

Uma análise comparativa dos perfis dos estados melhor e pior ranqueados no IDEB usando os microdados do ENEM 2022

Leide Ferreira¹, Edwin Monteiro¹, Raimundo Barreto¹

¹Instituto de Computação – Universidade Federal do Amazonas (UFAM)
CEP 69067-005 – Manaus – AM – Brasil

{lmsf, edwin, rbarreto}@icomp.ufam.edu.br

Abstract. *The Exame Nacional do Ensino Médio (ENEM) is recognized as the primary national assessment for admission to higher education. However, its lack of meaningful feedback for students regarding their performance has been a point of concern. This study proposes the utilization of Item Response Theory (IRT) and its models to be applied in the best and worst ranked state in the 2021 IDEB, for the collection and analysis of item difficulties in the exams, estimating the skills required to answer them, in order to provide feedback on the competencies and abilities demanded by these items.*

Resumo. *O Exame Nacional do Ensino Médio (ENEM) é reconhecido como o principal avaliador nacional para a admissão de estudantes no ensino superior. No entanto, sua falta de feedback significativo para os estudantes em relação ao desempenho tem sido um ponto de preocupação. Este estudo propõe a utilização da TRI e seus modelos para ser aplicada no melhor e pior Estado ranqueado no IDEB de 2021 para a coleta e análise das dificuldades dos itens das provas, estimando as habilidades necessárias para respondê-los, a fim de fornecer um feedback sobre as competências e habilidades exigidas por esses itens.*

1. Introdução

A avaliação da qualidade do ensino é uma preocupação constante em sistemas educacionais ao redor do mundo. O Índice de Desenvolvimento da Educação Básica [IDEB 2023] é um indicador sintético que busca medir a qualidade da educação básica no país. Ele é calculado a partir da combinação de dois componentes principais: as taxas de aprovação escolar e as médias de desempenho dos estudantes em avaliações nacionais, um desses é o Exame Nacional do Ensino Médio (ENEM) que tem se consolidado como uma importante ferramenta para mensurar o desempenho dos estudantes e avaliar a qualidade das instituições de ensino.

O ENEM (Exame Nacional do Ensino Médio) utiliza a Teoria de Resposta ao Item (TRI) proposta por [Lord 1980] como método estatístico para a aplicação e correção das provas no Brasil. Essa abordagem difere do sistema de correção tradicional, que considera apenas a contagem de acertos e erros. Este estudo utiliza a TRI como uma ferramenta para analisar e comparar o melhor e pior Estado levantados do IDEB de 2021 utilizando os microdados do ENEM 2022 como base de experimento. Esses microdados fornecem uma ampla gama de informações, incluindo as provas, os gabaritos, as informações sobre os itens, as notas e o questionário respondido pelos inscritos no ENEM [INEP 2022].

Além dos microdados, é utilizada a matriz de referência do ENEM [INEP 2023]. Essa matriz de referência disponibilizada no INEP descreve as competências e habilidades exigidas dos participantes do exame. Ela serve como um guia para a elaboração das questões e permite uma compreensão mais precisa das habilidades avaliadas em cada área do conhecimento.

A combinação dos microdados do ENEM e da matriz de referência proporciona uma base sólida para a análise dos perfis dos estados melhor e pior ranqueados. Essas informações específicas permitem uma investigação mais detalhada sobre as diferenças nos resultados obtidos, considerando tanto as características dos itens quanto as competências e habilidades demandadas dos participantes. Isso significa que é possível examinar quais competências estão mais fortemente associadas ao desempenho dos alunos e identificar possíveis lacunas no ensino e na aprendizagem em diferentes estados. Isso contribui para uma compreensão mais abrangente das diferenças observadas e auxilia na formulação de políticas e estratégias educacionais mais eficazes.

