



Universidade de Brasília
Departamento de Estatística

Avaliação da Prova de Biologia do Vestibular da UnB Utilizando Modelos de Teoria de Resposta ao Item

Fernando André Grasser de Souza

Orientador: Prof. Dr. Antonio Eduardo Gomes

Trabalho de Conclusão de Curso

Brasília
Novembro de 2021

Fernando André Grasser de Souza

Avaliação da Prova de Biologia do Vestibular da UnB Utilizando Modelos de Teoria de Resposta ao Item

Relatório apresentado ao Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Orientador: Prof. Dr. Antonio Eduardo Gomes

Brasília - DF, 2021

Resumo

Este trabalho apresenta diferentes formas de avaliar os resultados obtidos a partir das 34 questões de biologia do vestibular da UnB, elaborado pelo CEBRASPE. As formas apresentadas contam com a aplicação de modelos de Teoria de Resposta ao Item, incluindo Modelos de Resposta Gradual (MRG), como uma alternativa ao fator de correção presente na metodologia de pontuação do vestibular, onde uma questão respondida de forma incorreta subtrai pontos do escore final do respondente. Os Modelos de Resposta Gradual utilizados consideram a categoria de “não resposta” como uma das possíveis alternativas. Conclui-se que o Modelo de Resposta Gradual que considera a “não resposta” como a menor categoria apresenta resultados melhores do que modelos onde o erro é a categoria inferior. Além disso, foi observado que respondentes com menor habilidade em ciências biológicas apresentam tendência de não responder aos itens, enquanto indivíduos com habilidade média ou alta apresentam maior probabilidade de responder um item de forma incorreta eventualmente.

Palavras Chave: teoria de resposta ao item, TRI, modelo de resposta gradual, avaliação, vestibular, CEBRASPE, habilidade, proficiência.

Sumário

1 Introdução	8
2 Metodologia	10
2.1 Conjunto de Dados	10
2.2 Ferramentas Computacionais	10
2.3 Modelos de Teoria de Resposta ao Item	10
2.3.1 Modelo Logístico de um parâmetro	11
2.3.2 Modelo Logístico de dois parâmetros	11
2.3.3 Modelo Logístico de três parâmetros	12
2.3.4 Interpretação Gráfica	12
2.3.5 Função de Informação do Item	13
2.3.6 Função de Informação do Teste	14
2.3.7 Modelo de Resposta Gradual	14
2.4 Estimação dos Parâmetros de Modelos de Teoria de Resposta ao Item	16
2.4.1 Estimação dos Parâmetros dos Itens	16
2.4.2 Estimação das Habilidades	16
2.4.3 Estimação dos parâmetros dos itens e das Habilidades	17
3 Análise Descritiva	18
3.1 Análise Descritiva	18
3.1.1 Categorias das Respostas Tipo A	18
3.1.2 Categorias das Respostas Tipo C	19
3.1.3 Escores Convencionais	20
3.1.4 Análise descritiva dos itens	21
3.1.5 Estimativas não paramétricas para as curvas características dos itens	22
4 Resultados	24
4.1 Modelo Dicotomizado	24
4.2 Modelo de Resposta Gradual 1	25
4.3 Modelo de Resposta Gradual 1 sem itens do tipo C	29
4.4 Modelo de Resposta Gradual 2	31

4.5 Curvas de Informação do Teste	35
5 Conclusões.	37
Referências.	38

1 Introdução

Métodos não padronizados e informais foram por muito tempo, no início das civilizações, utilizados para medir habilidades cognitivas de um indivíduo, como raciocínio, memória, ou compreensão sobre determinada área. Porém, este tipo de variável, caracterizada como traço latente, não pode ser observada de forma direta, mas sim por meio de medidas de variáveis suplementares. Sempre procurando um meio de responder este problema, estudiosos e professores entenderam rapidamente a necessidade de se aplicar testes semelhantes para diferentes pessoas. Tornando-se assim o meio mais tradicional na área da educação. O método de teste começou a ser utilizado para diferentes meios, desde avaliar se o aprendizado dos alunos em uma classe ocorreu de forma bem sucedida, até selecionar os indivíduos com maiores habilidades para se preencher vagas em alguma instituição profissional ou de ensino.

Tendo como referência os modelos educacionais europeus, os exames de admissão se tornaram obrigatórios para instituições de ensino superior brasileiras no início do século XX. Entretanto, avaliar grupos de indivíduos submetidos a diferentes formas de ensino e diferentes realidades sociais não é uma tarefa fácil. A partir disso, sempre se procurou fazer destes métodos de seleção os mais confiáveis, justos e padronizados possíveis. Desta forma, esta procura tem se tornado material de estudo e de discussão em diversas áreas de conhecimento. Uma das metodologias mais recentes e bem-sucedidas para esta finalidade é a Teoria de Resposta ao Item (TRI), com seus primeiros modelos surgindo na década de 50.

Atualmente a TRI é reconhecida como um método bastante vantajoso e vem se tornando predominante na avaliação de níveis de habilidade. No Brasil, foi utilizada pela primeira vez em 1995 para analisar os dados do Sistema Nacional de Ensino Básico - SAEB. A partir disso, foi utilizada em outras avaliações de larga escala, como o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - SARESP (ANDRADE; TAVARES; VALLE, 2000). Além disso, o Exame Nacional do Ensino Médio - ENEM, a maior prova do país, utiliza da Teoria de Resposta ao Item para sua metodologia de escores.

Andrade, Tavares e Valle (2000) consideram que o maior avanço da Teoria de Resposta ao item em relação à Teoria Clássica é o fato de permitir a comparação entre indivíduos de populações diferentes que fizeram exames parcialmente distintos, com itens que se encontram na mesma escala. Além disso, a criação de escalas de conhecimento interpretáveis.

Entendendo a importância desta aprimoração constante, esta dissertação estuda a metodologia de escores utilizada nas provas dos vestibulares da Universidade de Brasília

(UnB) realizadas pelo Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos (CEBRASPE).

Como uma possível alternativa para o sistema tradicional de pontuação do vestibular da UnB, este trabalho propõe diferentes formas de avaliar os resultados obtidos a partir das questões de biologia do vestibular da UnB do segundo semestre do ano de 2014, que segue a mesma metodologia empregada em todos os vestibulares até o presente momento. A prova de 2014 será utilizada por conta da disponibilidade dos dados. As questões avaliadas, possuem dois grupos, o primeiro deles é de questões de certo ou errado. Já o segundo, é composto de questões de múltipla escolha. Para o cálculo, serão utilizados modelos de Teoria de Resposta ao Item (TRI) e, então, seu resultado será comparado com as notas convencionais utilizadas pelo CEBRASPE, que não é baseada na Teoria de Resposta ao Item.

2 Metodologia

2.1 Conjunto de Dados

O banco de dados utilizado neste trabalho foi disponibilizado pelo CESBRASPE. Nele estão contidas as respostas das 34 questões de biologia de 7232 estudantes que realizaram a prova de tipo I do vestibular da UnB do segundo semestre de 2014. As questões analisadas possuem duas classificações: tipo A e tipo C. Os itens do tipo A, são aqueles onde o indivíduo marca sua resposta como sendo “certo” ou “errado” para cada uma das afirmações. Já as questões do tipo C, o indivíduo marca apenas uma dentre quatro opções de resposta.

2.2 Ferramentas Computacionais

O programa computacional utilizado para obter os resultados deste trabalho foi o *software* de ambiente de desenvolvimento RStudio, versão 4.0.5. As funções utilizadas para a construção e ajuste dos modelos estão presentes nos pacotes **mirt** e **ltm**.

2.3 Modelos de Teoria de Resposta ao Item

Entendendo a necessidade de se utilizar medidas mais confiáveis na hora de se medir traços latentes, a Teoria Clássica dos Testes (TCT) foi desenvolvida na área da psicométrica para se tornar um dos primeiros instrumentos a trazer resultados mais confiáveis. Esta ferramenta utiliza o escore total de um indivíduo em um teste para medir um traço latente.

Levando em consideração as barreiras encontradas na TCT, a técnica de Teoria de Resposta ao Item (TRI) começou a ser desenhada na década de 50 e tem sido cada vez mais utilizada e aprimorada. A Teoria de Resposta ao Item - TRI é constituída por um conjunto de modelos matemáticos utilizados para representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros desse item e da habilidade do respondente. Esses modelos são desenvolvidos de tal forma que quanto maior a habilidade do indivíduo, maior a probabilidade de acerto no item (ANDRADE; TAVARES; VALLE, 2000).

Os modelos de TRI desenvolvidos até o momento dependem principalmente de três fatores. O primeiro deles é a natureza do item, ou seja, se é dicotômico ou não dicotômico. Em relação aos demais fatores, será considerada apenas uma população

envolvida e um traço latente mensurado para o presente trabalho. Na Seção 2.2.1 a Seção 2.2.3, serão apresentados modelos logísticos, utilizados para itens dicotômicos. Já a Seção 2.2.5 apresenta uma abordagem de modelo para itens não dicotômicos.

2.3.1 Modelo Logístico de um parâmetro

O modelo logístico de um parâmetro (ML1), também conhecido como modelo de Rasch, é um dos primeiros e mais simples da teoria de resposta ao item. Este considera apenas um parâmetro, sendo ele o parâmetro de dificuldade do item. O modelo assume que a probabilidade do j -ésimo indivíduo acertar o i -ésimo item é dada por:

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-(\theta_j - b_i)}}, \quad (2.3.1)$$

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

- U_{ij} assume o valor 1 quando o j -ésimo indivíduo responde de forma correta o i -ésimo item, ou 0 quando responde de forma incorreta.
- θ_j corresponde à habilidade do j -ésimo indivíduo.
- $P(U_{ij} = 1|\theta_j)$ é chamada de Função de Resposta do Item (FRI). É a probabilidade do j -ésimo indivíduo com habilidade θ_j responder o i -ésimo item de forma correta.
- b_i é o parâmetro de dificuldade do i -ésimo item, medido na mesma escala da habilidade do indivíduo.

