

Na figura 12, tem-se as curvas características das únicas questões do tipo C (múltipla escolha) representadas pelos itens 9, 12 e 29. Analisando o gráfico percebemos que os indivíduos com habilidades entre 0 e 2.5 obtiveram uma maior probabilidade de acerto.

Os itens 9 e 12 são bem parecidos, como citado na análise da tabela 3. Porém, o item 29 se destaca por apresentar uma curva pouco inclinada, indicando ser um item com baixa taxa de discriminação, mas apresenta o ponto de inflexão mais à esquerda, indicando ser o item mais difícil entre os do tipo C.

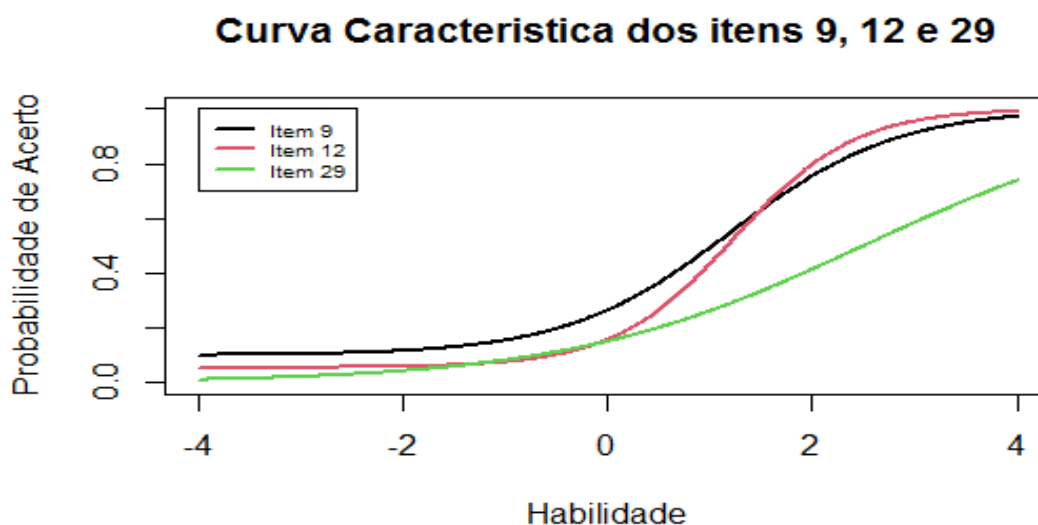


Figura 12: Curvas características dos itens 9, 12 e 29 (Tipo C)

Observando o gráfico de informações dos itens na figura 13, as curvas indicam que em geral os itens do tipo A são pouco informativos para pessoas com habilidades abaixo de 0. Porém, alguns itens se destacam, por exemplo, os itens 1, 5, 10, 15, 23, 30 e 31 são consideravelmente informativos para as pessoas com habilidades entre -2 e 2. Os itens 2, 4, 17, 18, 20, 21 e 32, fornecem quantidade relevante de informação para pessoas com proficiência acima de 0.

