

Mensuração da Proficiência dos Estudantes do Curso de Bacharelado em Sistemas de Informação, na Prova do ENADE (2017) via Teoria de Resposta ao Item

Mário Diego Rocha Valente¹, Dionne Cavalcante Monteiro¹, Heliton Ribeiro Tavares²

¹Faculdade de Computação – Universidade Federal do Pará (UFPA)

²Faculdade de Estatística – Universidade Federal do Pará (UFPA)

Av. Augusto Corrêa 01 – Guamá – Belém – PA – Brasil

diego.vatente@gmail.com, dionne@ufpa.br, heliton@ufpa.br

Abstract. *This study aimed to measure the performance of undergraduate students in Information Systems who took the ENADE exam using Item Response Theory (IRT). Student performance was measured using the three-parameter logistic model (3PL), based on data extracted from the website of the National Institute for Educational Studies and Research Anísio Teixeira (INEP), referring to the ENADE examination conducted in 2017, comprising a total of 11,523 students. The estimated proficiency scale allowed the classification of students into three performance levels (Basic, Intermediate, and Advanced). The results also indicated that the items included in the ENADE test presented a relatively high level of difficulty for the evaluated group. In particular, students showed lower performance in the areas of Probability, Programming, and Databases. Conversely, students demonstrated a relatively higher performance in Artificial Intelligence and Theory of Computation. These findings suggest the need for institutional actions and public policies capable of contributing to the improvement of students' academic performance in these areas.*

Keywords *Performance, Proficiency, ENADE, Item Response Theory*

Resumo. *Este estudo teve como objetivo mensurar o desempenho dos estudantes do curso de Sistemas de Informação que realizaram o Exame Nacional de Desempenho dos Estudantes (ENADE) por meio da Teoria de Resposta ao Item (TRI). A mensuração do desempenho foi realizada utilizando o modelo logístico de três parâmetros (ML3P), com dados extraídos do site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), referentes à prova do ENADE aplicada em 2017, contemplando um total de 11.523 estudantes. A escala estimada permitiu classificar os estudantes em três níveis de desempenho (Básico, Intermediário e Avançado). Os resultados indicaram que os itens presentes na prova do ENADE apresentaram um elevado grau de dificuldade para o grupo avaliado. Observou-se, ainda, baixo desempenho dos estudantes nos temas de Probabilidade, Programação e Banco de Dados, porém um desempenho superior relativo nas áreas de Inteligência Artificial e Teoria da Computação. Esses resultados sugerem a necessidade de ações institucionais e de políticas públicas voltadas à melhoria do desempenho acadêmico dos estudantes nessas áreas do conhecimento.*

Palavras-Chave *Desempenho, Proficiência, ENADE, Teoria de Resposta ao Item*

1. Introdução

O Exame Nacional de Desempenho dos Estudantes (ENADE) é um dos pilares da avaliação do Sistema Nacional de Avaliação da Educação Superior (SINAES), criado pela Lei nº 10.861/2004. O exame é aplicado anualmente aos alunos de cursos de graduação, sendo o único instrumento oficial que avalia os resultados do processo de ensino-aprendizagem (BRASIL, 2017).

A escolha da edição de 2017 para o curso de Sistemas de Informação justifica-se por representar o último ciclo avaliativo pleno antes das reformulações promovidas pelas novas Diretrizes Curriculares Nacionais (DCNs) da Computação. Assim, os dados deste ano servem como uma linha de base (baseline) histórica para avaliar o impacto de futuras adequações curriculares.

A Prova do ENADE (2017) apresentou um componente de Formação Geral (FG), correspondendo a 25% da nota final (10 questões), e um Componente Específico (CE), correspondendo a 75% (30 questões). Desta forma, este trabalho tem o objetivo de mensurar a proficiência dos estudantes de Sistemas de Informação na referida prova, por meio da Teoria de Resposta ao Item (TRI), provendo subsídios para o aprimoramento de Projetos Político-Pedagógicos.

2. Fundamentação Teórica

2.1. Teoria da Resposta ao Item (TRI)

A Teoria de Resposta ao Item (TRI) é um conjunto de modelos matemáticos que representam a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da habilidade (traço latente) do respondente (ANDRADE et al., 2000). Diferente da Teoria Clássica dos Testes (TCT), que avalia o escore bruto de forma dependente do teste, a TRI permite que desempenhos sejam comparados e colocados em uma escala única de conhecimento.

Os modelos da TRI dependem fundamentalmente da natureza do item (dicotômico ou politômico), do número de populações envolvidas e do número de traços latentes medidos (unidimensional ou multidimensional). Para a aplicação clássica da TRI, duas suposições são necessárias: a **unidimensionalidade** (existência de uma aptidão dominante responsável pelo desempenho) e a **independência local** (respostas a itens distintos são independentes quando o traço latente é controlado) (PASQUALI, 2013). Neste estudo, assume-se a premissa de uma "Proficiência Global" dominante, caracterizando-se como uma limitação analítica a ser considerada, dado que a prova contempla múltiplos domínios.