2.3.2 Modelo Logístico de dois parâmetros

Além de levar o parâmetro de dificuldade do item em consideração, o modelo logístico de dois parâmetros (ML2) também considera a discriminação do item. Em síntese, quanto maior o valor do parâmetro, maior é a capacidade do item de diferenciar respondentes com proficiências diferentes. A probabilidade de acerto do i -ésimo item por parte do j -ésimo respondente é dada por:

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \quad (2.3.2)$$

Além dos parâmetros análogos ao modelo logístico de um parâmetro, tem-se:

- a_i é o parâmetro de discriminação do i -ésimo item.

- D é um fator constante de escala, majoritariamente se atribui o valor 1. Quando se deseja minimizar a diferença máxima entre as funções de distribuição logística e normal, utiliza-se o valor 1,702.

2.3.3 Modelo Logístico de três parâmetros

O modelo logístico de 3 parâmetros (ML3) considera, além da dificuldade e a discriminação do item, a probabilidade de que um indivíduo que não possua a habilidade para responder corretamente um item o acerte por acaso. Este terceiro parâmetro é chamado de "parâmetro de acerto ao acaso". De forma análoga aos modelos anteriores, a probabilidade do j -ésimo indivíduo acertar o i -ésimo item é descrita por:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \quad (2.3.3)$$

Onde, além dos parâmetros mostrados anteriormente, tem-se:

- c_i é o parâmetro de acerto casual do i -ésimo item. Corresponde a probabilidade de um indivíduo com baixa habilidade acertar o item ao acaso.

2.3.4 Interpretação Gráfica

Como foi visto nas equações 2.3.1, 2.3.2 e 2.3.10, a probabilidade de um indivíduo com proficiência θ_j acertar o item i é dado por $P(U_{ij} = 1|\theta_j)$. De forma gráfica, é possível entender o comportamento entre $P(U_{ij} = 1|\theta_j)$ e os demais parâmetros do modelo pela Curva Característica do Item (CCI).

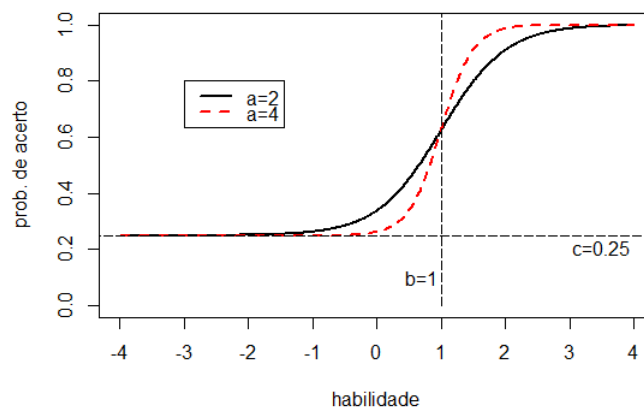


Figura 1: Exemplo de Curvas Características

A Figura 1 mostra o exemplo de CCIs para dois itens distintos. Através delas, é possível ver como a probabilidade de acerto de cada item, representada pelo eixo Y, aumenta de acordo com o parâmetro de habilidade, representada pelo eixo X.

Neste caso é importante observar que ambas as curvas possuem o parâmetro de dificuldade b iguais a 1. Este parâmetro é chamado de parâmetro de localização e representa a habilidade necessária para a probabilidade de acerto do item ser igual a $(1 + c)/2$. Desta forma, quanto maior o valor de b , mais difícil é o item.

O parâmetro c , também é o mesmo para os dois itens representados no gráfico e é igual a 0,25. Ele representa a probabilidade de um indivíduo com baixa habilidade acertar o item. Desta forma, para itens onde exista a possibilidade de se "chutar" dentre as alternativas de resposta, o parâmetro c nunca será efetivamente igual a 0.

Por fim, o parâmetro a representa o poder de discriminação da questão. Quanto menor o valor de a , mais próximas serão as probabilidades de alunos com habilidades muito diferentes acertarem o item. Na Figura 1, a diferença entre as curvas se dá pelos valores de a serem distintos. É possível notar que a curva com parâmetro $a = 2$ é menos íngreme que a curva com $a = 4$. Hipoteticamente, um item com parâmetro a igual a 0 é um item em que a habilidade do respondente não interfere na probabilidade de acerto, sendo assim um item de sorte ou azar, onde a resposta é escolhida ao acaso.

2.3.5 Função de Informação do Item

Segundo Andrade, Tavares e Valle (2000), a função de Informação do Item é uma medida utilizada para se entender para quais valores de habilidade do indivíduo o item terá uma estimativa de proficiência mais precisa. A função de informação do item é dada por:

$$I_i(\theta) = \frac{\left[\frac{d}{d\theta} P_i(\theta)\right]^2}{P_i(\theta)Q_i(\theta)} \quad (2.3.4)$$

onde,

- $I_i(\theta)$ é o parâmetro de discriminação do i -ésimo item.
- $P_i(\theta) = P(U_{ij} = 1 | \theta_j)$
- $Q_i(\theta) = 1 - P_i(\theta)$

Para o modelo logístico de 3 parâmetros, esta equação pode ser escrita da seguinte forma:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[\frac{P_i(\theta) - c_i}{1 - c_i} \right]^2 \quad (2.3.5)$$

Mostrando, assim, como os parâmetros interferem no montante de informação do item. Desta forma, a informação é maior:

- (i) quando b_i está próximo de θ ;
- (ii) quanto maior o valor de a_i ;
- (iii) quando c_i está próximo de 0.

2.3.6 Função de Informação do Teste

A Função Informação do Teste é dada pela soma das informações dos itens, ou seja:

$$I(\theta) = \sum_{i=1}^I I_i(\theta). \quad (2.3.6)$$

Na TRI, outra maneira de se representar a função de informação do teste é por meio do erro padrão de estimação, dado por:

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}}. \quad (2.3.7)$$

2.3.7 Modelo de Resposta Gradual

O modelo de resposta gradual de Samejima (1969), aplicado para itens não dicotômicos, procura obter mais informações sobre as respostas do que apenas se elas foram marcadas de forma correta ou errada. Para isso, o modelo propõe que as categorias de respostas podem ser ordenadas entre si, de forma que a escolha de uma resposta com categoria mais baixa agregue menos para o escore final do respondente do que a escolha de uma resposta de categoria mais alta (SAMEJIMA, 1969).

Este modelo pode ser aplicado também em questões do tipo A do vestibular da UnB. Para isso, é necessário considerar não duas, mas em três categorias: certo, errado e não resposta.

Para a aplicação do modelo, os escores das categorias do i -ésimo item são ordenados de forma crescente, representados por $k = 0, 1, \dots, m_i$ onde o número de categorias do item i é $(m_i + 1)$. Assim, a probabilidade de um indivíduo escolher uma categoria ou outra mais alta do i -ésimo item é expressa ao se incrementar o modelo logístico de 2 parâmetros da seguinte forma:

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}}, \quad (2.3.8)$$

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

- $b_{i,k}$ é o parâmetro de dificuldade da k -ésima categoria do i -ésimo item.

Para itens não dicotômicos, a discriminação de uma categoria de resposta é dada tanto pelo parâmetro de inclinação, quanto da distância das categorias de dificuldade adjacentes. Assim, é necessário a ordenação entre os níveis de dificuldade das categorias de um determinado item i , de acordo com a classificação de seus escores. Por definição, tem-se, então:

$$b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,m_i}.$$

A probabilidade do j -ésimo indivíduo receber um escore k para o i -ésimo item, é dada pela expressão:

$$P_{i,k}(\theta_j) = P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j) \quad (2.3.9)$$

onde,

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

e

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

Assim,

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

e

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

Então, pelas definições de Samejima, temos:

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k+1})}}. \quad (2.3.10)$$

É importante notar que para um item com $(m_i + 1)$ categorias, devem ser estimados m_i valores de dificuldades, além do parâmetro de inclinação do item. Desta forma, para cada item, o número de parâmetros estimados será igual ao número de categorias de respostas.

2.4 Estimação dos Parâmetros de Modelos de Teoria de Resposta ao Item

Para Andrade, Tavares e Valle (2000), a etapa de estimação dos parâmetros dos itens e das habilidades dos respondentes é uma das mais importantes da TRI. A probabilidade de uma resposta correta a um determinado item são dependentes apenas dos parâmetros do item e da habilidade do respondente. Porém, em muitos casos, estes fatores são desconhecidos. Em geral, são conhecidas apenas as respostas dos indivíduos aos itens do teste.

No ponto de vista teórico, é possível dividir o problema em três casos distintos, quando se conhece os parâmetros dos itens, é de interesse do estudo estimar as habilidades dos indivíduos, quando já se conhece as habilidades, é preciso estimar apenas os parâmetros dos itens e, por fim, na maioria dos casos, se deseja estimar simultaneamente as habilidades e os parâmetros dos itens.

2.4.1 Estimação dos Parâmetros dos Itens

Para se estimar unicamente os parâmetros dos itens, é necessário conhecer as habilidades, o que na maioria das vezes não é o caso. Andrade, Tavares e Valle (2000), citam que geralmente o Método da Máxima Verossimilhança é utilizado para a estimação através da aplicação de processos iterativos, como o algoritmo *Newton-Raphson* ou *Scoring de Fisher*.

2.4.2 Estimação das Habilidades

A estimação das habilidades é utilizada quando se deseja submeter indivíduos aos itens já estruturados com o objetivo de classificá-los ou selecioná-los. Segundo Andrade, Tavares e Valle (2000), a estimação é feita pelo Método da Máxima Verossimilhança com a aplicação de processos iterativos.