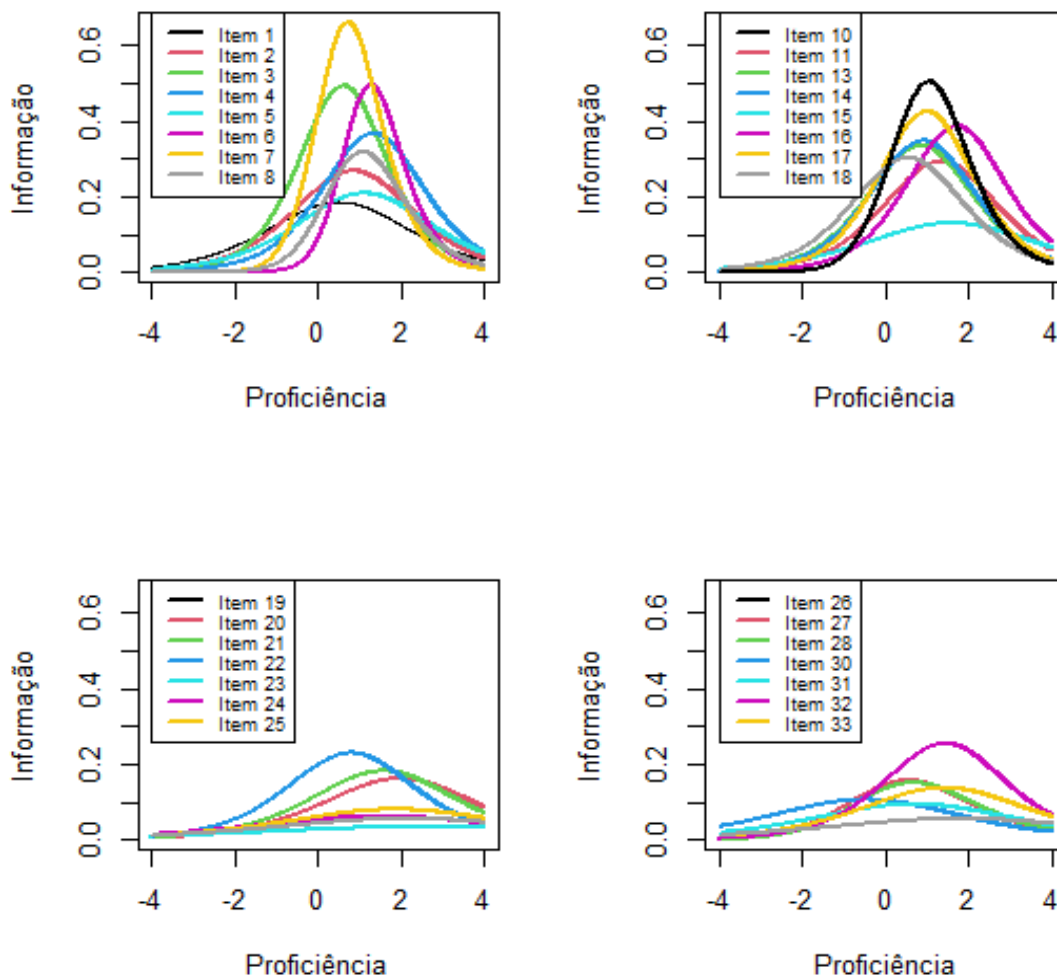


Figura 13: Função de informação dos itens (Questões do tipo A).

A figura 14 representa o gráfico da função de informação dos itens para questões do tipo C. Os itens 9 e 12 indicam boas informações para indivíduos com habilidades acima de 1. De acordo com o gráfico da função de informação dos itens, o item 29 em geral não é um bom avaliador devido a sua pouca discriminação dos itens.

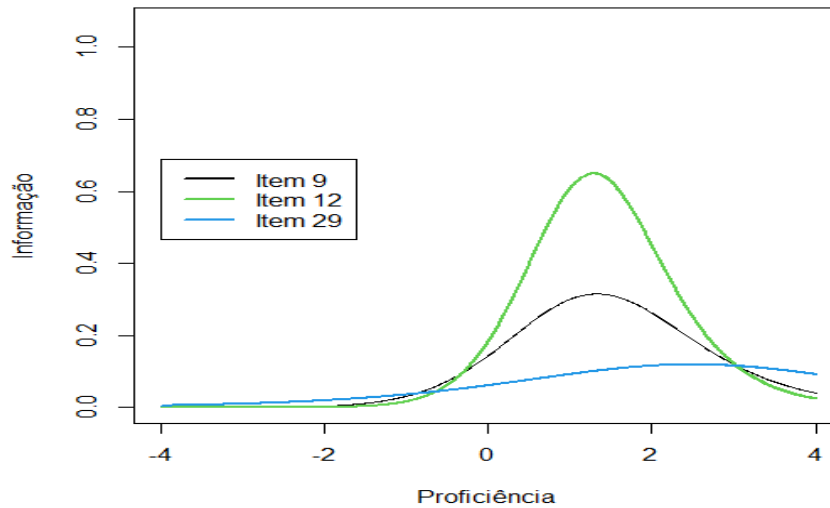


Figura 14: Função de informação dos itens (Questões do tipo C).

Na Figura 15 é apresentada a Função de Informação do Teste (FIT). Através desse gráfico podemos observar que o instrumento de medida tem maior informação para os valores da habilidade compreendidos entre aproximadamente -1 e 2. A curva ideal seria uma curva próxima ao formato de uma distribuição normal padrão, em que sua amplitude esteja incorporando a maior parte das habilidades dos indivíduos, ou seja, em geral o modelo se saiu bem, seguindo o que a literatura indica.

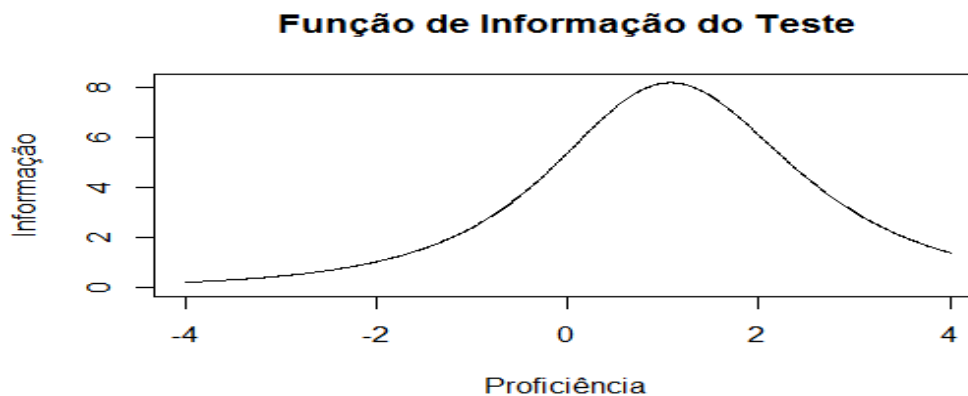


Figura 15: Função de Informação do Teste (FIT)

5.3 Análise para o Modelo de Resposta Gradual

Para verificar se o modelo de resposta gradual é melhor que o modelo de 3 parâmetros, fizemos a mesma análise. Foram estimados os parâmetros dos itens, as curvas características dos itens e por fim a curva de informação do teste.