2.2. Modelo Logístico de 3 Parâmetros (ML3P)

Neste estudo, adota-se um modelo unidimensional para itens dicotomizados. O Modelo Logístico de 3 parâmetros (LORD, 1980) é expresso pela Equação 1:

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

Em que:

- U_{ij} é uma variável dicotômica (1 = acerto, 0 = erro).
- θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.
- b_i é o parâmetro de **difículdade**. cabe ressaltar que, no ML3P, b_i não representa o nível associado a 50% de probabilidade de acerto, mas sim o ponto onde a probabilidade é $P(\theta = b_i) = \frac{1+c_i}{2}$
- a_i parâmetro de **discriminação** (inclinação da Curva Característica do Item). Espera-se $a_i > 1$ para boa discriminação (ANDRADE et al., 2000)
- c_i probabilidade de Acerto Casual (chute). Para 5 alternativas, espera-se valores próximos a 0,20 (SCHER et al., 2014).
- D fator de escala (constante 1,7).

3. Trabalhos Correlatos

No Brasil, alguns trabalhos abordaram a Teoria da Resposta ao Item (TRI) para avaliação de proficiência com dados do ENADE. Oliveira (2006) utilizou a TRI para proceder a uma análise das propriedades psicométricas da prova do ENADE de 2004, aplicada aos alunos dos cursos de medicina. Nogueira (2008) aplicou a TRI para avaliar as questões da prova de formação geral do ENADE de 2004 e 2005, a fim de estimar a proficiência dos estudantes e o ajuste dos itens ao modelo de Rasch.

Primi, Hutz e Silva (2011) realizaram análise psicométrica dos itens empregando os modelos Rasch e de créditos parciais na prova do ENADE de Psicologia de 2006. Camargo et al. (2016) mensuraram a proficiência dos estudantes do curso de Ciências Contábeis no ENADE utilizando TRI, demonstrando que os itens da prova representaram um alto grau de dificuldade para os estudantes e, em geral, os estudantes apresentaram proficiências muito baixas.

Figueiró e Mozzaquatro (2017) analisaram os microdados do Enade de 2014, para avaliar o desempenho dos estudantes do curso de Ciência da Computação do Estado do Rio Grande do Sul, bem como foram feitas comparações para agrupar as instituições públicas e privadas do mesmo estado de acordo com seus desempenhos

Em 2018, Lima et al., desenvolveram uma metodologia para classificação das questões do ENADE em temas. Araujo(2019), desenvolveu uma ferramenta web com dados de todos os cursos do Enade 2017, utilizando algoritmo de machine learning, para classificar as notas dos alunos por desempenho.

Charão et al. (2020) avaliaram o desempenhos dos alunos de Ciência da Computação da UFSM, na prova do Enade (2005 a 2017), explorando os resultados para subsidiar e reunir contribuições para a revisão do projeto pedagógico do curso.

Cunha et al. (2021) desenvolveram uma ferramenta de análise de dados automática com os microdados do Enade (2005 a 201) para o curso de Ciência da Computação da UFPA, via linguagem de programação Python. Rendeiro et al, (2022) analisou o desempenho dos estudantes de engenharia da computação da UFPA com base nos dados do Enade (2014,2017 e 2019).

4. Material e Métodos

4.1. População e Coleta de Dados

A população do estudo é constituída pelos alunos regularmente matriculados nos cursos de graduação em Sistemas de Informação, que fizeram a prova do ENADE em 2017.

Foi realizado o download dos microdados com extensão .txt por meio do site do INEP, como as respostas disponibilizados são apresentadas na forma de alternativas A, B, C, D ou E, estas foram comparadas ao gabarito. Assim, as respostas corretas receberam o valor 1(um) e as respostas erradas foram substituídas por 0(zero). Para o estudo foram consideradas as questões objetivas originais, ou seja 35 (trinta e cinco). Como a questão 14 foi anulada pelo INEP, ela foi excluída das análises de calibração e estimação das proficiências, restando 34 itens válidos.

4.2. Recursos Computacionais

Para mensuração do desempenho (proficiência) dos estudantes no ENADE foi implementada um script no software R-Project versão 4.5.5 (R DEVELOPMENT CORE TEAM, 2026) e um ambiente de desenvolvimento integrado chamado Rstudio versão 1.1.47 com uso dos pacotes: mirt (CHALMERS, 2012), ltm (RIZOPOULOS, 2006), irtoys (PARTCHEV, 2010).

4.3. Classificação dos Itens por Conteúdos Curriculares

Para realizar análises sobre um histórico de provas, é necessário adotar uma classificação comum para os conteúdos curriculares abordados nas 35 questões objetivas da prova do Enade. No quais os conteúdos são baseados nas diretrizes curriculares nacionais para os cursos de graduação, mas a cada edição estes são agrupados de forma diferente pelas comissões assessoras do INEP. Com isso, as cinco edições do exame para os cursos na área de computação, houve variação no agrupamento dos conteúdos de formação específica: 15 tópicos (2005 e 2008), 21 em (2011) e 17 em 2014 e 2017 (CHARÃO et al., 2020).

Para escolha de uma classificação para os itens, foi utilizado como referência o trabalho de Charão et al, (2020), tomando-se por referência as 17 categorias de conteúdos de formação específica do último exame em 2017 e 2014. Para os exames de 2005, 2008 e 2011, um especialista realizou manualmente a categorização de cada questão em até três temas, com base nas categorias do ENADE do 2017.

Para aprofundar o diagnóstico pedagógico, os itens também foram classificados sob a ótica da Taxonomia Revisada de Bloom (ANDERSON; KRATHWOHL, 2001), categorizando as questões em três estratos de exigência cognitiva: Básico (Lembrar e Compreender), Intermediário (Aplicar e Analisar) e Avançado (Avaliar e Criar).