2. Referencial Teórico

2.1. A Teoria de Resposta ao Item (TRI)

A Teoria de Resposta ao Item (TRI) propõe modelos matemáticos para inferir os traços latentes de um indivíduo com base em suas respostas a itens de um teste. Essa abordagem é amplamente aplicada em experimentos que utilizam modelos para itens dicotômicos ou de múltipla escolha, como os modelos logísticos de 1, 2 e 3 parâmetros. Esses modelos permitem a análise dos traços latentes dos indivíduos, como habilidades ou proficiências, por meio da relação entre a probabilidade de uma resposta correta e os parâmetros dos itens. O primeiro modelo é o modelo de 1 parâmetro de Rasch (1PLM) [Rasch 1980], que assume que a probabilidade chamada de Theta (θ) de um indivíduo responder corretamente a um item é determinada pela dificuldade do item b_j . Eq. (1)

Em contraste, o modelo de 2 parâmetro (2PLM), proposto por Birnbaum em 1957, considera dois parâmetros para descrever um item: a dificuldade do item b_j e a discriminação do item a_j . Além da habilidade do indivíduo, a discriminação do item influencia a probabilidade de acerto. A discriminação do item representa a capacidade do item de diferenciar entre os indivíduos com diferentes níveis de habilidade. Eq. (2)

Por sua vez, o modelo de 3 parâmetro (3PLM), desenvolvido por [Birnbaum 1968], difere dos modelos anteriores ao incluir um terceiro parâmetro chamado de chute pseudo (item pseudo-guessing). Enquanto o 1 parâmetro considera apenas a dificuldade do item e o 2 parâmetro inclui a dificuldade e a discriminação do item, o de 3 parâmetro adiciona o parâmetro de chute pseudo c_j . Esse parâmetro refere-se à probabilidade de um indivíduo que não possui habilidade suficiente para responder corretamente ao item acertá-lo por acaso. Esse parâmetro é particularmente relevante quando os itens possuem alternativas de múltipla escolha, onde o indivíduo pode adivinhar a resposta correta. Eq. (3)

$$P_{jk} = P(U_{jk} = 1 | \theta_k) = \frac{1}{1 + e^{-(\theta_k - b_j)}} \quad (1)$$

$$P_{jk} = P(U_{jk} = 1 | \theta_k) = \frac{1}{1 + e^{-D a_j (\theta_k - b_j)}} \quad (2)$$

$$P_{jk} = P(U_{jk} = 1 | \theta_k) = c_j + (1 - c_j) \frac{1}{1 + e^{-D a_j (\theta_k - b_j)}} \quad (3)$$

Para explicação mais abrangente, utilizaremos o modelo de 3PLM que engloba os modelos de 1 e de 2 parâmetros. A Eq. (3) será usada para descrever os parâmetros presentes na fórmula.

Parâmetro a_k : Também conhecido como parâmetro de inclinação, ele controla a inclinação da curva logística. Valores mais altos de a_k indicam uma curva mais íngreme, o que significa que a probabilidade de responder corretamente ao item aumenta mais rapidamente com a habilidade latente.

Parâmetro b_k : É o ponto de inflexão da curva logística, onde a probabilidade de resposta correta está no meio do intervalo entre 0 e 1. Valores mais altos de b_k indicam um deslocamento para a direita da curva, o que significa que a habilidade latente necessária para ter uma probabilidade de resposta correta moderada é maior.

Parâmetro c_k : É o deslocamento vertical da curva logística. Ele controla a posição vertical da curva e representa a probabilidade de uma resposta correta quando a habilidade latente é mínima. Valores mais altos de c_k indicam uma curva mais alta no eixo vertical, o que significa uma maior probabilidade de resposta correta para habilidades latentes mais baixas.

Além da interpretação dos modelos por meio de fórmulas matemáticas, é possível visualizá-los por meio de gráficos utilizando a Curva Característica do Item (CCI) ilustrada na Figura 1. A CCI ilustra a relação entre a probabilidade de resposta correta e o nível de habilidade, fornecendo informações sobre as propriedades dos itens em um teste ou escala de mensuração.