Na tabela 4, é possível observar que os valores mínimo e máximo, referentes ao parâmetro de discriminação foram, respectivamente, 0,6090 e 2,020, indicando que os itens tiveram um boa discriminação usando o modelo MRG, mostrando ser até mais eficiente que o modelo de 3 parâmetros.

Quanto aos parâmetros de dificuldade dos itens obtivemos um mínimo de -2.1377 e máximo de 1.9019 . De acordo com a literatura o ideal seria entre -2 e $+2$, porém, os valores estão bem próximos desse intervalo. Em relação a proficiência, o mínimo foi de -2.8095 , indicando que há candidatos com habilidade muito baixa, ainda mais baixa quando estimado pelo modelo de 3 parâmetros. Por outro lado, o máximo foi de 2.9057 , indicando que também há candidatos que se sobressaíram muito bem.

Parâmetro	Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo
Discriminação	0.690	0.992	1.137	1.149	1.350	2.020
Dificuldade b_{i1}	-2.1277	-0.5002	-0.1661	-0.3276	0.0858	0.7148
Dificuldade b_{i2}	-0.3956	0.6047	0.9695	0.8913	1.1567	1.9019
Proficiência	-2.8095	-0.6292	0.0672	-0.0005	0.7227	2.9057

Tabela 4: Tabela dos quartis para os parâmetros dos itens e proficiências.

Na tabela 5 estão apresentados os valores dos parâmetros dos itens estimados e a frequência das respostas dos indivíduos para cada item, tanto para as questões do Tipo A quanto para o Tipo C. Analisando primeiro os itens do Tipo A, assim como no modelo 3 de parâmetros, o item 28 foi o item em que os indivíduos mais acertaram. Há indícios de uma discriminação razoavelmente baixa dos indivíduos ($a_{28} = 0.69$), mas percebe-se que ele é um item muito fácil, pois tem baixos valores dos parâmetros de dificuldade ($b_{28} = -0.40$), fazendo com que muitos indivíduos acertem a questão.

Ainda na tabela 5, observamos os itens 9, 12 e 29 que, por sua vez são os únicos itens de múltipla escolha (Tipo C). Dentre os 3 itens citados, o item 12 é considerado o mais difícil por ser seu índice de dificuldade o maior de todos os itens ($b_{12} = 1.16$), com uma taxa de acerto de apenas 23.70%. Além disso o item discrimina bem os indivíduos ($b_{12,2} = 1.3$). O item 9 é parecido com o item 29, porém o item 9 discrimina melhor os indivíduos, com parâmetro $a_9 = 1.03$. Ambos também apresentam o parâmetro de dificuldade bem parecidos e podem ser consideradas questões difíceis com taxas de acertos

de apenas 31.55% e 17.13%, respectivamente.

Itens	Categoria -1 (%)	Categoria 0 (%)	Categoria 1 (%)	a_i	b_{i1}	b_{i2}
Item 1	23,01	35,84	41,15	0,89	-0,69	0,55
Item 2	33,99	33,28	32,73	1,04	-0,75	0,90
Item 3	10,83	54,11	35,07	1,35	0,25	0,69
Item 4	12,36	65,80	21,83	1,69	0,65	1,16
Item 5	15,64	54,59	29,77	1,32	0,27	0,93
Item 6	39,03	19,44	41,52	0,72	-2,14	0,59
Item 7	23,09	38,81	38,09	1,14	-0,44	0,60
Item 8	34,43	21,07	44,50	0,74	-1,93	0,40
Item 9	30,14	38,30	31,55	1,03	-0,50	0,97
Item 10	25,88	50,58	23,53	1,52	0,09	1,13
Item 11	12,26	56,29	31,44	1,44	0,32	0,81
Item 12	35,23	41,07	23,70	1,16	-0,33	1,30
Item 13	27,52	42,31	30,17	1,40	-0,23	0,88
Item 14	29,84	43,69	26,47	1,14	-0,20	1,18
Item 15	14,93	69,47	15,60	2,02	0,71	1,34
Item 16	27,23	45,46	27,31	1,36	-0,11	1,02
Item 17	23,51	37,97	38,52	1,17	-0,45	0,59
Item 18	25,54	44,47	29,99	1,28	-0,15	0,93
Item 19	19,84	60,25	19,91	1,51	0,46	1,32
Item 20	19,04	57,16	23,80	1,54	0,32	1,10
Item 21	25,18	40,04	34,78	1,09	-0,40	0,77
Item 22	33,97	38,81	27,21	0,76	-0,62	1,49
Item 23	21,39	45,66	32,95	0,84	-0,17	1,04
Item 24	25,50	47,21	27,30	1,05	-0,07	1,18
Item 25	22,37	49,96	27,67	1,05	0,05	1,15
Item 26	25,43	23,33	51,24	0,72	-1,76	0,00
Item 27	18,27	31,17	50,57	0,69	-1,19	0,04
Item 28	15,38	27,36	57,26	0,69	-1,49	-0,40
Item 29	34,64	48,23	17,13	0,99	-0,03	1,90
Item 30	14,28	45,12	40,60	1,08	-0,16	0,50
Item 31	34,40	42,60	23,00	1,23	-0,26	1,29
Item 32	23,31	48,52	28,17	1,21	0,01	1,05
Item 33	17,27	51,74	30,99	1,04	0,15	0,99