4.3.1. Modelo de Mensuração da TRI

A mensuração do desempenho foi desenvolvida por meio da Teoria da Resposta ao Item (TRI) em que foi construída uma escala para avaliação dos níveis de proficiência dos alunos no Enade. Nessa forma de mensuração, todos os itens que compõem o instrumento de avaliação (prova) são colocados em uma mesma escala de medida da proficiência.

A estimação dos parâmetros dos modelos baseados na Teoria da Resposta ao Item foi realizada pelo Método de Máxima Verossimilhança Marginal, com a aplicação conjunta de um processo iterativo chamado de algoritmo Newton-Raphson ou Scoring de Fisher, conforme Andrade et al. (2000).

Para as comparações entre os modelos gerados utilizou-se o Teste da Razão de Verossimilhança por meio da Anova e os critérios AIC (Akaike Information Criterion)

e BIC (Bayesian Information Criterion), além de serem produzidos os gráficos com as Curvas Características do Item, Curvas de Informação do Item, Função de Informação Total do Teste.

5. RESULTADOS E DISCUSSÕES

Nos resultados obtidos pelos respondentes neste teste, observa-se notas variando entre 0 e 30 acertos, de um total de 35 itens. Destaca-se que ocorreu escore nulo (nenhum acertos, somente 284 alunos), como também não existiram respondentes que obtiveram escore total.

Tabela 1. Medidas Estatísticas do Escore Bruto Geral (contagem de acertos) para a Prova do ENADE/2017 referentes aos concluintes de Sistemas de Informação no Brasil (n=11.523).

Escore (0 a 35)	Formação Geral	Componente Específico	Escore Total
Nº de Itens	8	27	35
Média	3,99	10,90	14,90
Mínimo	0	0	0
Máximo	8	23	30
1º Quartil (25%)	3	8	12
3º Quartil (75%)	5	13	18

Para obter percentual de acerto/erro, foi considerado o vetor de resposta do estudante, checando cada questão e comparando com o total de questões de determinado tema. Analisando a distribuição de proporção de acertos, notou-se que, Processos Migratórios (84,5%), Teoria da Computação (83.1%) e Inteligência Artificial (73,8%) são os temas de maior percentual de acertos entre os estudantes concluintes do curso de Bacharelado em Sistemas de Informação. Para uma análise do ponto de vista de nota, a tabela 2 traz as médias.

Tabela 2. Medidas Estatísticas da Nota Bruta Geral (escala 0 a 100) para a Prova do ENADE/2017 referentes aos concluintes de Sistemas de Informação no Brasil (n=11.523).

Notas (0 a 100)	Formação Geral	Componente Específico	Nota Final
Mínimo	0,00	0,00	0,00
Máximo	100,00	95,50	87,30
Média	49,95	46,71	44,65
1º Quartil (25%)	37,50	36,40	35,50
Mediana	50,00	45,50	44,50
3º Quartil (75%)	62,50	59,10	53,70

Uma análise exploratória inicial dos dados sugere verificar, ainda de forma empírica, o grau de dificuldade de cada item, o que pode ser feito por meio da observação dos percentuais de acertos, mostrado na Tabela 3, onde se observa que itens com altos percentuais de acertos podem ser considerados como mais fáceis do que aqueles com baixo percentual de acerto.

Na Tabela 3 merecem destaques os itens 6 e 33 por apresentarem percentuais de acertos acima de 80% o que classifica, em princípio, como itens mais fáceis. De forma análoga, os itens 9, 10, 25 e 29 foram os itens mais difíceis, com percentuais de erro também acima de 80%.

Na verdade este fato sugere fortemente a análise de um especialista em itens para, de posse de outras informações, compreender e avaliar os itens em questão, inclusive identificando aqueles que não contribuem para o modelo e que, portanto, não devem ser considerados no processo de análise final.

Tabela 3. Proporção de Acertos e Erros empíricos para os 34 Itens na Prova do ENADE/2017 referentes aos alunos de Sistemas de Informação (n=11.523).

Caderno	Item	Erros (%)	(n)	Acertos (%)	(n)	Gabarito
FG	01	76,60	8.827	23,40	2.696	C
	02	49,97	5.758	50,03	5.765	C
	03	64,55	7.438	35,45	4.085	B
	04	36,74	4.233	63,26	7.290	B
	05	38,37	4.421	61,63	7.102	C
	06	15,46	1.782	84,54	9.741	E
	07	65,17	7.510	34,83	4.013	A
	08	53,54	6.169	46,46	5.354	D
CE	09	82,90	9.552	17,10	1.971	C
	10	89,60	10.325	10,40	1.198	E
	11	78,43	9.037	21,57	2.486	A
	12	48,13	5.546	51,87	5.977	C
	13	60,70	6.995	39,30	4.528	E
	15	49,83	5.742	50,17	5.781	A
	16	27,76	3.199	72,24	8.324	D
	17	26,21	3.020	73,79	8.503	B
	18	62,20	7.167	37,80	4.356	B
	19	54,27	6.254	45,73	5.269	B
	20	64,32	7.412	35,68	4.111	E
	21	69,96	8.061	30,04	3.462	B
	22	58,34	6.722	41,66	4.801	D
	23	54,46	6.276	45,54	5.247	D
	24	72,46	8.349	27,54	3.174	B
	25	84,63	9.752	15,37	1.771	B
	26	50,98	5.874	49,08	5.649	E
	27	49,41	5.694	50,59	5.829	A
	28	59,79	6.890	40,21	4.633	E
	29	80,01	9.220	19,99	2.303	A
	30	45,40	5.232	54,60	6.291	C
	31	50,76	5.849	49,24	5.674	B
	32	70,24	8.094	29,76	3.429	D
	33	16,90	1.947	83,10	9.576	C
	34	56,56	6.517	43,44	5.006	C
	35	45,38	5.229	54,62	6.294	C