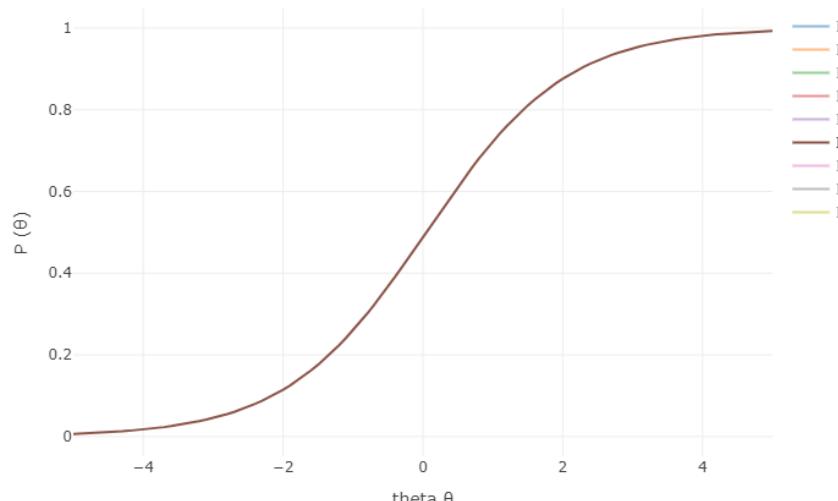


Figura 1. Curva Característica de um Item

Neste estudo, utilizaremos esses modelos para examinar as características dos itens acumulativos, com o objetivo de compreender e comparar as habilidades dos participantes em uma mesma escala de medida. Além disso, utilizaremos a CCI como uma ferramenta de interpretação visual desses itens com tais características, representando-os em um gráfico para uma melhor compreensão.

3. Trabalhos Correlatos

O objetivo do trabalho de [Monteiro et al. 2021] é desenvolver um método de avaliação automática que proporcione feedback, com o intuito de aprimorar o processo de ensino-aprendizagem. Por meio do uso do modelo Rasch, o experimento busca identificar as habilidades dos estudantes e as dificuldades dos itens, a fim de gerar feedback formativo que possa ser direcionado tanto aos alunos quanto aos professores.

Segundo [Araujo et al. 2018] apresentou o uso da TRI numa pesquisa exploratória para investigar a adequação das questões produzidas para a competição Bebras como instrumento de avaliação do Pensamento Computacional (PC). O desafio de avaliar o PC reside na complexidade de mensurar habilidades cognitivas. A Teoria de Resposta ao Item é utilizada como método para estimar os parâmetros dos itens avaliados. Os resultados indicam que as questões analisadas possuem níveis moderados e altos de discriminação, assim como níveis de dificuldade fáceis e moderados. No entanto, o conjunto de questões analisado ainda não atinge um nível de confiabilidade adequado para ser utilizado como instrumento de avaliação. Com base nos estudos anteriores, a principal contribuição desta pesquisa é fornecer um comparativo entre o Estado com o melhor e pior índice no Índice de Desenvolvimento da Educação Básica (IDEB), utilizando a TRI para avaliar possíveis diferenças na aprendizagem entre eles. O uso dessa abordagem permite uma análise mais precisa e detalhada das habilidades e conhecimentos dos estudantes, permitindo identificar quais aspectos podem estar contribuindo para o desempenho educacional superior de determinados Estados em comparação com outros.

4. Metodologia

Para realizar a comparação experimental, utilizou-se uma base de dados contendo os resultados do Exame Nacional do Ensino Médio (ENEM) de 2022. O objetivo deste estudo foi comparar os perfis educacionais de dois estados, o Paraná e o Pará, que se destacaram como os mais bem e mal ranqueados, respectivamente, no referido exame. Essa análise tem como propósito examinar diferenças significativas entre esses dois estados, considerando variáveis relacionadas a matriz de referência do ENEM disponibilizada pelo INEP.

4.1. Características dos dados e pré-processamento

A base de dados inclui os questionários respondidos pelos participantes em cada uma das áreas de conhecimento do ENEM, que são Linguagens, Códigos e suas Tecnologias (LC), Matemática e suas Tecnologias (MT), Ciências da Natureza e suas Tecnologias (CN) e Ciências Humanas e suas Tecnologias (CH). Todos esses dados foram reunidos em um único arquivo, contemplando as informações disponibilizadas sobre os participantes do ENEM, bem como informações gerais sobre os itens das provas nas quatro áreas de conhecimento.

Tabela 1. Amostra de respostas dos alunos da área de Matemática e suas Tecnologias

NU_INSCRIÇÃO	TX_RESPOSTAS_MT	TX_GABARITO_MT	NU_NOTA_MT
210051404132	CBAECCCBCBCBAEBBCECCBBDCECAADDDEACCCBDBCD	EEXEBABDDCECCBCCEDCAEBADEABBDBDEDCCADDCA	584.2
210052030424	EEAEBCBDCBCAABEDCCCBDBDDBACECBBACBCCAEDBCBDBCB	EEXEBABDDCECCBCCEDCAEBADEABBDBDEDCCADDCA	527.6
210052034229	DDABEAECBCDCBDBCDABEDABCDDABDBBEDAABC	ECCCABBDDBDDCEDDCABXEEEBAEDCEDCAEBADEAB	530.5
210053102185	BAEAEAEBBDEADEACAEDBDBAECDBCBBEDEACBCDECBCA	DCCAEBABDDCABECCBCCEADDCEFCDEBADCABBDBDEDCE	571.1
210053703555	EDDCBABABEEBDBDAEADDAEABDEBADDACDBACDEABCD	EEXEBABDDCECCBCCEDCAEBADEABBDBDEDCCADDCA	432.2
210054102731	DCCADDCEAEDBEBBDBDACAEDCECCABCCECDCBEADABEE	DDCABDCCAEDCEBBDBDAEBAEDCAECDDBC	730.1