2.4.3 Estimação dos parâmetros dos itens e das Habilidades

Na maioria dos casos, não se sabe os parâmetros dos itens e nem das habilidades. Segundo Andrade, Tavares e Valle (2000), Bock e Lieberman (1970) desenvolveram o modelo de estimação dos itens pelo Método da Máxima Verossimilhança Marginal. A proposta do método é realizar a estimação em duas etapas: primeiro estimando os parâmetros dos itens e, em seguida, as habilidades. Este método considera uma determinada distribuição para a habilidade dos indivíduos, cuja a função densidade de probabilidade é dada a partir do conjunto de parâmetros associados e integrado em relação a θ . Esta estimação é feita a partir da aplicação de métodos numéricos.

3 Análise Descritiva

3.1 Análise Descritiva

Com o intuito de entender previamente as informações utilizadas no estudo, nesta seção serão apresentados os resultados obtidos a partir da análise descritiva do banco de dados.

Como citado anteriormente, o banco de dados, disponibilizado pelo CEBRASPE, conta com 34 questões de biologia respondidas por 7232 indivíduos que realizaram a prova do tipo 1 do vestibular do segundo semestre de 2014 da Universidade de Brasília. Entre as 34 questões, 30 são classificadas como do tipo A e 4 são do tipo C.

Desta forma, no total serão 245888 respostas, que podem variar entre "corretas", "não respondidas" e "incorretas". A partir disso, o banco de dados foi manipulado para facilitar a análise do ponto de vista de escores, atribuindo as seguintes pontuações, como define o Edital N°1, para cada categoria dos itens do Tipo A: resposta correta (1), não respondida (0) e resposta incorreta (-1). Em relação aos itens do Tipo C, as pontuações atribuídas foram: resposta correta (2), não respondida (0) e resposta incorreta (-0,667), totalizando então 216960 respostas para questões do tipo A e 28928 respostas para questões do tipo C.

3.1.1 Categorias das Respostas Tipo A

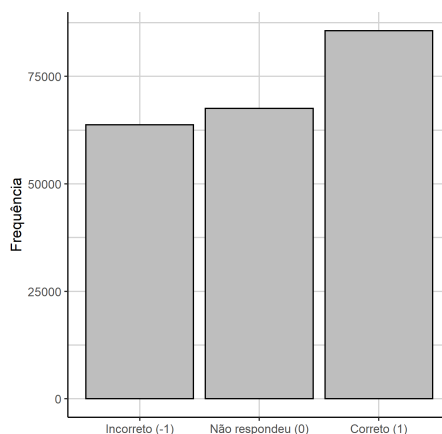
A Tabela 1 mostra a distribuição de frequências para cada categoria de respostas do tipo A. Primeiramente, observa-se que a maior frequência encontrada é a de respostas corretas (39,48%). Por outro lado, a frequência de questões marcadas de forma incorreta é de 29,38%. Por fim, a frequência de não resposta fica entre as duas, com 31,14%. Estas informações mostram a preferência por parte dos indivíduos de se não responder às questões de tipo A de biologia quando não se há certeza do acerto, já que o número de respostas incorretas é menor do que o número de não respostas.

Tabela 1: Frequência de Pontuações

Respostas	Freq. Absoluta	Freq. Relativa
Incorreta	63739	29,38%
Não respondeu	67560	31,14%
Correta	85661	39,48%
Total	216960	100%

A distribuição de respostas do tipo A também está representada graficamente na Figura 2.

Figura 2: Gráfico de barras das frequências de pontos



3.1.2 Categorias das Respostas Tipo C

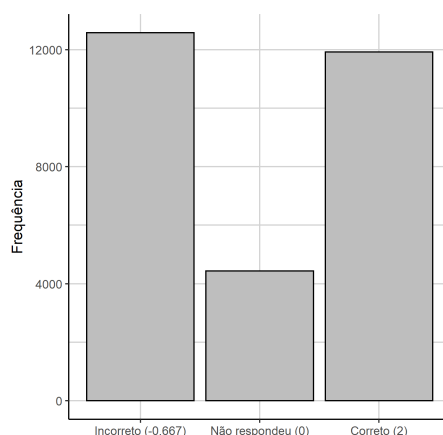
Em relação às respostas para os itens de tipo C, a Tabela 2 mostra a distribuição de frequência de cada categoria. Neste caso, se observa uma frequência bem menor para o caso de não resposta (15,32%). A categoria mais presente é a de respostas incorretas (43,49%), enquanto a de respostas corretas possui uma frequência inferior, porém próxima (41,19%). Esta distribuição de frequências traduz a preferência dos respondentes de se arriscar o acerto para os itens de tipo C. Esta lógica faz sentido, já que o score que se agrega a sua nota ao se acertar o item é aproximadamente três vezes maior que a pontuação de penalidade ao se errar.

Tabela 2: Frequência de Pontuações

Respostas	Freq. Absoluta	Freq. Relativa
Incorreta	12580	43,49%
Não respondeu	4434	15,32%
Correta	11914	41,19%
Total	28928	100%

A distribuição de respostas do tipo C pode ser observada graficamente na Figura 3.

Figura 3: Gráfico de barras das frequências de pontos



3.1.3 Escores Convencionais

As notas convencionais considerando apenas as questões de biologia foram calculadas, para cada respondente. A nota é constituída ao se atribuir as pontuações mostradas anteriormente para cada uma das respostas de cada indivíduo, formando, assim, um escore para cada um dos 7232 respondentes. Caso um candidato acerte todas as questões de biologia, sua nota atribuída seria de 38 pontos. Por outro lado, ao se errar todas as 34 questões de biologia, a nota atribuída seria de -30,66 pontos.

No Quadro 1, é apresentado as medidas descritivas das notas convencionais. O menor escore é de -16,67 pontos. Enquanto isso, nota-se que a maior nota registrada é de 30,00 pontos. Entre o primeiro e o terceiro quartil se encontram 50% de todas as notas, logo, pode-se observar que metade de todos os escores estão entre 0,99 e 9,00 pontos. Observa-se, também, que a nota convencional média é de 5,16 pontos.

Quadro 1: Quadro Resumo de Pontuação Total

Mínimo	-16,67
1° Quartil	0,99
Mediana	4,99
3° Quartil	9,00
Máximo	30,00
Média	5,16
Desvio Padrão	5,95

A Figura 4, apresenta o histograma dos escores convencionais. É possível ver que a curva formada pelas notas é deslocada para a direita, indicando uma concentração

de pontuações positivas para as questões de biologia. Também, observa-se uma maior frequência de notas entre 0 e 10 pontos.

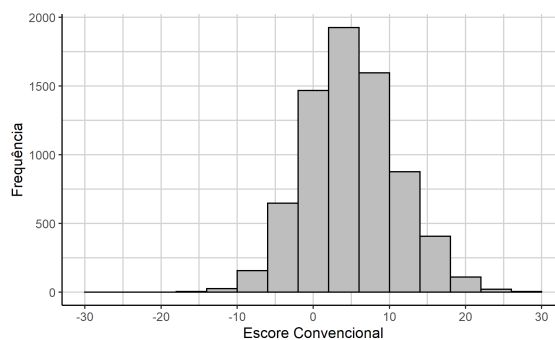


Figura 4: Histograma de Pontuação Total

3.1.4 Análise descritiva dos itens

Figura 5 apresenta em ordem crescente a taxa de acerto de cada um dos 34 itens analisados. Assim, é possível obter uma primeira análise sobre a dificuldade dos itens. É possível observar que o item 20 é o que possui menor taxa de acerto (10,41%). Por outro lado, o item 3 é o item com maior taxa de acerto (70,65%). Além disso, é possível perceber que 27 dos 34 itens possuem uma taxa de acerto inferior a 50%.

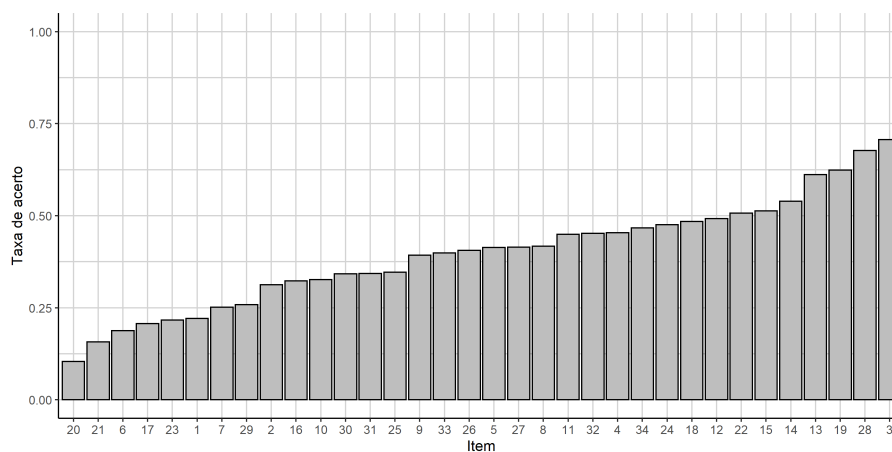


Figura 5: Taxa de acerto

Para a detecção de possíveis itens incoerentes à prova, foi calculado o coeficiente de correlação ponto-bisserial para cada um dos 34 itens. Este coeficiente fornece a correlação linear entre as respostas do indivíduos para o item (0 ou 1) e seus respectivos escores dicotomizados (número de itens com resposta positiva). Dessa forma, é esperado que um item coerente com prova possua um coeficiente de correlação superior a 0. Como

é possível observar na Figura 6, os valores dos coeficientes calculados são todos positivos e menores do que 1, significando assim consistência dos itens com a prova aplicada.