Tabela 5: Parâmetros dos itens para o modelo MRG

A Figura 16 apresenta a CCI do item 6 (Tipo A). Neste gráfico é possível observar que indivíduos com habilidade até aproximadamente -1.6 têm maior probabilidade de errar a questão. Já indivíduos com habilidade entre -1.7 até, aproximadamente, 0.2 , têm maior probabilidade de não responder a questão. E por fim, indivíduos com habilidade acima de 0.6 , aproximadamente, têm maior probabilidade de acertar a questão. Percebe-se que o item de fato tem boa discriminação, possuindo as curvas bem espaçadas entre si, fazendo com que seja possível visualizar a diferença entre respostas dos indivíduos de acordo com a sua respectiva proficiência.

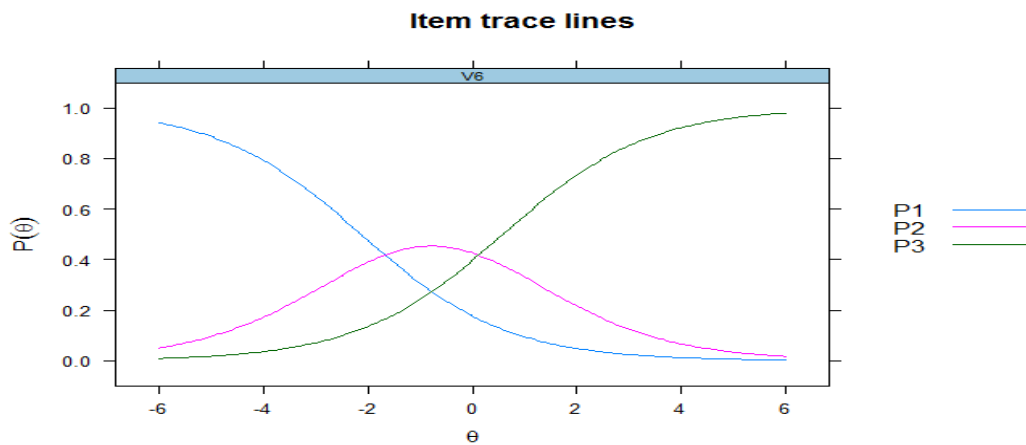


Figura 16: Curva característica do item 6 (Tipo A)

A figura 17 apresenta a CCI para os itens do Tipo C (9,12 e 29). Analisando os gráficos, podemos observar que o item 12 é mais difícil que o item 9 e 29. Isso ocorre pelo fato de que a habilidade necessária para um indivíduo acertar a questão 12 (aproximadamente 0.8) é maior do que a habilidade necessária para um indivíduo acertar as questões 9 e 29 (aproximadamente 0.2 e 0.6 , respectivamente) e também o fato de o ponto de inflexão (cor verde) do item 12 estar mais à direita, indicando ser um item com maior nível de dificuldade.

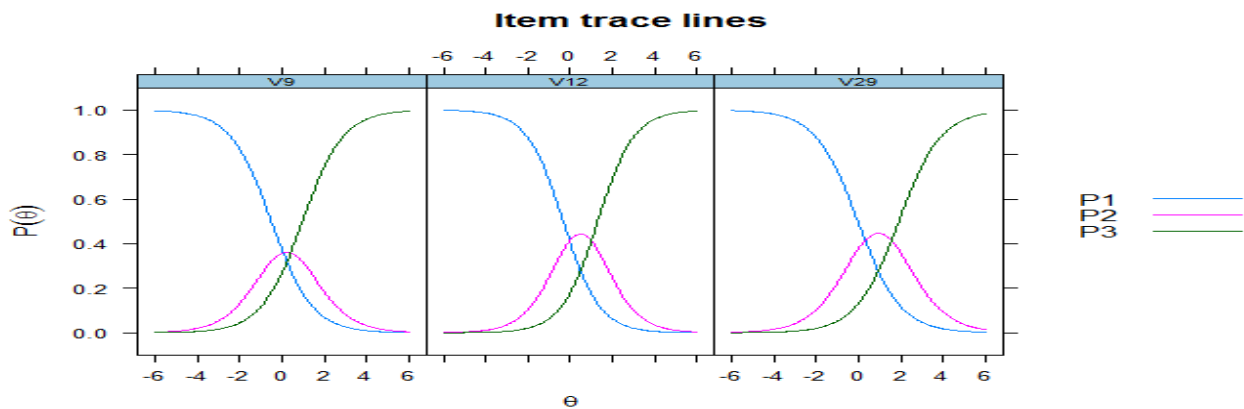


Figura 17: Curva característica dos itens 9, 12 e 29 (Tipo C)

A figura 18 representa a função de informação do teste usando o modelo de resposta gradual. Através desse gráfico podemos observar que o instrumento de medida tem maior informação para os valores da habilidade compreendidos entre aproximadamente -1.5 e 2.5 . Ou seja, isso indica que a prova de matemática é mais adequada para avaliar habilidades que estão dentro desse intervalo. A título de comparação, o MRG abrangeu um intervalo maior de habilidade do que o modelo dicotômico de 3 parâmetros, indicando ser também um modelo adequado.