5.1. Curvas Características da TRI

Na TRI, a Curva de Informação do Item (CII) indica o montante de informação psicométrica que cada questão fornece ao longo do traço latente. A Figura 1 apresenta as CIIs da prova. Observa-se que os itens 33, 11 e 1 são aqueles que fornecem maior quantidade de informação, contribuindo significativamente para a precisão da medida em determinados níveis de habilidade.

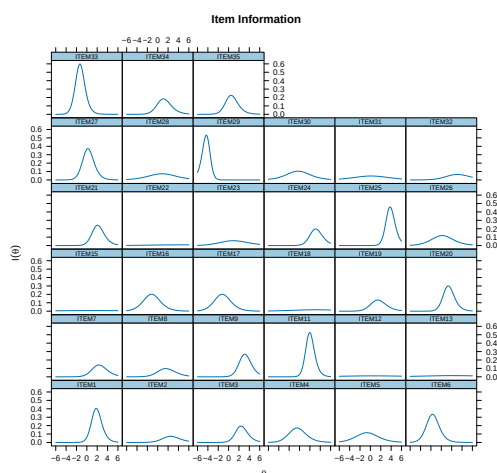


Figura 1. Curvas de Informação dos Itens (CII) para o curso de Sistemas de Informação.

A informação total fornecida pelo teste (CIT) é a soma das informações de cada um dos itens. A Figura 2 ilustra a CIT, evidenciando que a prova traz mais informação e estima o traço latente com maior precisão no intervalo de proficiência entre 0 e 4. Isso indica que a prova foi mais rigorosa e adequada para avaliar, com menor margem de erro, estudantes de nível intermediário a avançado.

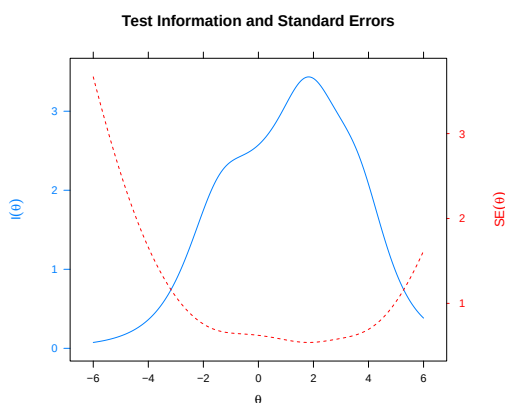


Figura 2. Curva de Informação Total do Teste (ENADE/2017).

Por fim, a Figura 3 apresenta as Curvas Características dos Itens (CCI), que descrevem a relação direta entre o nível de habilidade do estudante (eixo X) e a probabilidade de acerto em cada item (eixo Y). As diferentes inclinações e pontos de inflexão refletem graficamente as variações nos parâmetros de discriminação (a_i) e dificuldade (b_i) estimados no modelo.

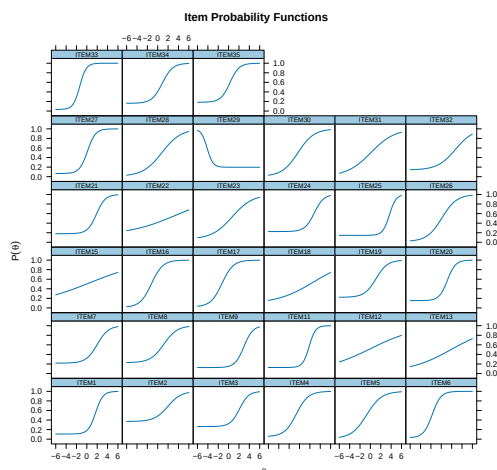


Figura 3. Curvas Características dos Itens (CCI) da Prova do ENADE/2017.

5.2. Critérios para Seleção dos Modelos

Dados dois ou mais modelos ajustados é necessário que se determine o que melhor explica a variável resposta, se ajustando melhor aos dados. A comparação dos modelos é feita via o Teste de Razão de Verossimilhança e os valores dos critérios de ajuste AIC (Akaike Information Criterion), criado por Akaike (1974) e BIC (Bayesian Information Criterion), descrito por Schwarz (1978).

Tabela 4. Comparação dos Modelos da TRI para a Prova do Enade/2017 referentes aos alunos matriculados no Curso de Sistemas de Informação do Brasil (n=11.523)

Comparação	Modelos	AIC	BIC	Verossimilhança	P-valor
Rasch x ML2PL	Rasch	417259.8	417480.3	-208599.9	0.000
	ML2PL	414313.1	414739.5	-207098.5	0.000
Rasch x ML3PL	Rasch	417259.8	417480.3	-208599.9	0.000
	ML3PL	414073.2	414712.8	-206949.6	0.000
ML2PL x ML3PL	ML2PL	414313.1	414739.5	-207098.5	0.000
	ML3PL	414073.2	414712.8	-206949.6	0.000

Para avaliar a adequação do modelo que está sendo testado tradicionalmente o qui-quadrado (χ^2), ou algum derivado, é o tipo de índice estatístico que é mais utilizado. Apesar desta estatística ter a vantagem de ter uma distribuição conhecida, o (χ^2) é fortemente influenciado pelo tamanho da amostra, pela normalidade dos dados e pela complexidade do modelo.