Para a condução da etapa experimental, é necessário criar matrizes binárias de erros ou acertos. Essas matrizes são geradas por meio da comparação das respostas fornecidas pelos participantes com o gabarito correspondente a cada área do conhecimento, como visto na Tabela 1, foi realizado um pré-processamento desses dados para geração das matrizes binárias. A Tabela 2 apresenta a correção binária (erros ou acertos) para uma amostra de 5 alunos da área Matemática e suas Tecnologias do Estado do Paraná.

Tabela 2. Amostra de matriz de correção binária da área de Matemática e suas Tecnologias do Paraná

matricula	Q39	Q4	Q5	Q27	Q13	Q17
210052175473	1	0	0	0	0	0
210051153661	1	0	0	1	0	0
210052381783	0	1	1	0	0	1
210053851175	0	1	0	1	0	0
210052538084	1	1	0	0	0	0
210053858644	0	0	1	0	0	1
210051089649	1	0	1	0	1	0

4.2. Experimentação utilizando a Base de Dados

Esta seção descreve a etapa de experimentação, cujo objetivo é determinar os parâmetros de dificuldade dos itens, as probabilidades de acerto ao item e as habilidades dos alunos. Primeiramente é necessário realizar a calibração dos dados que é um paradigma introduzido por Allan Birnbaum em 1968 e implementado em (Baker e Kim 2017), essa calibração é um processo essencial na obtenção dos parâmetros mencionados. Por meio desse processo, é possível ajustar os parâmetros dos modelos de acordo com as respostas observadas dos participantes, levando em consideração as características dos itens e as habilidades demonstradas pelos estudantes. Utiliza-se as bibliotecas ERM (Mair P.; Hatzinger, 2007) para o modelo de 1 PLM e a ltm (Latent Trait Models, Rizopoulos, 2006) para os modelos de 2PLM e 3PLM. Ambas as bibliotecas disponíveis na linguagem R oferecem funcionalidades específicas que facilitam a calibração dos dados e a obtenção dos parâmetros desejados. Para explicação da experimentação o modelo de 3PLM será utilizado.

Na biblioteca ltm, a calibração do modelo de três parâmetros é realizada usando o método de Estimação da Máxima Verossimilhança Marginal (Marginal Maximum Likelihood Estimation - MMLE) [Bock and Aitkin 1981]. Esse método busca encontrar os valores dos parâmetros (ak , bk , ck) que maximizam a verossimilhança dos dados observados.

5. Resultados e Discussões

O experimento foi aplicado na base de dados de dois Estados Paraná e Pará levantados no IDEB no ranking nacional de melhor e pior Estado respectivamente. Entre os artefatos gerados estão as dificuldades dos itens, as habilidades estimadas e as curvas características do item. Dada a quantidade de itens por área, a análise dos artefatos leva em consideração uma amostra contendo três itens com cada nível de dificuldade, considerando difícil, média e fácil para fins de comentário.

5.1. Experimentação utilizando modelo Rasch

Nessa sessão é utilizado a base de matemática e suas Tecnologias da prova amarela. Na

Tabela 3. Amostra das habilidades estimadas com o Modelo Rasch para a área de MT

Pontuação MT Pará	habilidade	Pontuação MT Paraná	habilidade
40	2,40	42	3,28
38	1,96	39	2,25
30	0,82	34	1,36
17	-0,60	20	-0,19
12	-1,17	8	-1,67
4	-2,44	3	-2,87