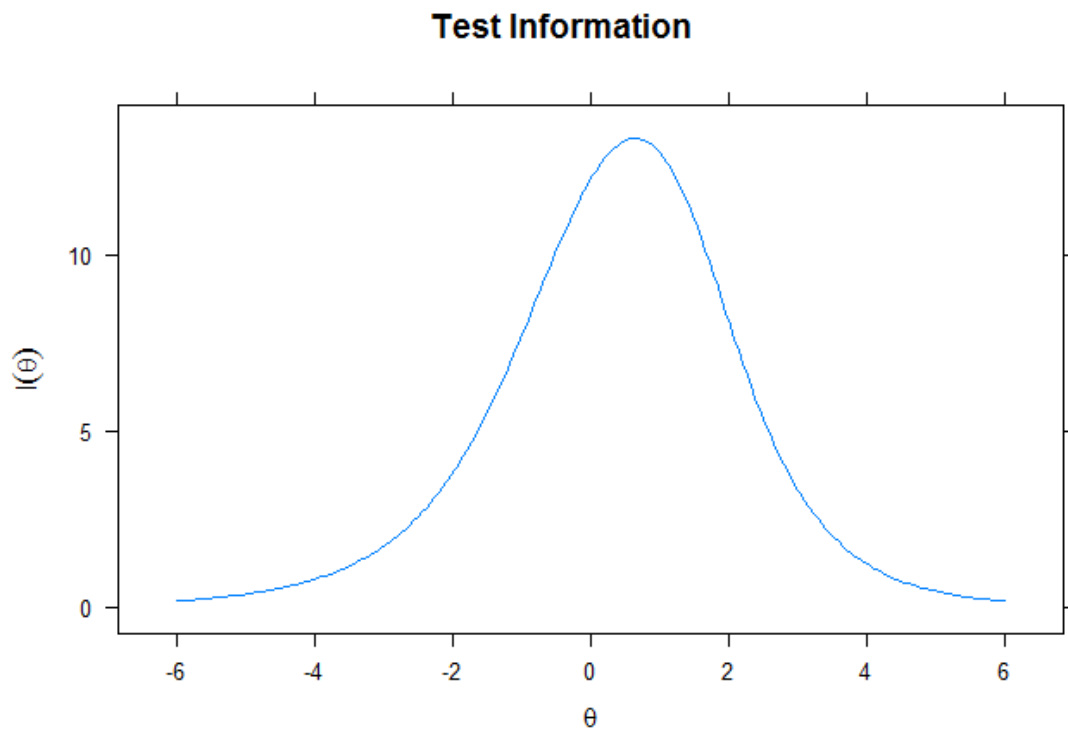


Figura 18: Função de Informação do Teste (FIT)

5.4 Comparação entre as notas das provas corrigidas pelo Método Convencional, pelo Modelo Dicotômico de 3 parâmetros e pelo MRG

Conforme citado anteriormente, a correção convencional padronizada das provas do vestibular da UnB elaboradas pelo CEBRASPE é feita através da soma das pontuações adquiridas pelos alunos ao responderem as questões, subtraindo a média e dividindo pelo desvio padrão de todas as notas. Vale lembrar que o tipo de banco de dados estudado aqui se baseia na pontuação dos alunos somente para as questões do tipo A e C para a prova de matemática, a qual tem as seguintes características: se o aluno acertar a questão, ele ganha 1 ponto; se o aluno não responder a questão, ele não recebe pontuação nenhuma; e por fim, se ele errar a questão, ele recebe a pontuação -1.

Fizemos a comparação entre o método de correção convencional padronizado e o modelo de resposta gradual. O modelo de resposta gradual distribui a pontuação das categorias da seguinte forma: se o aluno acertar a questão, ele ganha 3 pontos; se o aluno não responder a questão, ele ganha 2 pontos; e por fim, se ele errar a questão, ele recebe a pontuação 1.

Também fizemos a comparação entre os dois modelos estudados, o modelo dicotômico de 3 parâmetros e o modelo de resposta gradual. Para o modelo de dicotômico a pontuação foi distribuída da seguinte forma: se o aluno acertar a questão, ele ganha 1 ponto; se o aluno não responder a questão ou se ele errar a questão, ele recebe a pontuação -1.