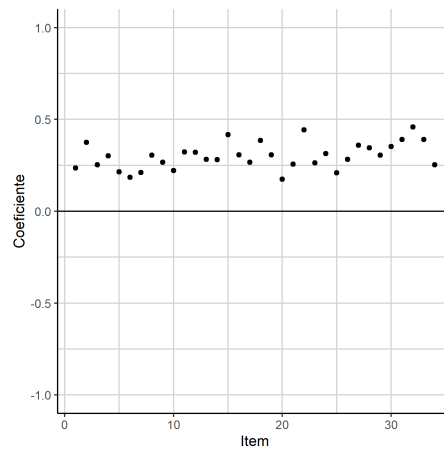


Figura 6: Coeficiente de correlação ponto-biserial

3.1.5 Estimativas não paramétricas para as curvas características dos itens

Com o objetivo de se obter uma primeira estimativa mais simplificada e sem a utilização da estimação de parâmetros, para cada um dos 34 itens, foi observado a distribuição da proporção de acerto para cada escore considerando a pontuação dicotomizada, onde um acerto equivale a 1 ponto e um erro ou não resposta equivale a 0 pontos. Dessa forma, a Figura 21 mostra como é a relação entre estes dois fatores para cada um dos itens.

Analisando a distribuição dos pontos para os três itens com menores taxas de acerto, os itens 20, 21 e 6, respectivamente, é possível observar que as três curvas são as que possuem menor inclinação dentre os itens do estudo. Percebe-se, por exemplo, que a taxa de acerto do item 20 para indivíduos com escore superior a 25 foi abaixo de 50%.

Por outro lado, os três itens com maior taxa de acerto, 3, 28 e 19 apresentam uma inclinação elevada na região com escores mais baixos. No caso do item 3, indivíduos com o escore próximo de 5 obtiveram, em média, taxas de acerto próximas de 50%. Enquanto isso, para o mesmo item, respondentes com escore superior a 25 apresentaram taxas de acerto próximas de 100%, demonstrando, dessa forma, uma grande diferença entre o item 3 e o item 20.

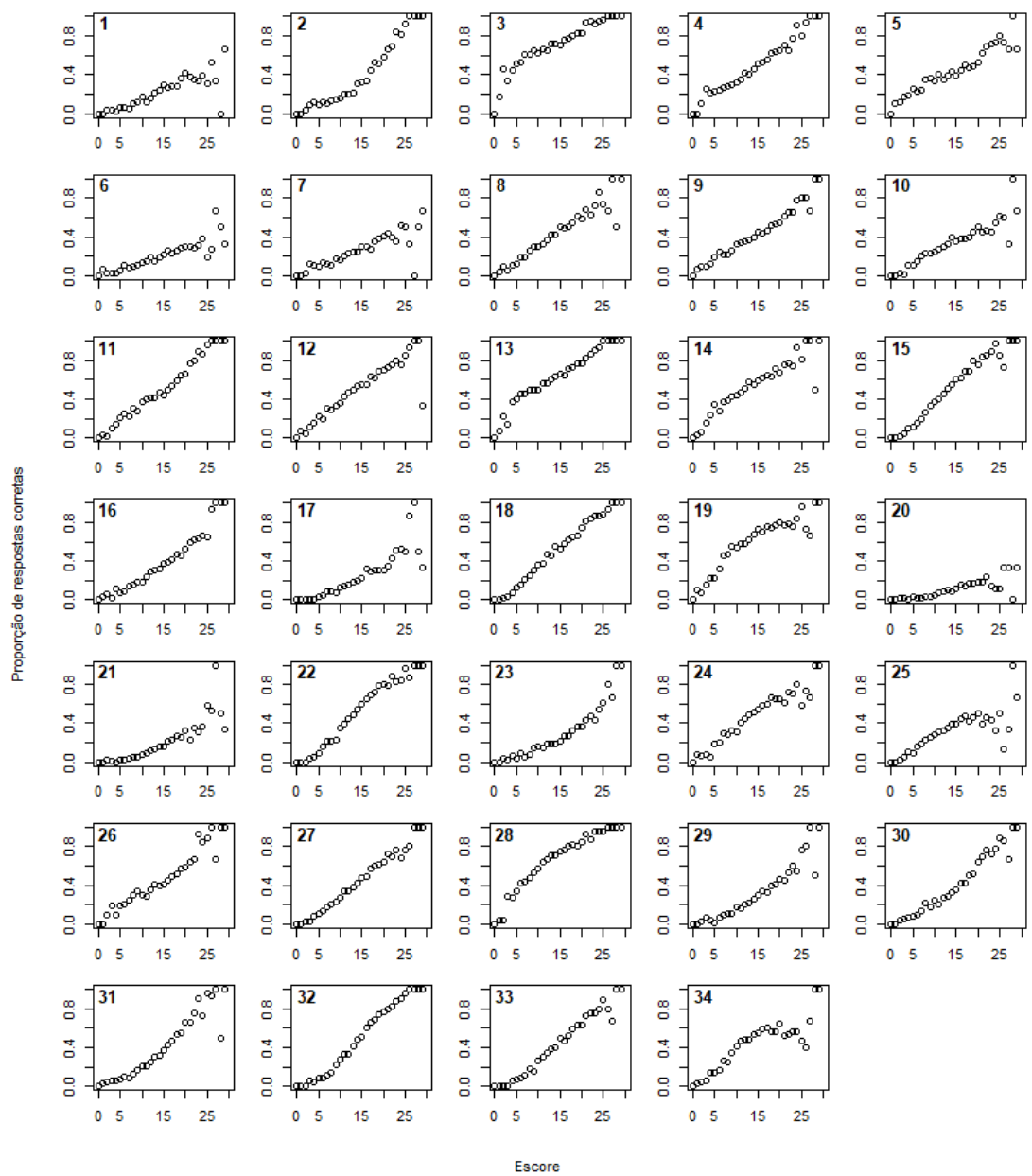


Figura 7: Estimativas não paramétricas para as Curvas Características dos Itens

4 Resultados

4.1 Modelo Dicotomizado

Tendo como intuito estimar as proficiências dos indivíduos e comparar os resultados com aplicações de modelos mais complexos, foi aplicado o modelo logístico de 3 parâmetros. Este modelo considera respostas dicotômicas, ou seja, não há distinção entre itens não respondidos e itens respondidos de forma incorreta. O Quadro 2 apresenta medidas descritivas para os parâmetros dos itens e proficiência estimada a partir do ML3. A partir disso, é possível observar que o valor mínimo para o parâmetro de discriminação (a_i) é superior a zero, o que é esperado para os valores referentes a esse parâmetro. Caso um item apresente valor de discriminação negativo, quanto maior for a proficiência do respondente, menor seria a probabilidade dele acertar o item. Além disso, é possível notar que existem diferentes níveis de dificuldades entre os itens, com valor mínimo para o parâmetro de dificuldade (b_i) sendo -1,191 e o valor máximo chegando a 4,828.

Em relação ao parâmetro de acerto casual (c_i), observa-se que a maioria dos itens apresentou o valor igual a zero, o que é de se esperar, já que o método de avaliação do vestibular penaliza respondentes que marcam uma questão errada pelo fator de correção. Desta forma, a maioria dos indivíduos com baixa habilidade optam por não responder as questões ao invés de tentar acertar ao acaso. Caso não houvesse penalidade, aqueles que não soubessem a resposta responderiam ao acaso e teriam probabilidade de 0,5 de acerto. Por outro lado, o item com o valor máximo para o parâmetro c_i é o item 3, e apresenta valor para o parâmetro igual a 0,665. O alto valor para o parâmetro pode ser correlacionado com a identificação da resposta correta para o item a partir de outras habilidades não relacionadas à biologia, como lógica ou conhecimentos gerais.

Quadro 2: Quadro Resumo de parâmetros estimados

Medida	Parâmetro de Discriminação (a_i)	Parâmetro de Dificuldade (b_i)	Parâmetro de Acerto Casual (c_i)	Proficiência Estimada (θ_i)
Mínimo	0,301	-1,191	0,000	-2,305
1 Quartil	0,477	0,252	0,000	-0,554
Mediana	0,643	1,129	0,000	0,064
3 Quartil	0,938	2,078	0,000	0,607
Máximo	15,776	4,828	0,665	2,67
Média	2,334	1,242	0,073	0,003

O gráfico de dispersão apresentado na Figura 8 mostra a relação entre o parâmetro de dificuldade estimado e a proporção de acerto para cada um dos 34 itens. Percebe-se, desta forma, que no geral há uma tendência de que itens com maior valor de parâmetro

de dificuldade apresentem uma menor proporção de respostas corretas com coeficiente de correlação de Pearson igual a $-0,807$, indicando uma forte correlação negativa.

Observando a Figura 9, é possível entender como é a relação entre as proficiências estimadas dos 7232 respondentes e o escore padronizado calculado a partir do escore convencional. O gráfico evidencia uma tendência linear entre os dois tipos de escores utilizados, o que é verificado a partir do alto valor para o coeficiente de correlação de Pearson, igual a $0,952$. Entretanto, é possível observar uma grande concentração de estimações com valores próximos a 1 para escores padronizados mais elevados, evidenciado pelo formato apresentado na extremidade da faixa formada pelos pontos, podendo significar que o modelo dicotomizado não seja adequado para estimação de proficiências mais elevadas.

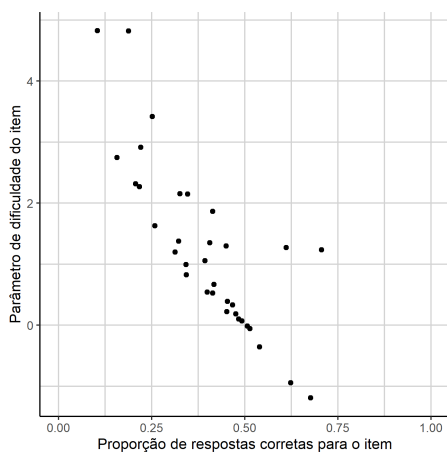


Figura 8: (a)

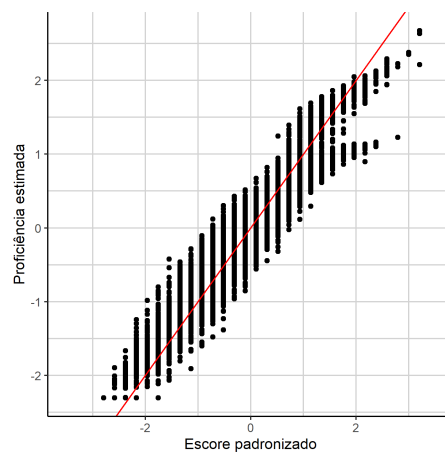


Figura 9: (b)

(a) Parâmetro de dificuldade do item versus proporção de respostas corretas para o modelo logístico de 3 parâmetros (b) Escore padronizado versus proficiência estimada para o modelo logístico de 3 parâmetros

4.2 Modelo de Resposta Gradual 1

A primeira alternativa apresentada para a aplicação de um modelo de resposta gradual leva em consideração as categorias seguindo a disposição 1=errado, 2=não resposta e 3=certo. O Quadro 3 apresenta as medidas resumo para os parâmetros estimados dos itens e a proficiência estimada dos respondentes.