Tabela 3, são apresentadas as pontuações brutas e habilidades correspondentes para os Estados do Paraná e Pará nas áreas de Linguagens e Códigos e Matemática e suas tecnologias. Na coluna de Linguagens e Códigos, a pontuação máxima alcançada foi de 33 pontos, com uma habilidade correspondente de 1,34 para esse Estado. Já em Matemática e suas tecnologias, foi observado o máximo de 42 acertos com habilidade estimada de 3,28, para o Estado do Paraná. No entanto, ao analisar os resultados do Estado do Pará na Tabela 3, observa-se que suas pontuações máximas de acertos foram inferiores em comparação com o Estado do Paraná. Em Linguagens e Códigos, o máximo de acertos foi de 27 pontos, enquanto em Matemática e suas Tecnologias, o máximo de acertos foi de 40 pontos. Essa diferença de desempenho entre os dois Estados evidencia disparidades na performance dos estudantes nas respectivas áreas de conhecimento. Percebe-se na Tabela

Tabela 4. Dificuldade do Item com o Modelo Rasch para a área de MT

Itens - MT	Dificuldade Pará	Dificuldade Paraná
12	1.425	1.909
27	1.21	1.517
39	0.756	0.985
21	0.144	0.105
17	0.066	0.023
19	0.093	-0.011
9	-1.599	-2.11
29	-1.253	-1.517
11	-0.965	1.476

que os itens mais desafiadores de Matemática e suas Tecnologias são os itens 12, 27 e 39, os quais estão representados com uma distância maior dos demais. Isso ocorre devido

ao fato de que esses itens possuem uma dificuldade superior. O item 27, por exemplo, apresenta uma dificuldade de 1,21 para o Estado do Pará, em comparação com o item 11, que possui uma dificuldade mais baixa de -0,965. Analisando os gráficos da Curva Característica do Item (CCI) apresentados na Figura 2 e 3, que ilustra as curvas desses itens, é notável que, para que um participante tenha pelo menos 51% de chance de acertar um item difícil de Matemática e suas Tecnologias, é necessário que ele possua uma habilidade estimada acima de 1,4. Em contrapartida, para o item de menor dificuldade, item 11, a chance de um participante acertar 51% é alcançada com uma habilidade acima de -1. Ao analisar a matriz de competência do ENEM [INEP 2023], verifica-se que um participante que acertou o item 27 demonstra competência na área relacionada à Competência 7. Essa competência está associada à *"Competência de área 7 - Compreender o caráter aleatório e não-determinístico dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas, determinação de amostras e cálculos de probabilidade para interpretar informações de variáveis apresentadas em uma distribuição estatística."*. Essa capacidade permite ao participante interpretar informações de variáveis apresentadas em uma distribuição estatística. Essa conquista indica que o participante possui um conhecimento sólido nessa competência específica, o que reflete sua compreensão e habilidade para lidar com situações relacionadas à aleatoriedade e probabilidade.

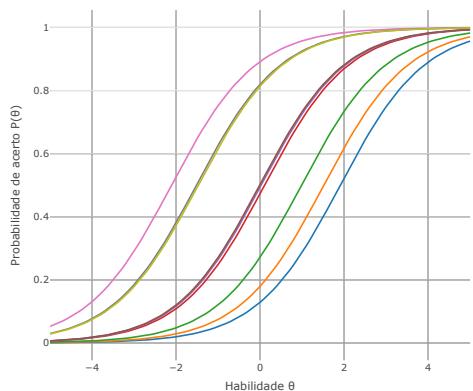


Figura 2. CCI MT Estado do Paraná

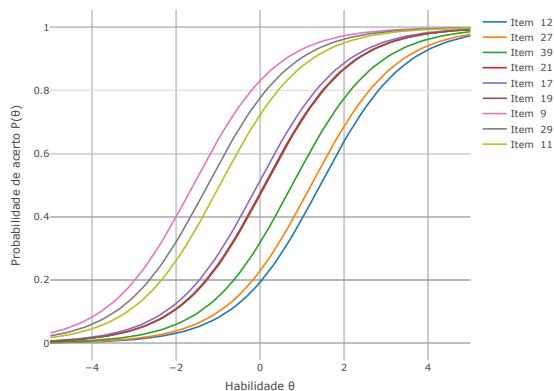


Figura 3. CCI MT Estado do Pará

5.2. Experimentação utilizando modelo de 2 parâmetros

Tabela 5. Dificuldade do Item com o Modelo de 2 parâmetro para a área de LC

Itens - MT	Dificuldade Pará	Discriminação Pará	Dificuldade Paraná	Discriminação Paraná
31	12.93	0.10	2.67	0.45
26	10.51	0.12	7.54	0.17
24	8.75	0.27	7.60	0.33
33	0.82	0.97	0.19	1.23
17	0.75	0.54	0.93	0.54
32	-1.10	1.64	-1.85	1.59
18	-1.05	1.46	-2.06	0.97