De acordo com a tabela 6, as notas obtidas pelo modelo de resposta gradual apresentaram valores dos quartis bem próximos aos corrigidos pelo método convencional padronizado. A nota mínima e máxima obtida pelo MRG foi de -3.1835 e 4.5406 , enquanto pelo método convencional padronizado a mínima foi de -3.4602 e a máxima 4.4775 , ou seja, o modelo de resposta gradual apresentou coerência quando comparado ao método convencional. Já o modelo dicotômico de 3 parâmetros teve os extremos mais distantes do que o esperado.

Escores	Mínimo	1° Q	Mediana	Média	3° Q	Máximo
Notas Conv. Padronizadas	-3.4602	-0.6342	-0.1355	0	0.5295	4.4775
Notas Modelo Dicotômico	-1.8519	-0.7992	-0.0974	0	0.6044	3.7626
Notas MRG	-3.1835	-0.7118	-0.0938	0	0.5240	4.5406

Tabela 6: Tabela dos quartis para as notas calculadas pelos modelos propostos.

Na Figura 19, são apresentados os histogramas das notas convencionais padronizadas e das notas do MRG. Podemos observar que as notas convencionais padronizadas ficaram concentradas entre -1.5 e 1.5 , e as notas obtidas pelo MRG ficaram mais distribuídas próximas de -1.5 e 1 . De forma geral os gráficos indicam haver bastante semelhança na forma de suas distribuições.

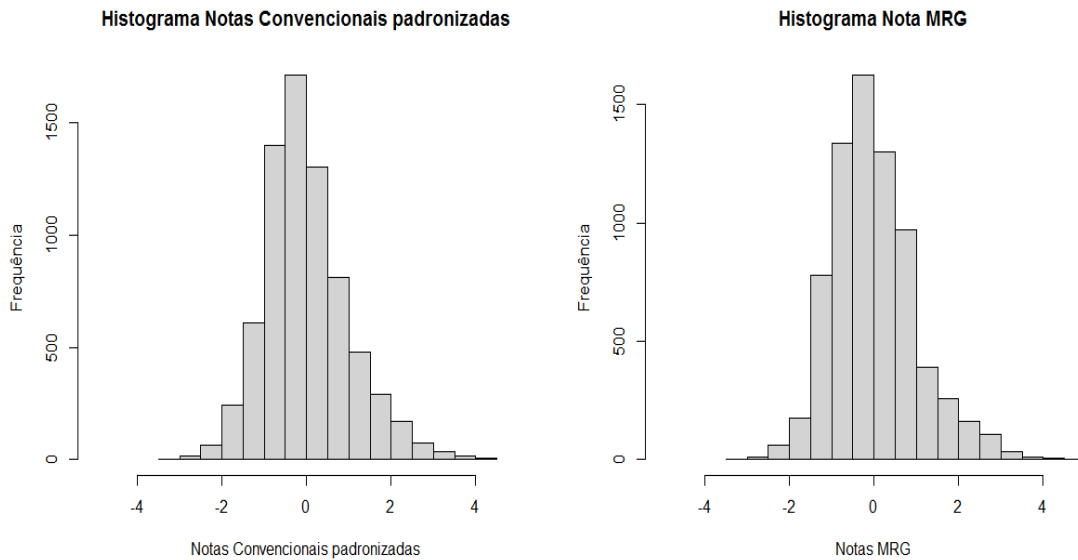


Figura 19: Histograma das Notas Convencionais Padronizadas e das Notas pelo MRG.

Para ter um melhor entendimento, plotamos o gráfico de dispersão entre essas notas, que notavelmente apresenta uma correlação linear alta (0.9882) como vemos na figura 20.

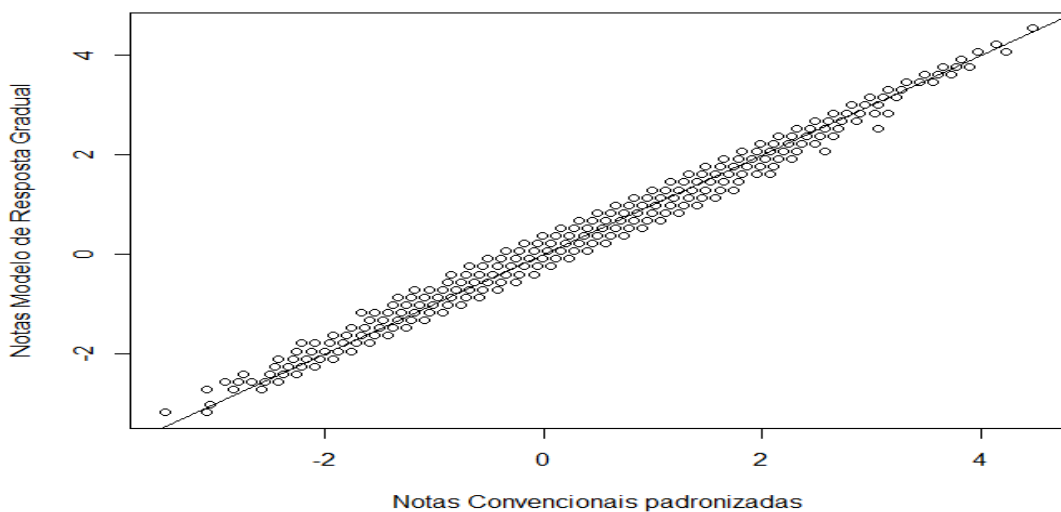


Figura 20: Diagrama de Dispersão entre as Notas Convencionais Padronizadas e as Notas pelo MRG.