Primeiramente, é possível observar a existência de valores negativos para o parâmetro de discriminação (a_i), o que não é esperado, já que o valor do parâmetro deve ser sempre superior a zero. Um valor negativo para o parâmetro de discriminação indica que a probabilidade de acerto do item diminua com o aumento da habilidade do respondente, o que vai contra a proposta do teste.

É possível observar também pontos destoantes para os parâmetros de dificuldade dos itens (b_{i1}) e (b_{i2}), com os mínimos sendo $-83,861$ para b_{i1} e $-367,419$ para b_{i2} e os

máximos sendo 263,272 e 14,666 respectivamente.

Quadro 3: Quadro Resumo de parâmetros estimados

Medida	Parâmetro de Discriminação (a_i)	Parâmetro de Dificuldade (b_{i1})	Parâmetro de Dificuldade (b_{i2})	Proficiência Estimada (θ_i)
Mínimo	-0,588	-83,861	-367,419	-2,436
1 Quartil	0,003	-6,155	-1,263	-0,547
Mediana	0,227	-2,060	0,134	-0,079
3 Quartil	0,426	0,040	1,225	0,514
Máximo	1,100	263,272	14,666	2,602
Média	0,256	2,342	-14,242	0,000

A Figura 10 apresenta o gráfico de dispersão entre a proficiência estimada considerando o modelo dicotomizado, apresentado anteriormente, e o MRG 1. Observa-se um formato de nuvem, com grande variabilidade e sem um padrão claro que correlacione as proficiências estimadas pelos dois métodos. O valor para o coeficiente de correlação de Pearson é de 0,447, indicando uma correlação moderada.

Comparando as proficiências estimadas pelo MRG com os escores padronizados, mostrado na Figura 11, é possível notar um comportamento similar ao gráfico anterior. Os pontos distribuídos pelo gráfico também formam uma nuvem que dá indícios de que os valores não estão fortemente correlacionados, o que pode ser verificado a partir do valor do coeficiente de correlação de Pearson, igual a 0,496.

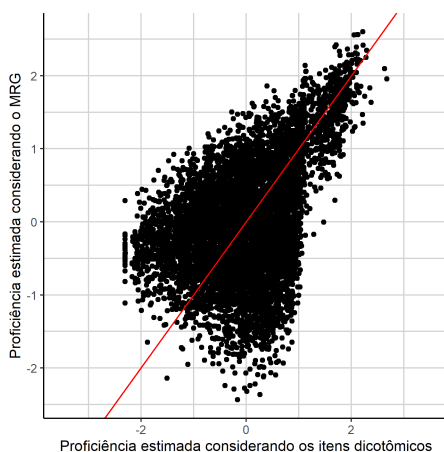


Figura 10: (a)

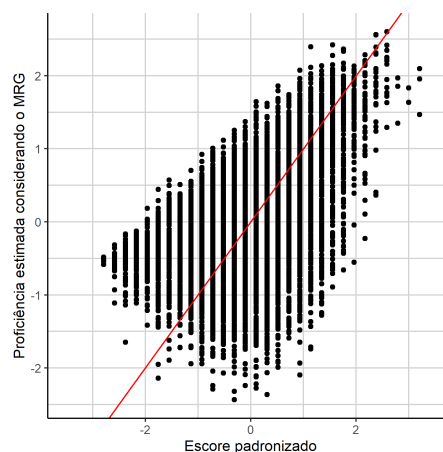


Figura 11: (b)

(a) Proficiência estimada considerando o Modelo Dicotomizado versus proficiência estimada considerando o Modelo de Resposta Gradual 1 (b) Escore padronizado versus proficiência estimada considerando o Modelo de Resposta Gradual 1

Com o objetivo de se observar de forma mais detalhada as estimações para os parâmetros de discriminação, a Figura 12 apresenta o valor dos parâmetros para cada um dos 34 itens. Para o caso do parâmetro de discriminação, é esperado que os valores

estimados sejam superiores a zero, entretanto isso não acontece com alguns itens. Os itens 1, 6, 10, 20, 21, 24, 25, 29 e 34 apresentam valores negativos para este parâmetro, indicando uma possível inconsistência entre o modelo e as questões utilizadas no teste. É importante notar que as questões com menor proporção de acerto, os itens 20, 21 e 6, estão presentes entre os itens com valor de discriminação negativo.

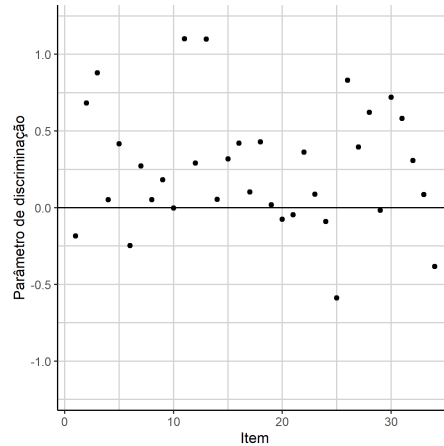


Figura 12: Parâmetro de discriminação considerando o Modelo de Resposta Gradual 1, com categorias 1=errado, 2=não resposta e 3=certo

Levando em consideração as 3 categorias para o Modelo de Resposta Gradual, 1=errado, 2=não resposta, 3=certo, são estimados 2 parâmetros de dificuldade ($b_{i,1}$ e $b_{i,2}$). A Figura 13, apresenta o gráfico para os parâmetros de dificuldade das categorias para os itens do teste. Para uma melhor visualização, a Figura 14 apresenta uma escala menor para o eixo Y, ocultando os pontos referente ao item 10, que apresenta valores discrepantes. Dessa forma, como é possível observar a partir das imagens, para a maioria dos itens, o parâmetro $b_{i,2}$ estimado apresenta um valor superior ao parâmetro estimado $b_{i,1}$. Isto é esperado, já que por definição, $b_{i,1} \leq b_{i,2}$ (SAMEJIMA, 1969). Entretanto, esse não é o caso para os itens que apresentaram estimação de valores negativos para o parâmetro de discriminação, os itens 1, 6, 10, 20, 21, 24, 25, 29 e 34.

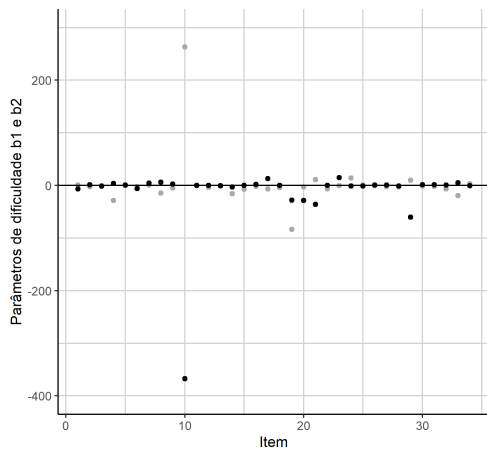


Figura 13: (d)

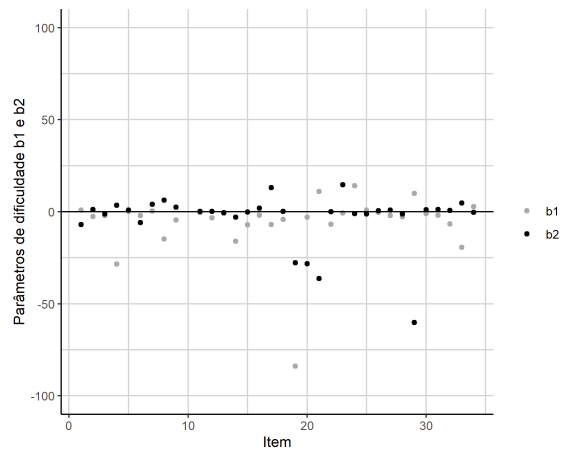


Figura 14: (e)

- (a) Parâmetros de dificuldade considerando o Modelo de Resposta Gradual 1 (b) Parâmetros de dificuldade considerando o Modelo de Resposta Gradual 1 sem a presença do item 10

A Figura 15 apresenta as curvas características das três categorias para cada uma das 34 questões. A curva P1, de cor azul, representa a probabilidade de se errar o item, enquanto a curva P2, que apresenta cor rosa, traduz a probabilidade de não resposta. Por fim, a curva P3, de cor verde, apresenta a probabilidade de acerto.

Utilizando o item 2 como exemplo, observa-se que indivíduos com parâmetros de habilidade inferior a -2 possuem maior probabilidade de errar a questão (P1). Enquanto isso, indivíduos com habilidades entre -2 e 1, apresentam maior probabilidade de não responder ao item (P2). De forma complementar, respondentes com habilidade superior a 1 apresentam maior probabilidade de marcar a alternativa correta para o item (P3).

O item 34 pode também ser utilizado como exemplo. Neste caso, é possível perceber que indivíduos com parâmetro de proficiência inferior a zero apresentam maior probabilidade de se acertar a questão. Analogamente, respondentes com habilidade superior a zero apresentam maior probabilidade de errar o item.