Neste experimento, iremos considerar os parâmetros de dificuldade e discriminação do item na base de dados de Linguagens e Códigos da prova de cor amarela. Ao analisar os itens considerados mais fáceis, como o item 32 da prova de Matemática e

Códigos, Tab 5, o mesmo exibe um padrão de probabilidade de acerto de 89,37% visualizando a CCI da Fig 4. Isso significa que 7.888 participantes acertaram essa questão, de um total de 8.826 participantes. Portanto, a habilidade necessária para responder corretamente essa questão está entre -2 e 0 na escala de habilidade (tabela de dificuldade). A competência desse participante está relacionada à Competência de área 3: "*Construir noções de grandezas e medidas para a compreensão da realidade e a solução de problemas do cotidiano*". Além disso, sua habilidade permite "*Avaliar o resultado de uma medição na construção de um argumento consistente*", conforme indicado pela habilidade H13. Agora, a análise é aplicada a uma questão com maior dificuldade, como o item 31, cuja habilidade H23 visa "*Inferir em um texto quais são os objetivos de seu produtor e quem é seu público-alvo, pela análise dos procedimentos argumentativos utilizados*", Figura 4 e 5. Observamos que no Paraná Figura 4, essa questão teve uma probabilidade de acerto de 50% quando a habilidade (θ) do participante é acima de 2, enquanto no Estado do Pará, a probabilidade de acerto acima de 50% da questão ocorre quando a habilidade é acima de 4. Analisando a base de dados, constatamos que 23% dos participantes do Paraná e 20% do Pará acertaram esse item, embora as porcentagens sejam muito próxima, a habilidade varia devido à discriminação do item (tabela de dificuldade e discriminação) gerada pelo modelo de 2PLM, sendo 0,80 para o Pará e 0,29 para o Paraná. Isso significa que no Estado do Pará, com uma discriminação de 0,8, o modelo teve uma maior capacidade de diferenciar os participantes com habilidades distintas, pois quanto maior o valor da discriminação, mais eficaz é a diferenciação. Por outro lado, a base do Paraná apresentou um valor muito baixo de discriminação, tornando-a menos eficaz nesse aspecto.

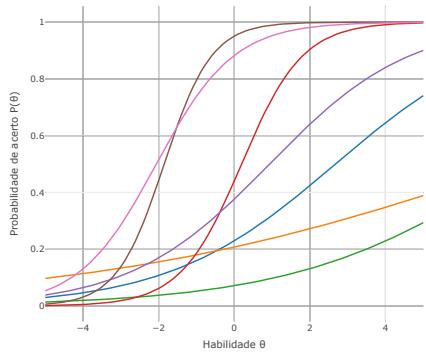


Figura 4. Linguagens e Códigos - Estado do Paraná

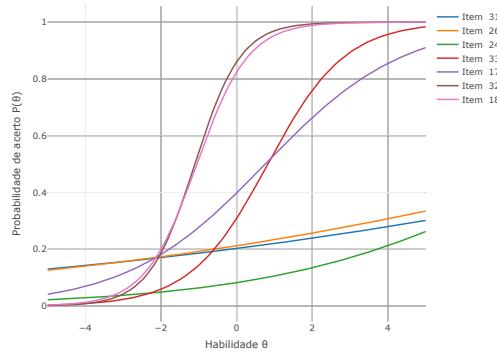


Figura 5. Linguagens e Códigos - Estado do Pará

5.3. Experimentação utilizando modelo de 3 parâmetros

Nesta sessão, iremos analisar os parâmetros de dificuldade, discriminação e chute ao acaso do item Tabela 6, utilizando a base de dados de Ciências Humanas. Ao observar as curvas, podemos notar que além de uma distância consideravelmente maior devido ao parâmetro de chute ao acaso (parâmetro c), as curvas também apresentam um perfil mais íngreme, indicando uma alta capacidade de discriminação entre os respondentes com diferentes níveis de habilidade. Em outras palavras, a curva íngreme sugere que o item é eficiente em diferenciar claramente entre aqueles com alta habilidade e aqueles