Analisando os gráficos da figura 21, percebemos que a maior parte das proficiências estimadas estão no intervalo de -2 e 1 , com algumas proficiências discrepantes em ambos, sendo o modelo de resposta gradual o que tem maior amplitude e maior média, o que já era esperado, sendo que a não resposta considerada como erro aumenta a punição, consequentemente diminuindo a estimativa pelo modelo dicotômico de 3 parâmetros.

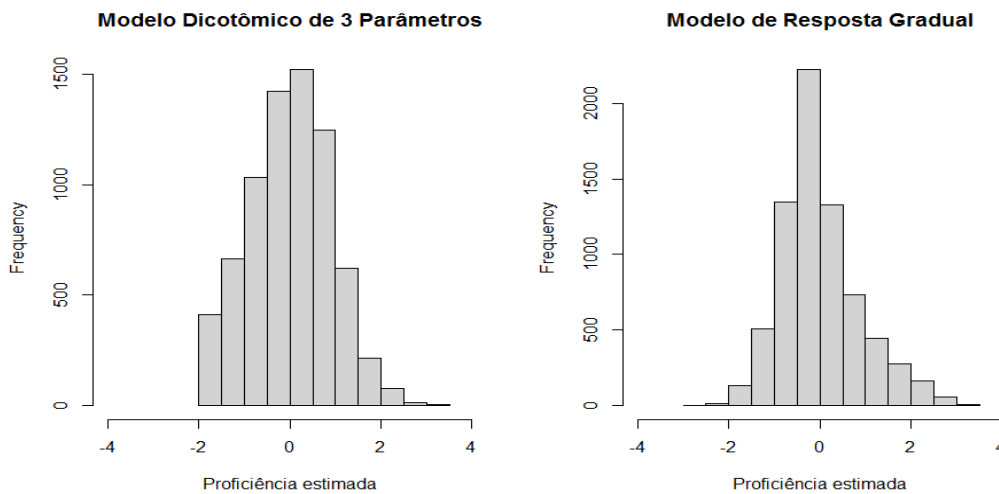


Figura 21: Histogramas das proficiências estimadas pelo modelo dicotômico e pelo MRG.

Na figura 22 temos um gráfico de dispersão entre a diferença das proficiências obtidas pelos 2 modelos (politômico menos o dicotômico) e o escore total obtido pelos indivíduos que realizaram a prova, no qual podemos perceber que há uma tendência de que o modelo politômico favorece os indivíduos com maior escore, ou seja, a presença da categoria não resposta faz com que os escores sejam aumentados. Podemos ver que escores acima de 10, em sua grande maioria, possuem diferença positiva nas estimativas.

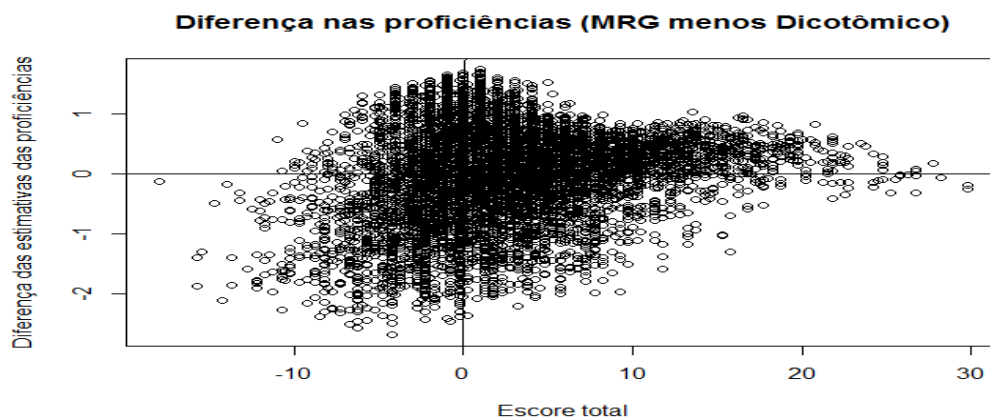


Figura 22: Gráfico de dispersão do escore total versus a diferença nas proficiências estimadas pelos 2 modelos.

Por fim, também fizemos um gráfico de dispersão entre as estimativas dos dois modelos estudados, a partir do qual observamos uma correlação forte (0.6641), e é notável que a estimativa é maior para o modelo de resposta gradual para a maior parte dos respondentes. Mas, nota-se que a correlação foi prejudicada devido à presença de alguns valores distantes na cauda inferior, ocasionando a perda de relação entre os modelos.