Analisando os gráficos, percebe-se que nos casos dos itens 1, 6, 10, 20, 21, 24, 25, 29 e 34, a probabilidade de acerto decai com o aumento da proficiência dos respondentes. Conforme explicado anteriormente, isso significa que segundo o modelo de resposta gradual utilizado, quanto maior for a proficiência de um respondente, menor será a probabilidade dele acertar este item. Desta forma, é provável que existam limitações no modelo utilizado ou incoerências entre estes itens e o teste realizado. Além disso, é possível notar que alguns itens apresentam uma baixa capacidade de discriminação, como é o caso do item 4, 8, 14, 17, 19, 23 e 33, mesmo que estes não apresentem incoerências no parâmetro de discriminação.

Em contrapartida aos itens citados anteriormente, os itens 2, 3, 5, 7, 11, 12, 13, 15, 16, 18, 22, 26, 27, 28, 30, 31 e 32 apresentam estimações para os parâmetros de

difficuldade e discriminação dentro do esperado para itens apropriados para o modelo. A adequação destes itens pode também ser observada a partir das curvas características, que apresentam coerência entre o parâmetro de habilidade e as probabilidades de erro, não resposta e acerto.

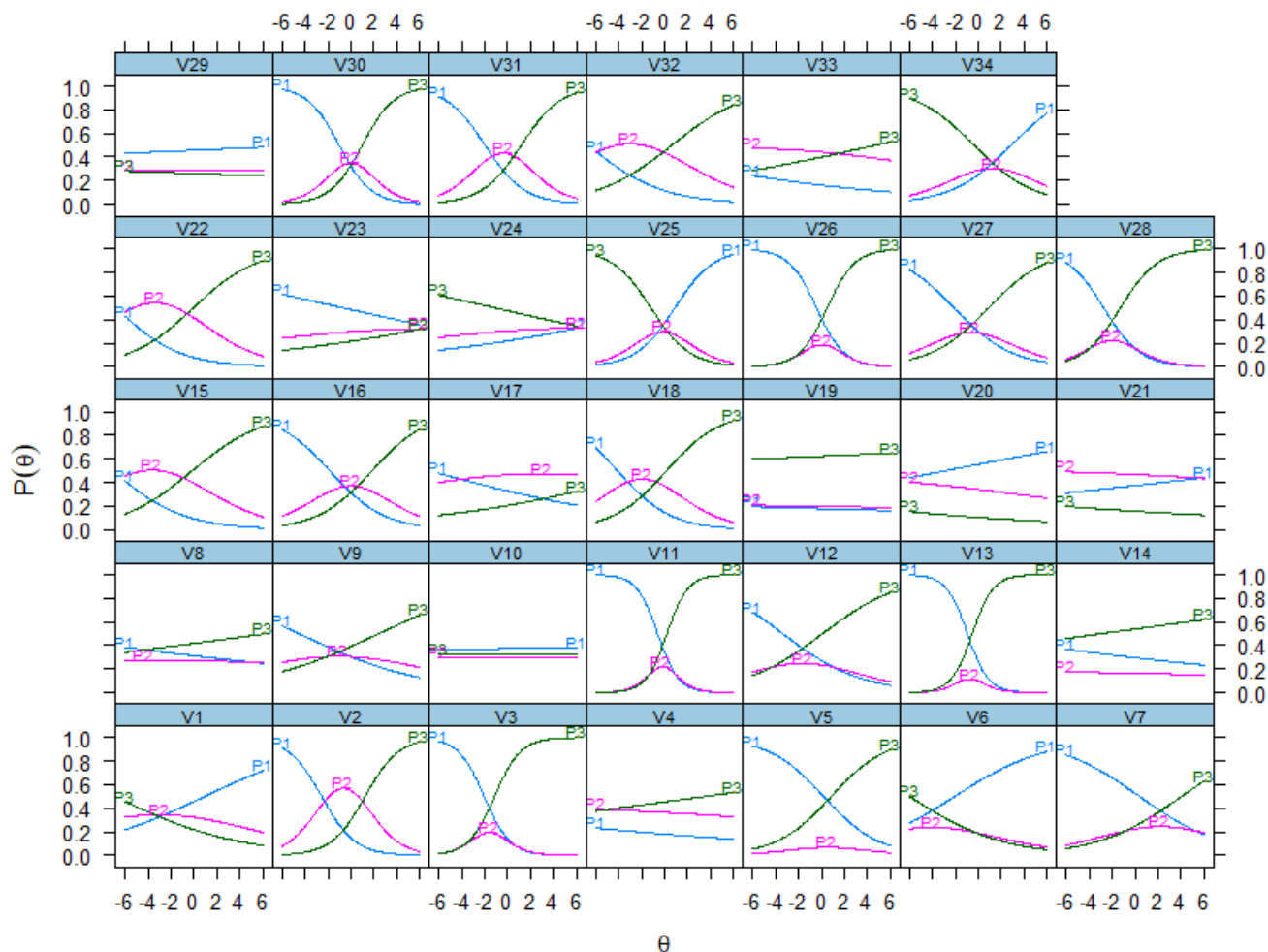


Figura 15: Curvas Características para o Modelo de Resposta Gradual 1, com categorias 1=errado, 2=não resposta e 3=certo

4.3 Modelo de Resposta Gradual 1 sem itens do tipo C

Como foi mostrado na etapa de análise descritiva, existe uma grande diferença entre a tendência de não responder os itens do tipo A e os itens do tipo C. Enquanto 31,14% dos itens do tipo A foram deixados em branco, apenas 15,32% dos itens do tipo C não foram respondidos. Isso pode ser reflexo das pontuações atribuídas a cada tipo de item, já que as questões de tipo C somam 2 pontos para o escore em caso de acerto e só tiram 0,667 pontos em caso de erro. Enquanto isso, as questões do tipo A contribuem

com 1 ponto e também descontam 1 ponto.

Levando em consideração a grande diferença entre as taxas de resposta e as pontuações atribuídas, o modelo de resposta gradual com categorias 1=errado, 2=não resposta e 3=certo foi aplicado apenas para as questões de tipo A com o intuito de se investigar uma maior adequação do modelo ao padrão de resposta e, conseqüentemente, aos itens. Desta forma, os itens 5, 13, 23 e 26 foram retirados do conjunto de itens e outro modelo foi construído.

O Quadro 4 e a Figura 16 apresentam os resultados referentes aos parâmetros e as curvas características. É possível perceber, a partir da análise dos valores e das curvas, que mesmo sem os itens de tipo C, ainda existem valores inconsistentes para os parâmetros estimados de discriminação (a_i) e os parâmetros de dificuldade ($b_{i,1}$ e $b_{i,2}$). Sendo assim, a utilização de um MRG com categorias 1=errado, 2=não resposta e 3=certo não apresenta uma forte adequação aos itens estudados mesmo se desconsiderarmos os itens de tipo C.

Quadro 4: Quadro Resumo de parâmetros estimados

Medida	Parâmetro de Discriminação (a_i)	Parâmetro de Dificuldade ($b_{i,1}$)	Parâmetro de Dificuldade ($b_{i,2}$)	Proficiência Estimada (θ_i)
Mínimo	-0,536	-20,980	-183,745	-2,362
1 Quartil	-0,002	-5,111	-1,351	-0,520
Mediana	0,169,	-2,150	0,050	-0,064
3 Quartil	0,498,	-0,164	1,258	0,475
Máximo	0,972	131,638	23,310	2,476
Média	0,224	1,295	-8,251	0,000

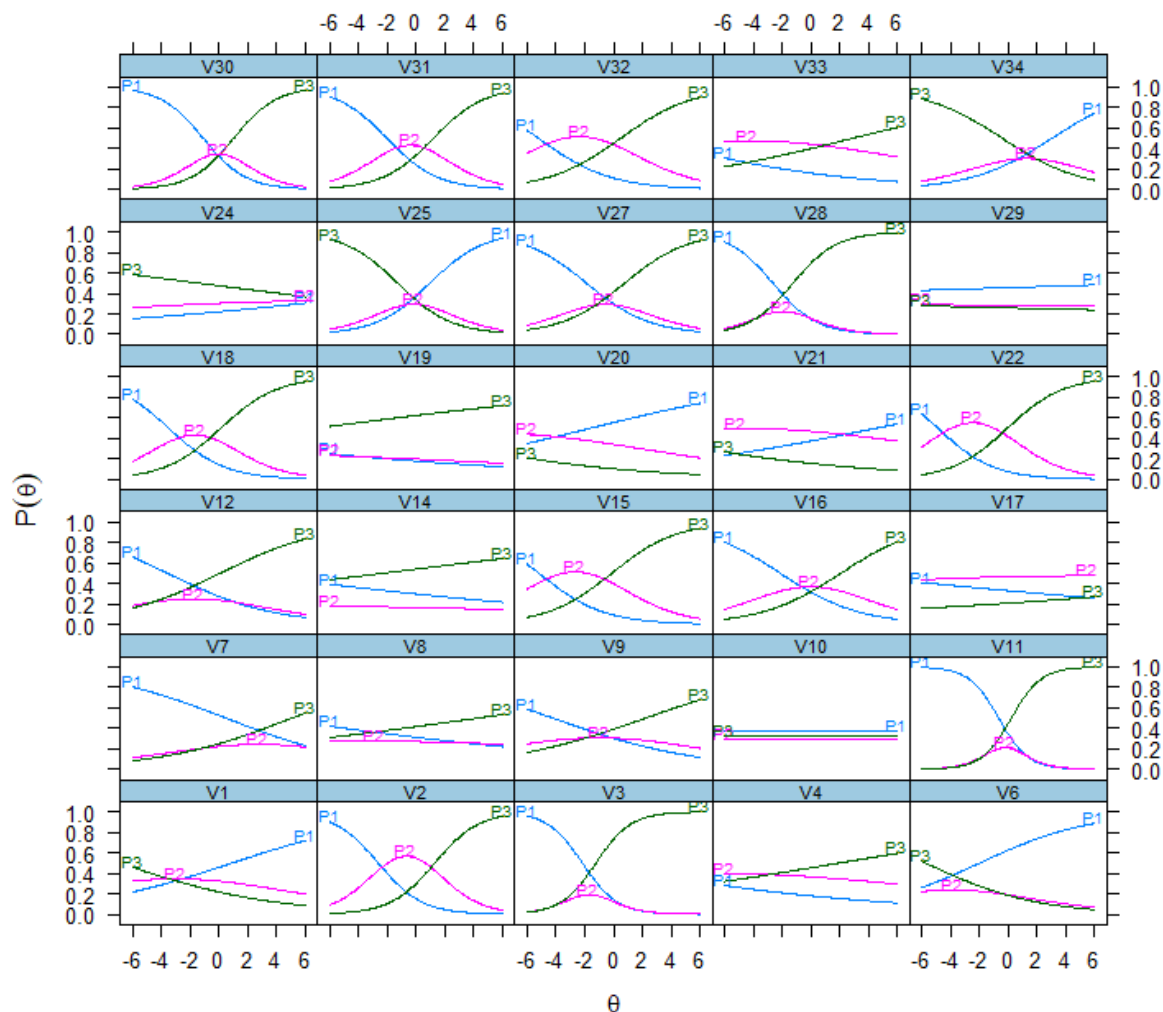


Figura 16: Curvas Características para o Modelo de Resposta Gradual 1, com categorias 1=errado, 2=não resposta e 3=certo