Para sanar tal desvantagem, muitos índices foram desenvolvidos. Porém, ao invés de confiar em apenas uma forma de mensuração, Nye e Drasgow (2011) afirmam que é altamente recomendado que pesquisadores examinem conjuntamente mais de um índice ao avaliarem seus modelos.

Para a presente pesquisa, a avaliação da qualidade dos ajustes dos modelos foi feita pelas estatísticas RMSEA (erro quadrático médio de aproximação, em inglês Steiger–Lind root mean square error of approximation: STEIGER, 1990), SRMR (resíduo quadrático médio padronizado, em inglês Standardized Root Mean Square Residual), CFI (índice de ajuste comparativo, em inglês Bentler’s Comparative Fit Index: BENTLER, 1990) e TLI (índice de Tucker-Lewis), usando a função M^2 do pacote mirt do software $R_{4.5}$.

Tabela 5. Estatísticas de Qualidade do Ajuste dos Modelos da TRI para a Prova do Enade/2017 referentes aos alunos matriculados no Curso de Sistemas de Informação do Brasil (n=11.523).

Modelos	Iterações	M^2	G.L	P-valor	RMSEA	SRMSR	TLI	CFI
Rasch	15	5550.83	405	0.000	0.033	0.043	0.8044	0.8004
ML2PL	11	2069.60	377	0.000	0.019	0.019	0.9309	0.9358
ML3PL	21	1489.91	348	0.000	0.016	0.017	0.9595	0.9667

Na concepção de Cai e Hansen (2013), o Índice de Tucker-Lewis(TLI) têm a vantagem de ser pouco sensíveis ao tamanho da amostra, ao contrário do teste quiquadrado de comparação de modelos, que, usualmente, rejeita a adequação de modelos quando a amostra é muito grande, mesmo nos casos em que os modelos são adequados.

De acordo com Kline (2005), RMSEA, SRMR, TLI e CFI são todos índices de ajuste aproximados, e não delimitam onde terminam as variações ao acaso nos desvios de covariância entre o esperado pelo modelo e observado na amostra e onde começam as evidências de um modelo mal ajustado.

As metodologias consideram o ajuste adequado quando $RMSE < 0.05$ e TLI e CFI são maiores que 0.90 (THIMOTY, 2015).

Segundos os critérios acima definidos, o melhor modelo para os dados do ENADE/2017 parece ser o ML3P, pois apresenta os menores valores para os critérios AIC e BIC, conforme Tabela 4.

A Tabela 6 apresenta os parâmetros de discriminação (a_i), dificuldade (b_i) e acerto ao acaso (c_i) dos itens avaliados. Tais parâmetros foram estimados no software $R_{4.5}$ com escala (0,1).

Os parâmetros a_i indicam o poder de discriminação de cada item. Após a exclusão de itens que apresentaram violações metodológicas severas como o item Q10 (falha de convergência) e o item Q29 (discriminação negativa, indicando possível erro de formulação) o modelo estabilizou-se de forma substancial. No conjunto calibrado, os itens com maior capacidade de diferenciar estudantes de maior e menor habilidade foram: Q25 (Teoria dos Grafos, $a_i = 1,67$), Q11 (Paradigmas de Linguagens de Programação, $a_i = 1,64$) e Q33 (Inteligência Artificial, $a_i = 1,60$). Sob a ótica pedagógica, isso demonstra que o domínio de fundamentos teóricos e paradigmas de programação atua como o principal divisor de águas na proficiência dos concluintes.

O parâmetro b_i representa a dificuldade do item avaliada na mesma métrica do traço latente. Os itens que exigiram o maior nível de proficiência para o acerto foram

fortemente marcados pelo rigor matemático e estrutural: Q22 (Lógica e Matemática Discreta, $b_i = 3,58$), Q25 (Teoria dos Grafos/Estruturas de Dados, $b_i = 3,56$) e Q09 (Algoritmos/Estruturas de Dados, $b_i = 2,95$). O fato de o item Q25 ser simultaneamente o mais difícil e o mais discriminativo indica que apenas os estudantes do nível avançado conseguem resolvê-lo. Em contrapartida, os itens mais fáceis da prova foram: Q06 (Processos Migratórios, $b_i = 1,75$), Q33 (Inteligência Artificial, $b_i = 1,38$) e Q17 (Ética e Sociedade, $b_i = 1,29$). A facilidade do item de Inteligência Artificial pode estar atrelada a uma abordagem predominantemente conceitual e introdutória do tema no ciclo de 2017.

Adicionalmente, observa-se na Tabela 6 uma forte convergência entre a exigência cognitiva e a métrica estatística: os itens classificados como *Avançados* na Taxonomia de Bloom (a exemplo das questões Q22 e Q25) coincidem diretamente com os maiores parâmetros empíricos de dificuldade (b_i) identificados pela modelagem.

O parâmetro c_i representa a probabilidade de acerto ao acaso. Todos os itens retidos apresentaram valores estatisticamente coerentes com a probabilidade teórica de acerto casual em testes de múltipla escolha com cinco alternativas (em torno de 0,20). Esse comportamento valida a confiabilidade geral do instrumento, comprovando que a adivinhação (chute) não distorceu significativamente a medida da proficiência latente dos estudantes.