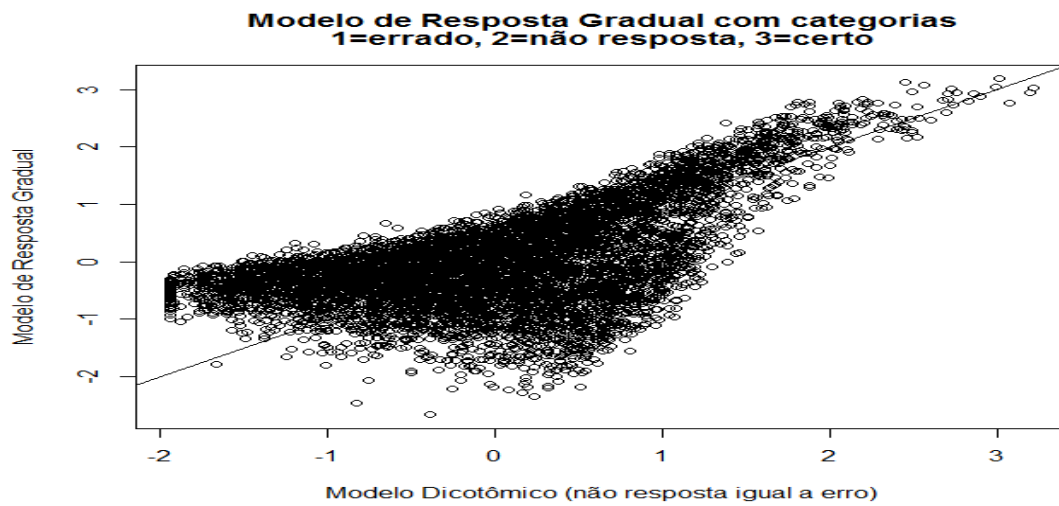


Figura 23: Diagrama de Dispersão entre as proficiências estimadas pelo modelo dicotômico e pelo MRG.

6 Conclusão

Neste trabalho foi realizada uma análise das notas de alunos submetidos a uma prova do vestibular da UnB, elaborada pelo CEBRASPE. O estudo foi somente das questões do tipo A e C para a prova de matemática, constituída por 33 questões e mais de 7 mil respondentes. Essa análise constituiu-se, principalmente, de uma comparação entre as notas desses participantes e pelo método de avaliação proposto, correção pelo modelo de resposta gradual da TRI, pelo modelo dicotômico de 3 parâmetros e a notas obtidas pelo método convencionalmente realizado pelo CEBRASPE.

Com relação à análise dos parâmetros dos itens estimados pelo Modelo de Resposta Gradual (MRG), observamos que os parâmetros dos itens estavam dentro do intervalo adequado e de acordo com a distribuição de respostas de cada categoria e com suas respectivas CCI's. Para o modelo dicotômico, foram observados também itens com parâmetros bem estimados, assim como também suas respectivas CCI's. Entretanto, foram observados valores dos parâmetros de alguns itens fora do intervalo ideal, gerando, assim, estimativas ruins.

Com relação ao cálculo dos escores, chegamos à conclusão de que as notas obtidas pelo método convencional e pelo MRG estão bem próximas entre si, apresentando alto nível de correlação (0.98), demonstrado pela diagrama de dispersão na figura 20. Já quando comparamos as estimativas segundo o MRG e o modelo dicotômico de 3 parâmetros, sendo que nesse último a não resposta foi considerada como erro, percebemos que a correlação foi aproximadamente 0.7, indicando que, quando não aglutinamos a não resposta ao erro, melhora a semelhança dos escores com o método convencional.

Dessa forma, podemos concluir que os resultados finais demonstram que o modelo de resposta gradual apresentou resultados coerentes, bem próximos do método convencional, com a vantagem de considerar graus distintos de discriminação e dificuldade entre os itens.

Referências

- ANDRADE, D. F. d.; TAVARES, H. R.; CUNHA, V. R. *Teoria da resposta ao item: conceitos e aplicações*. [S.l.: s.n.], 2000.
- ANDRADE D.F.; ANJOS, A. *Teoria da Resposta ao Item com uso do R*. [S.l.]: SINAPE, 2012.
- ANDRICH, D. P. *A rating formulation for ordered response categories*. [S.l.: s.n.], 1978.
- BAKER F. B. KIM, S. *Item response theory: parameter estimation techniques*. [S.l.]: Nova York: Marcel Dekker., 2004.
- LORD, F. *Applications of item response theory to practical testing problems*. [S.l.]: Hillsdale: Lawrence Erlbaum Associates, Inc, 1980.
- SAMEJIMA, F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, v. 34, p. 1–97, 1969.
- SANTOS V.L.F.; DOS, G. *Teoria de resposta ao item: uma abordagem generalizada das curvas características dos itens*. [S.l.]: Universidade Federal do Rio de Janeiro., 2009.
- VENDRAMINI C.M.M; DIAS, A. Teoria de resposta ao item na análise de uma prova de estatística em universitários. *Psico-USF*, v. 10, p. 201–210, 2005.