Tabela 6. Dificuldade do Item com o Modelo de 3 parâmetro para área de CH

Itens - CH	Dificuldade Pará	Discriminação Pará	chute ao acaso Pará	Dificuldade Paraná	Discriminação Paraná	chute ao acaso Paraná
38	4.01	0.18	0.20	2.21	3.09	0.31
3	3.12	0.32	,2.21	2.37	0.30	0.002
16	3.10	1.58	0.10	2.20	1.67	0.09
29	0.96	0.48	0	0.48	0.44	0.017
39	0.95	0.58	9.35	1.18	0.41	0.002
36	1.01	1.44	0.146	-1.83	1.47	1.26
5	-1.64	0.98	9.81	-2.60	0.78	0.0003

com baixa habilidade. Ao analisarmos o item 16, identificado em verde na Figura 6 a CCI do Estado do Paraná, sua dificuldade é 2.20, discriminação de 1.67 e acerto ao acaso de 0.09, podemos observar que a curva se torna íngreme a partir das habilidades acima de 1. Isso significa que abaixo desse intervalo, a probabilidade de acerto permanece em torno de 20%, o que torna o item desafiador para os participantes com habilidades inferiores. Por outro lado, ao analisarmos o mesmo item 16 na Figura 7 a CCI do Estado do Pará, sua dificuldade é de 3.10 com discriminação 1.58 e acerto ao acaso de 0.10, notamos que a curva só se torna íngreme quando a habilidade está acima de 2. Ao compararmos o total de acertos nos dois estados, verificamos que o item teve 1342 acertos para 8816 participantes no Estado do Paraná, enquanto no Estado do Pará o item registrou apenas 754 acertos para 6286 participantes. Com base na análise das curvas e nos totais de acertos, podemos concluir que esse item foi considerado mais difícil no Estado do Pará em comparação ao Estado do Paraná. Os participantes que acertaram esse item são considerados com a "*Competência de área 1 - Compreender os elementos culturais que constituem as identidades.*" e a habilidade "*H2 - Analisar a produção da memória pelas sociedades humanas.*". Ao analisarmos itens de dificuldade mediana, podemos focar no item 29. Esse item apresenta uma curva mais plana, o que indica uma taxa de mudança constante e gradual ao longo do seu percurso. Observamos uma dificuldade de 0.9 para o Estado do Pará e 0.4 para o Estado do Paraná, com discriminações de 0.48 e 0.44, respectivamente. Esses valores sugerem que tanto no Pará quanto no Paraná, as curvas correspondentes a esses itens não apresentaram variações bruscas nos valores representados.

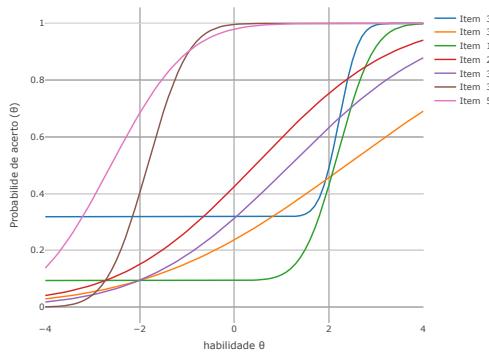


Figura 6. CCI Ciências Humanas - Estado do Paraná

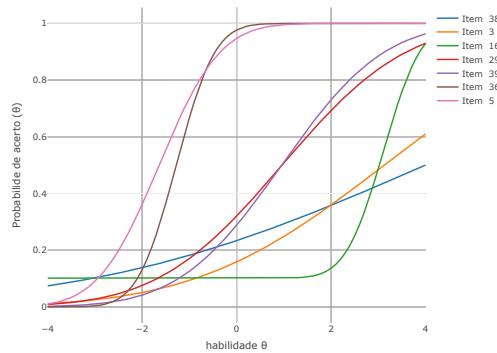


Figura 7. CCI Ciências Humanas - Estado do Pará

5.4. Discriminação Negativa

Nas sessões anteriores, foram apresentados apenas itens com discriminação positiva. No entanto, nesta sessão, citaremos um item considerado de dificuldade fácil, mas com

discriminação negativa. Esse é o caso do item 4 da prova de Ciências Humanas, conforme ilustrado na Figura 8. Itens com discriminação negativa podem ocorrer de duas maneiras. Primeiro, a resposta incorreta a um item de escolha entre duas opções sempre terá um parâmetro de discriminação negativo se a resposta correta tiver um valor positivo. Segundo, em alguns casos, a resposta correta a um item resultará em um parâmetro de discriminação negativo. Isso indica que algo está errado com o item, seja na sua formulação ou devido a algum equívoco prevalente entre os alunos de alto nível de habilidade. Em qualquer caso, isso indica que o item requer atenção [Baker 2017]. Uma discriminação negativa significa que a probabilidade de uma resposta correta diminui à medida que o nível de habilidade aumenta. Essa situação ocorre em itens que têm uma relação inversa com a habilidade medida, ou seja, quanto mais habilidoso o indivíduo, menos provável ele é de responder corretamente ao item. O item 4 precisa ser revisado e ajustado para garantir sua validade e fornecer informações úteis sobre a habilidade do estudante. Portanto, ele não deve ser considerado como um dado de comparação para avaliar a probabilidade de acerto.