Com base na distribuição da escala contínua de proficiência (θ) equalizada, definiram-se pontos de corte empíricos para classificar os estudantes em três níveis, os quais se aproximam do comportamento de uma distribuição normal baseada em desvios padrões: Nível Básico ($\theta < -1$), concentrando cerca de 16% dos alunos que dominam apenas conhecimentos rudimentares e interdisciplinares; Nível Intermediário ($-1 \leq \theta \leq 1$), englobando a grande maioria da amostra (aproximadamente 68%) capaz de resolver problemas de dificuldade média; e Nível Avançado ($\theta > 1$), contemplando os 16% de estudantes com alto desempenho, que se provaram os únicos capazes de responder aos itens de elevada complexidade estrutural e matemática exigidos pelo exame.

Essa tendência de dificuldade em disciplinas de base lógica e algorítmica vem sendo amplamente confirmada por pesquisas recentes na área de educação em Computação no Brasil. Ao analisarem o desempenho de concluintes no ENADE, estudos como o de Ramos, Oliveira e Braga (2024) identificaram que conhecimentos específicos exigindo alta abstração matemática, bem como arquitetura e infraestrutura de redes, continuam apresentando os menores índices de acerto em nível nacional. No mesmo sentido, investigações debatidas em edições recentes do Workshop sobre Educação em Computação, a exemplo de Trujillo (2025), evidenciam que essa defasagem estrutural nas ciências exatas não é um fenômeno isolado, mas sim um desafio crônico e sistêmico nos cursos de Tecnologia da Informação.

Aprofundando essa perspectiva com o uso de metodologias automatizadas para a exploração de microdados educacionais, trabalhos recentes como o de Matias (2026) constatarem, de forma análoga, que o raciocínio algorítmico representa o maior gargalo de desempenho dos estudantes nas avaliações do INEP. A convergência dos resultados obtidos em nosso estudo com essas pesquisas atuais (2024 a 2026) reforça a necessidade de modernização contínua das estruturas curriculares. Dessa forma, a aplicação da Teoria de Resposta ao Item demonstra ser um termômetro empírico essencial e de alta precisão

para orientar intervenções pedagógicas direcionadas e guiar a atualização dos Projetos Político-Pedagógicos (PPCs) em resposta às exigências contemporâneas de formação.

Tabela 6. Estimação dos Parâmetros Psicométricos (ML3P) e Classificação Cognitiva (Bloom) para os Itens da Prova do ENADE/2017 (n=11.523).

Itens	Conteúdo Classificado	Bloom (Cognitivo)	Discrim. (a_i)	Dificuldade (b_i)	Chute (c_i)
Q01	Globalização e Política Internacional	Avançado	1,41	1,69	0,11
Q02	Vida Urbana/Rural/Trabalho; CTS.	Avançado	0,78	1,94	0,37
Q03	Meio Ambiente; Responsabilidade Social	Avançado	1,14	2,10	0,26
Q04	TIC/Ciência/Tecnologia/Sociedade.	Básico	0,87	-0,61	0,05
Q05	Inovação Tecnológica/Meio Ambiente	Básico	0,69	-0,73	0,01
Q06	Processos Migratórios; Relações/Trabalho	Básico	1,19	-1,75	0,03
Q07	Cultura e Arte	Avançado	0,92	2,01	0,22
Q08	Responsabilidade Social/Ética e Cidadania	Intermediário	0,78	1,18	0,23
Q09	Algoritmos/Estruturas de Dados	Avançado	1,16	2,95	0,13
Q11	Paradigmas de Linguagens de Programação.	Avançado	1,64	1,85	0,13
Q12	Arquitetura de Computadores/Sistemas Operacionais.	Intermediário	0,22	0,09	0,05
Q13	Sistemas Digitais.	Avançado	0,24	2,04	0,02
Q15	Lógica e Matemática Discreta.	Intermediário	0,18	0,54	0,05
Q16	Redes de Computadores	Básico	0,91	-1,21	0,01
Q17	Ética, Computador e Sociedade.	Básico	0,92	-1,29	0,02
Q18	Inteligência Artificial e Computacional.	Avançado	0,28	2,66	0,08
Q19	Algoritmos/Estruturas de Dados.	Intermediário	0,90	1,08	0,22
Q20	Banco de Dados.	Intermediário	1,27	1,17	0,15
Q21	Redes de Computadores.	Avançado	1,16	1,86	0,18
Q22	Lógica e Matemática Discreta.	Avançado	0,21	3,58	0,14
Q23	Fundamentos e Técnicas de Programação.	Intermediário	0,51	0,71	0,07
Q24	Teoria da Computação.	Avançado	1,10	2,85	0,22
Q25	Teoria dos Grafos/Algoritmos/Estruturas.	Avançado	1,67	3,56	0,15
Q26	Teoria da Computação/Algoritmos/Estruturas.	Intermediário	0,70	0,11	0,01
Q27	Computação Gráfica; Processamento de Imagem.	Intermediário	1,30	0,13	0,07
Q28	Computação Gráfica; Processamento de Imagem.	Intermediário	0,55	0,81	0,01
Q30	Arquitetura de Computadores/Sistemas Operacionais.	Básico	0,65	-0,29	0,01
Q31	Compiladores.	Intermediário	0,44	0,11	0,01
Q32	Sistemas Distribuídos.	Avançado	0,59	2,74	0,14
Q33	Inteligência Artificial e Computacional.	Básico	1,60	-1,38	0,03
Q34	Teoria da Computação/Algoritmos/Estruturas.	Intermediário	1,00	0,88	0,16
Q35	Banco de Dados	Intermediário	1,13	0,25	0,18

*Itens Q10 e Q29 excluídos por anomalia metodológica; Q14 anulada. Taxonomia baseada em Anderson e Krathwohl (2001).