4.4 Modelo de Resposta Gradual 2

O Quadro 5 evidencia as medidas para os valores dos parâmetros dos itens e proficiências estimadas levando em consideração a aplicação do modelo de resposta gradual com categorias dispostas de forma distinta, sendo elas 1=não resposta, 2=errado e 3=certo. Neste caso, é possível observar que não ocorre a estimação de valores negativos para nenhum dos parâmetros de discriminação. Além disso, os valores para os parâmetros de dificuldade apresentam uma variação bem menor do que nos casos anteriores.

Quadro 5: Quadro Resumo de parâmetros estimados

Medida	Parâmetro de Discriminação (a_i)	Parâmetro de Dificuldade (b_{i1})	Parâmetro de Dificuldade (b_{i2})	Proficiência Estimada (θ_i)
Mínimo	0,390	-6,933	-2,229	-3,676
1 Quartil	0,724	-1,930	0,177	-0,536
Mediana	0,885	-0,998	0,638	0,160
3 Quartil	1,124	-0,579	1,127	0,715
Máximo	1,512	0,155	2,165	2,204
Média	0,923	-1,480	0,588	0,000

A Figura 17 apresenta o gráfico de dispersão entre as proficiências estimadas pelo modelo dicotômico e as proficiências estimadas seguindo o MRG proposto. Analisando a imagem, percebe-se que as proficiências calculadas pelos modelos apresentam uma tendência de correlação, diferente do modelo de resposta gradual apresentado na seção anterior. O coeficiente de correlação de Pearson para este caso é de 0,915, significando assim uma forte correlação entre as proficiências obtidas pelos dois métodos.

É importante notar que o formato apresentado pelos pontos no canto superior direito do gráfico evidencia que para uma grande variedade de parâmetros estimados pelo MRG, o parâmetro estimado pelo modelo dicotômico foi próximo de um.

A relação entre as proficiências estimadas pelo MRG e os escores padronizados pode ser observada pelo gráfico de dispersão apresentado na Figura 18. A imagem apresentada traz indícios de que existe uma forte correlação entre as proficiências estimadas pelo MRG e os escores convencionais padronizados. Além disso, é possível perceber que a relação entre os valores não apresenta o problema de valores concentrados em torno de 1, como em casos em que a comparação foi feita com o modelo dicotomizado. O coeficiente de correlação neste caso é de 0,911, comprovando uma correlação forte.

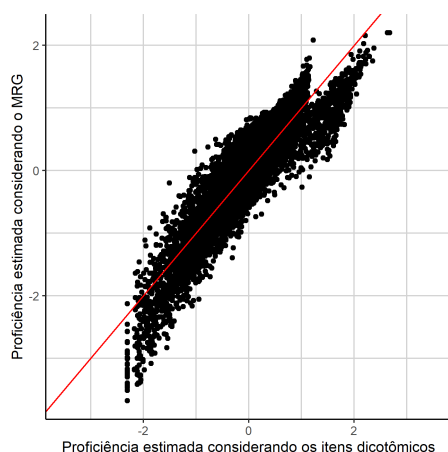


Figura 17: (a)

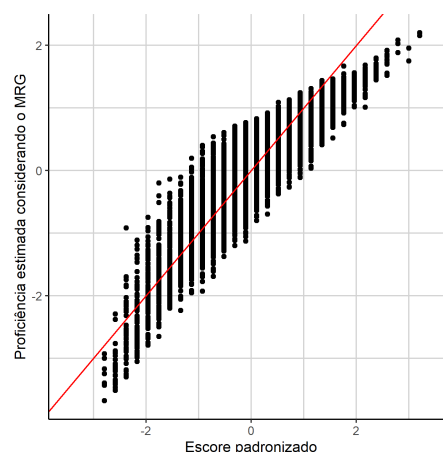


Figura 18: (b)

(a) Proficiência estimada considerando o Modelo Dicotomizado versus proficiência estimada considerando o Modelo de Resposta Gradual 2 (b) Escore padronizado versus proficiência estimada considerando o Modelo de Resposta Gradual 2

Conforme o gráfico apresentado na Figura 19 e as medidas dos parâmetros de discriminação no Quadro 5, o MRG apresenta estimativas negativas para os parâmetros de discriminação. Além disso, é possível perceber que os itens com menor taxa de acerto apresentam altos valores para os parâmetros de discriminação estimados. Desta forma, há indícios de que o MRG com categorias 1=não resposta, 2=errado e 3=certo apresente maior adequabilidade do que o MRG com categorias 1=errado, 2=não resposta e 3=certo.

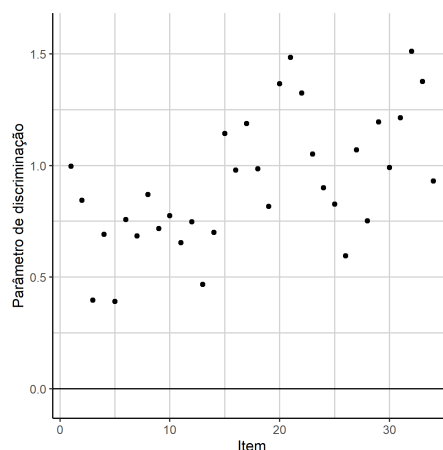


Figura 19: Parâmetros de discriminação considerando o Modelo de Resposta Gradual 2, com categorias 1=não resposta, 2=errado e 3=certo

O gráfico evidenciado pela Figura 20 apresenta os valores para os parâmetros de dificuldade ($b_{i,1}$ e $b_{i,2}$) para os 34 itens considerando o MRG com categorias 1=não resposta, 2=errado, 3=certo. Em contrapartida ao caso do modelo de resposta gradual anterior, o parâmetro de dificuldade $b_{i,2}$ é superior ao parâmetro $b_{i,1}$ para todos os 34

itens.

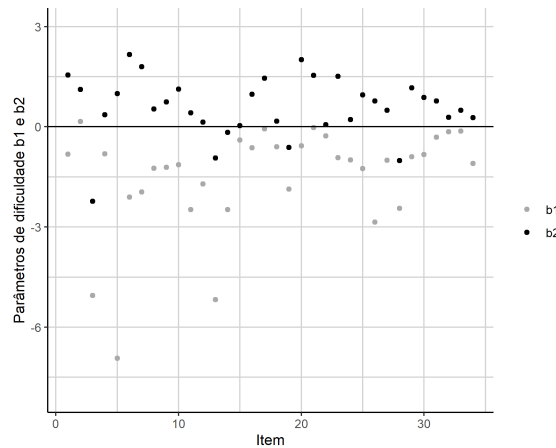


Figura 20: Parâmetros de dificuldade considerando o Modelo de Resposta Gradual 2, com categorias 1=não resposta, 2=errado e 3=certo

Na Figura 21, para os 34 itens de biologia estão apresentadas as curvas características referente ao modelo de resposta gradual considerando as categorias 1=não resposta, 2=errado, 3=certo. Para os gráficos construídos neste caso, a curva de cor azul, P1, representa a probabilidade de não se responder ao item, enquanto a curva de cor rosa, P2, traduz a probabilidade de se errar o item. A curva P3, de cor verde, permanece representando a probabilidade de se acertar o item. É possível perceber que nesta situação as curvas características dos itens são mais definidas e se sobressaem em regiões bem delimitadas.

Observa-se que itens com menores taxas de acerto, como os itens 20, 21, 6, 17 e 23, apresentam regiões formadas pelas curvas P1 (probabilidade de não resposta) e P2 (probabilidade de errar) maiores do que a região formada pela curva P3 (probabilidade de acerto). Em contrapartida, itens com maior chance de acerto, como os itens 3, 28, 19, 13 e 14, possuem uma região abaixo da curva P3 maiores que os demais.