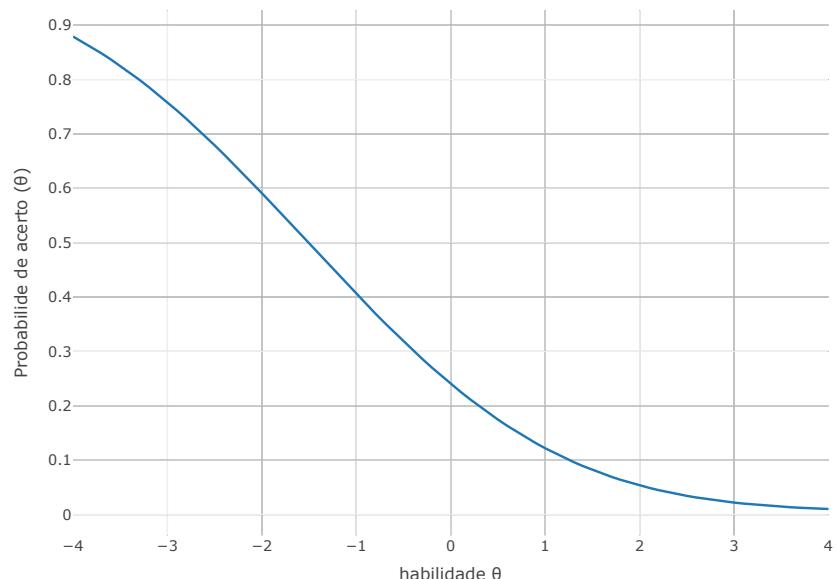


Figura 8. CCI Item 4 da área de Ciências Humanas

6. Conclusão

Por meio da aplicação de modelos da TRI nos resultados das provas do ENEM 2022 em cada área de conhecimento, pudemos realizar comparações entre os Estados com o melhor e pior desempenhos no IDEB 2021. Esse experimento permitiu gerar artefatos de habilidades estimadas, dificuldades dos itens e a geração da Curva Característica do Item, revelando a probabilidade de acerto que um candidato tem em relação a um determinado item dada uma área de conhecimento e prova específica. A análise comparativa entre os Estados revelou que o Paraná apresentou resultados superiores em relação ao Pará quando se tratava de itens de alta dificuldade. No entanto, para itens de dificuldade mediana e fácil, as habilidades estimadas foram bastante próximas entre os dois estados. Ao observar as curvas, foi possível notar que, no modelo de 3PLM, os parâmetros de discriminação e

acerto ao acaso desempenham um papel importante, as curvas se tornam mais ingremes ao chegar num certo nível de habilidade mostrando uma diferenciação entre habilidades distintas. Esses resultados ressaltam a relevância de considerar não apenas a dificuldade dos itens, mas também os parâmetros específicos de cada modelo da TRI ao analisar e interpretar os resultados do ENEM.

Referências

- Araujo, A. L. S. O., Andrade, W. L., Guerrero, D. D., Melo, M. R. A., and de Souza, I. M. L. (2018). Como identificar habilidades do pensamento computacional? um estudo empregando análise fatorial. In *IV Workshop de Ensino em Pensamento Computacional, Algoritmos e Programação*.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*.
- Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an em algorithm. *Psychometrika*, 46(4):443–459.
- IDEB (2023). <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/ideb>. Acesso em: 29-06-2023.
- INEP (2022). Microdados do enem 2022. <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>. Acessado em: 29-06-2023.
- INEP (2023). Matriz de referência do enem. https://download.inep.gov.br/download/enem/matriz_referencia.pdf. Acessado em: 29-06-2023.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.
- Monteiro, E. J. L. B. et al. (2021). Um método baseado na teoria da resposta ao item para avaliação e feedback automático no contexto de educação digital.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. The University of Chicago.