6. Considerações Finais

Este estudo teve como objetivo mensurar o desempenho (proficiência) dos estudantes de Computação no Enade, por meio da TRI. A partir da estimação realizada com ML3P, decorrente da TRI, foi realizada a mensuração da proficiência dos estudantes de Computação que fizeram a prova Enade/2017 e criada uma escala de medida padronizada. A análise dos itens (questões) evidenciou a capacidade da prova em mensurar

a proficiência dos estudantes dos diferentes níveis de domínio cognitivo exigido pelo exame. Nesse contexto, a TRI demonstrou capacidade de capturar a distribuição da proficiência dos estudantes de Computação ao longo dos níveis exigidos pela prova.

Os resultados da pesquisa apontaram que os itens contidos na prova apresentaram elevado grau de dificuldade em relação ao domínio cognitivo compreendido pela escala. Este resultado corrobora o baixo desempenho dos estudantes apontado pelo relatório do Inep (2017) para esta prova e aponta, especificamente, em quais aspectos e conhecimentos podem ser encontrados fragilidades de aprendizagem.

Vale ainda destacar que, enquanto estudos realizados com base na TCT abordam a questão do desempenho de maneira agregada, a análise pela TRI permitiu a identificação

pontual dos conhecimentos, capacidades e habilidades exigidas em cada nível da escala. Ao desmembrar a proficiência em níveis, evidenciando os conhecimentos exigidos, é possível uma atuação pontual de professores, IES e respectivas autoridades nas questões específicas em que foram demonstradas deficiências na aprendizagem.

Dados os benefícios mencionados e apresentados pela TRI neste estudo, pesquisas posteriores, relacionadas a determinantes do desempenho de alunos em Computação poderiam adotar como base para mensuração do desempenho dos estudantes as medidas fornecidas por modelos baseados na TRI. Essa análise possibilitaria estudar determinantes baseados em diferentes níveis de proficiência, contribuindo para o avanço nos estudos relacionados a esta área.

7. Declaração de Uso de Inteligência Artificial

Em conformidade com as diretrizes de transparência na pesquisa científica e com as políticas editoriais vigentes, os autores declaram o uso de ferramentas de Inteligência Artificial (IA) generativa como suporte na elaboração deste trabalho. A IA foi utilizada exclusivamente de maneira auxiliar nas etapas de revisão gramatical, aprimoramento da formatação estrutural em $\text{\LaTeX}$ e suporte na verificação de roteiros computacionais de análise de dados. Todo o processamento estatístico, embasamento teórico, interpretação psicométrica e conclusões pedagógicas são frutos originais e de inteira responsabilidade intelectual dos autores. Todo o material resultante do suporte da IA foi criteriosamente revisado e validado cientificamente antes da submissão.

8. Referências Bibliográficas

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control.*, Boston, v.19, n.6, p.716-723, Dec. 1974.

ANDERSON, L. W.; KRATHWOHL, D. R. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman, 2001.

Andrade, D.F., Tavares, H.R., Cunha, R.V. (2000). *Teoria da Resposta ao Item: Conceitos e Aplicações*. São Paulo: Associação Brasileira de Estatística.

Araújo, R. A. (2019). *Análise dos microdados do Enade: proposta de uma ferramenta de exploração utilizando mineração de dados*. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Goiás, Goiânia.

Birnbaum, A. (1968). Some latent trait models and their use in inferring and examinee's ability. In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley, ch. 17-20.

Bock, R. D. and Zimowski, M. F. (1997). Multiple Group IRT. In *Handbook of Modern Item Response Theory*. W.J. van der Linder e R.K. Hambleton Eds. New York: Springer-Verlag.

Baker, F.B.(2001). *The basics of item response theory*. Washington, DC: ERIC.

BORGATTO, A. F.; ANDRADE, D. F. de. *Análise clássica de testes com diferentes graus de dificuldade*. *Estudos em Avaliação Educacional*, São Paulo, v. 23, n. 52, p. 146–156, 2012.

BRASIL Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira Sinaes. Brasília, 2017.

CHARÃO, Andrea S.; WIECHORK, Karina; RODRIGUES, Marlon L. S.; BARBOSA, Fernando P.. Explorando Resultados por Questão no Enade em Ciência da Computação para Subsidiar Revisão de Projeto Pedagógico de Curso. In: WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO (WEI), 28. , 2020, Cuiabá. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020 . p. 16-20. ISSN 2595-6175.

CUNHA, Renan; SALES, Claudomiro; SANTOS, Reginaldo. Análise Automática com os Microdados do ENADE para Melhoria do Ensino dos Cursos de Ciência da Computação. In: WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO (WEI), 29., 2021, Evento Online. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2021 . p. 208-217. ISSN 2595-6175.

Camargo, R. V. W., Camargo, R. de C. C. P., Andrade, D. F. de, & Bornia, A. C. (2016). Desempenho dos alunos de ciências contábeis na prova ENADE/2012: uma aplicação da Teoria da Resposta ao Item. Revista De Educação E Pesquisa Em Contabilidade (REPeC), 10(3). <https://doi.org/10.17524/repec.v10i3.140>.