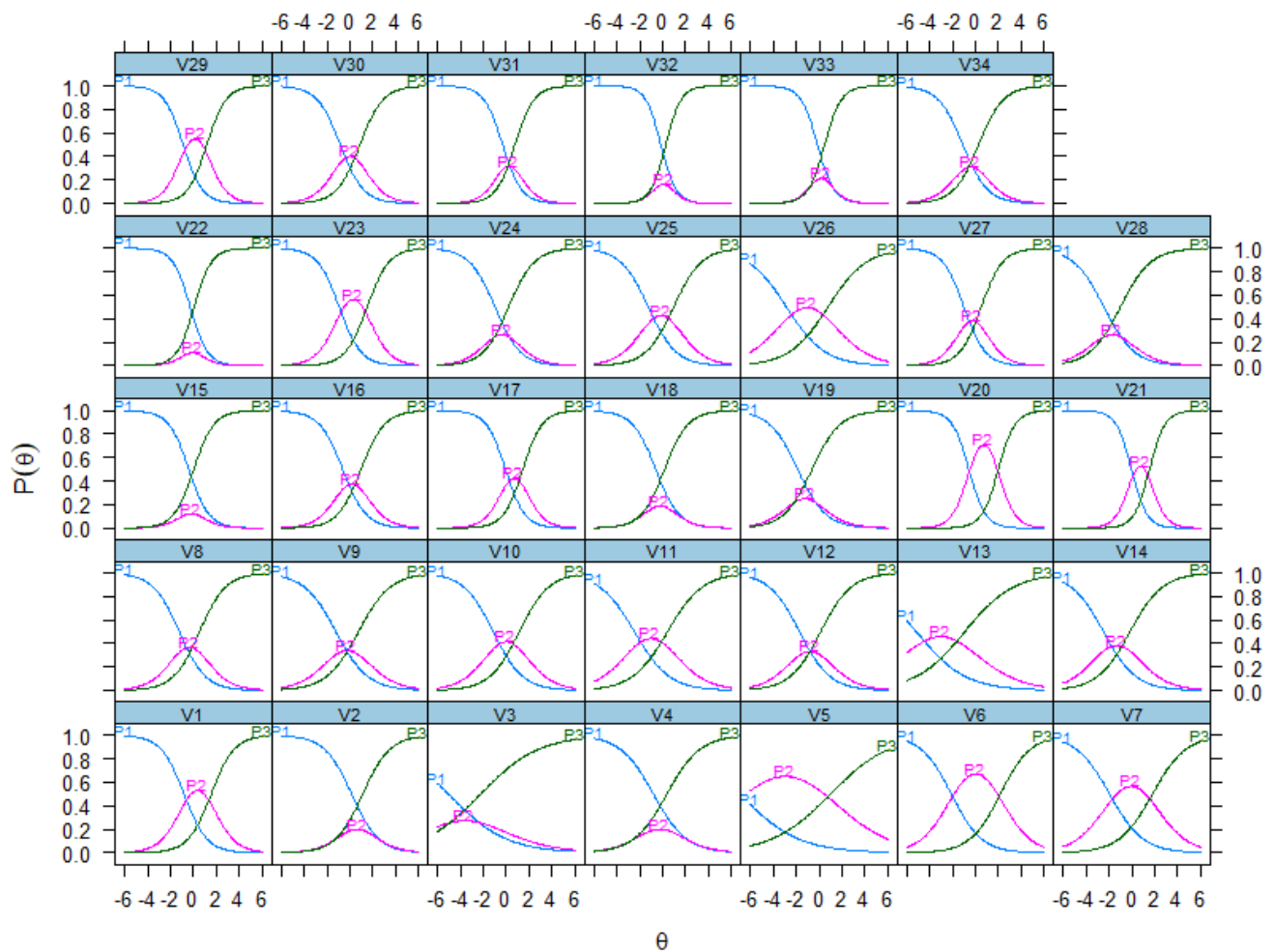


Figura 21: Curvas Características para o Modelo de Resposta Gradual 2, com categorias 1=não resposta, 2=errado e 3=certo

4.5 Curvas de Informação do Teste

Com o intuito de se comparar os três modelos de resposta gradual, a Figura 22 apresenta as curvas de informação do teste obtidas a partir da aplicação dos modelos. A partir da análise do gráfico, é possível perceber que o MRG com categorias 1=não resposta, 2=errado e 3=certo é capaz de gerar maior informação para a avaliação dos indivíduos no geral. Além disso, é possível notar que as curvas formadas pelos MRGs com categorias 1=errado, 2=não resposta e 3=certo apresentam maior informação para valores de proficiência entre -2,5 e 1,5. Porém, o modelo que desconsidera os itens de tipo C não é capaz de fornecer tanta informação quanto o primeiro modelo. Enquanto isso, a outra curva referente ao modelo de categorias 1=não resposta, 2=errado e 3=certo é levemente deslocada para a direita e possui maior informação para valores de habilidade entre -2 e 2.

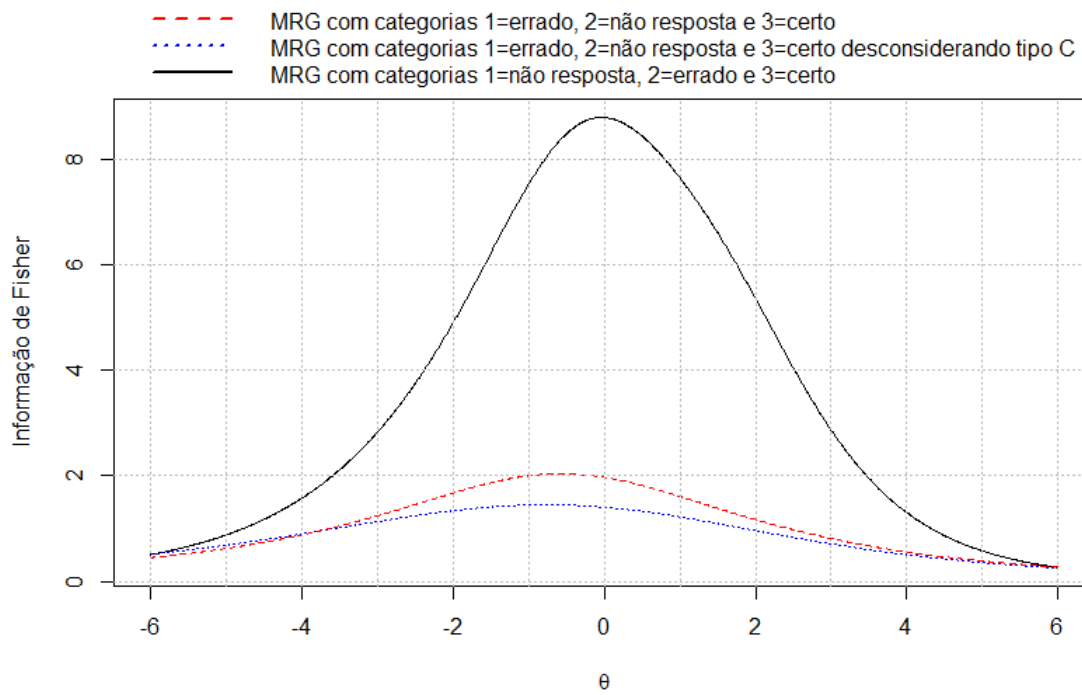


Figura 22: Curvas de Informação do Teste

5 Conclusões

O trabalho proposto teve como objetivo avaliar as notas referentes às questões dos tipos A e C da prova de Biologia do vestibular da UnB do segundo semestre de 2014, realizado pelo CEBRASPE. O estudo é feito a partir da comparação entre os escores obtidos pelo método convencional utilizado pelo CEBRASPE e os modelos de TRI propostos, sendo eles o modelo dicotomizado de três parâmetros e os modelos de resposta gradual com diferentes categorias.

É necessário levar em consideração que o comportamento dos indivíduos ao responderem as questões do vestibular foi condicionado pela metodologia de pontuação evidenciada no Edital N°1 - Vestibular UnB 2014, que aplica uma penalidade para respostas erradas. Este critério busca desestimular estudantes com pouco conhecimento sobre o assunto a tentarem acertar os itens respondendo ao acaso.

Ao se analisar as respostas dos alunos, foi possível perceber a tendência de alunos com menor proficiência não responderem aos itens, enquanto indivíduos com proficiência média e elevada tendem a eventualmente responder incorretamente o item, provavelmente por possuírem algum conhecimento sobre o assunto. Desta forma, os resultados obtidos a partir da análise mostram que dentro os modelos apresentados, o que melhor explica o comportamento das notas é o modelo de resposta gradual com categorias 1="não resposta", 2="errado" e 3="certo". Enquanto o modelo de resposta gradual com categorias 1="errado", 2="não resposta" e 3="certo" apresentou resultados inconsistentes e, inclusive, piores ao modelo dicotomizado.

Referências

- ALMEIDA, R. A. de. **Modelo de resposta gradual para testes com penalização para itens dicotômicos**. *Dissertação (Mestrado em Estatística)-Universidade de Brasília*, Brasília, 2015.
- ANDRADE, D. F. d.; TAVARES, H. R.; VALLE, R. d. C. **Teoria da Resposta ao Item: Conceitos e Aplicações**. [S.l.]: São Paulo: ABE, 2000. 8, 10
- ANJOS, A. dos; ANDRADE, D. F. d. **Teoria da Resposta ao Item com uso do R**. [S.l.]: João Pessoa, 2012.
- BAKER, F. B. **The Basics of Item Response**. [S.l.]: ERIC, 2001.
- BAKER, F. B.; KIM, S.-h. **Item Response Theory, Parameter Estimation Techniques**. Second edition. [S.l.]: Marcel Dekker, 2004.
- BAKER, F. B.; KIM, S.-h. **The Basics of Item Response Theory Using R**. [S.l.]: Springer, 2017.
- BOCK, R. D.; LIEBERMAN, M. **Fitting a response model for n dichotomously scored items**. *Psychometrika*, 1970.
- ENEM - EXAME NACIONAL DO ENSINO MÉDIO. **INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira**. Acessado em 16/04/2021. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>>.
- SAMEJIMA, F. A. Estimation of latent ability using a response pattern of graded scores. *Psychometric Monography*, 1969. 14, 27
- SANTOS, V. L. F. dos. Teoria de resposta ao item: uma abordagem generalizada das curvas características dos itens. **Universidade Federal do Rio de Janeiro**, 2009.
- UNIVERSIDADE DE BRASÍLIA. **Edital Nº 1 - Vestibular UnB 2014**. Acessado em 17/04/2021. Disponível em: <http://www.cespe.unb.br/vestibular/VESTUNB_14_2/>.