Coelho, E. C., Ribeiro Junior, P. J. & Bonat, W. H. (2014). Exame nacional de desenvolvimento de estudantes de estatística-desafios e perspectivas pela TRI. Revista da Estatística da Universidade Federal de Ouro Preto, 3(2), pp. 323-337.

Chalmers, R., P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software, 48(6), 1-29.

Hambleton, R.K., Swaminathan, H., Rogers, H.J. (1991). Fundamentals of Item Response Theory. Newbury Park: Sage Publications.

KATAOKA, V. Y. et al. O uso do r no ensino de probabilidade na educação básica: Animation e teachingdemos (the use of r in probability teaching at basic education: Animation and teaching demos). Proc. 18o Simpósio Nacional de Probabilidade e Estatística, São Paulo, 2008.

Lord, F.M. (1952). A theory of test scores (Psychometric Monograph No. 7). Iowa City, IA: Psychometric Society.

Lord, F.M. (1953). The relation of test score to the trait underlying the test. Educational and Psychological Measurement, 13, 517-549.

Lord, F.M., Norvick, M.R. (1968). Statistical Theories of Mental Test Score. Reading: Addison-Wesley.

Lord, F.M. (1980). Applications of Item Response Theory to Practical Testing Problems. Hillsdale: Lawrence Erlbaum Associates.

Lima, P. D. S., Ambrosio, A. P., Brancher, J. D., and Felix, I. (2018). Sysenadeanálise das questões de provas do enade organizadas pelos temas abordados. In Anais dos Workshops do Congresso Brasileiro de Informática na Educação, volume 7, page 419.

MATIAS, J. P. S. Uma ferramenta para visualização do desempenho dos cursos da UFC Quixadá no ENADE. Trabalho de Conclusão de Curso (Ciência da Computação)

– Universidade Federal do Ceará, Quixadá, 2026.

NOGUEIRA, S. O. ENADE: Análise de itens de formação geral e de estatística pela TRI. Dissertação de Mestrado, Programa de PósGraduação em Psicologia, Universidade São Francisco, Itatiba, 2008.

PASQUALI, L. Psicometria: teoria dos testes na psicologia e na educação. Petrópolis: Vozes, 2003.

Pestana, MH, & Gageiro, JG (2003). Análise de dados para ciências sociais: A complementaridade de SPSS [Data Analysis for Social Sciences: The Complementarity of SPSS] (3ª ed.). Lisboa: Edições Silabo.

PASQUALI, L. TRI - Teoria de Resposta ao Item: teoria, procedimentos e aplicações. Brasília: LabPAM/UnB, 2007.

PRIMI, Ricardo et al. Análise do funcionamento diferencial dos itens do Exame Nacional do Estudante (Enade) de Psicologia de 2006. Psico-USF, São Paulo, SP, v. 15, n. 3, p. 379-393, 2010.

PRIMI, R.; HUTZ, C. S.; SILVA, M. C. R. A prova do ENADE de psicologia 2006: concepção, construção e análise psicométrica da prova. Avaliação Psicológica, v. 10, n. 3, p. 271-294, 2011.

Rasch, G. (1960, 1980). Probabilistic models for some intelligence and attainment tests. Chicago, IL: MESA Press.

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. Journal of Statistical Software, 17(5), 1–25.

RABELO, M. Avaliação Educacional: Fundamentos, Metodologia e Aplicações no Contexto Brasileiro. Rio de Janeiro. SBM, 2013.

R DEVELOPMENT CORE TEAM. R: A Language and Environment for Statistical Computing, 2026. R Foundation for Statistical Computing. Vienna, Austria. ISBN: 3-90005107-0.

RAMOS, M. S.; OLIVEIRA, P. T. G.; BRAGA, A. R. Análise do Desempenho dos Estudantes do Curso Redes de Computadores da Universidade Federal do Ceará no ENADE. In: Anais do XXXII Workshop sobre Educação em Computação (WEI). Porto Alegre: SBC, 2024. p. 658-668.

RSTUDIO. RStudio: Integrated development environment for R (Versão 4.2.2) [Computer software]. Boston, MA.

Samejima, R. (1969). Estimation of latent ability using a response pattern of graded scores (Psychometric Monograph No. 17). Psychometric Society.

Samejima, F. (1972). A general model for tree-response data (Psychometric Monograph, No. 18). Psychometric Society.

SCHWARZ, G. Estimating the dimensional of a model. Annals of Statistics, Hayward, v.6, n.2, p.461-464, Mar. 1978.

SCHER, V. T. et al. Uma aplicação da teoria da resposta ao item na avaliação do ENADE do curso de Administração. XIV Colóquio Internacional de Gestão Universitária,

Florianópolis-SC, 3 a 5 de dezembro de 2014.

SIEVERT, C. et al. Plotly: Create Interactive Web Graphics via plotly.js. 2021. Disponível em: Acesso em: 03 fev. 2021.

THIMOTY, A. B. Confirmatory factor analysis for applied research. 2. ed. New York: The Guilford Press, 2015.

TRUJILLO, D. E. S. Análise Automatizada de Dados do ENADE com aplicação para os Cursos de Ciência da Computação. In: Anais do Workshop sobre Educação em Computação (WEI). Porto Alegre: SBC, 2025.