

Teoria de Resposta ao Item aplicada à prova do ENADE 2014 para Engenharia da Computação

Débora Spenassato¹, Rafael Tezza²

Resumo

Esse estudo visa apresentar as vantagens da utilização da Teoria de Resposta ao Item (TRI), em conjunto com a análise das questões objetivas do ENADE 2014 e do desempenho dos estudantes de Engenharia da Computação, comparando esses resultados com aqueles divulgados pelo INEP. Para isso, 26 itens foram analisados, sendo oito do componente formação geral (FG) e 18 de conhecimento específico (CE), respondidos por 2.424 estudantes do Brasil. O modelo logístico de três parâmetros foi utilizado, que envolve a dificuldade, a discriminação e a probabilidade de acerto ao acaso dos itens. Após análise pela TRI, seis itens foram eliminados e os níveis de dificuldade dos itens obtiveram ordens semelhantes às encontradas pelos índices de facilidade. Sugere-se adicionar itens mais difíceis para FG e itens mais fáceis para CE. Obteve-se uma correlação positiva e forte entre os escores das metodologias comparadas. Para FG, a maioria apresentou desempenho superior à média dos estudantes, já em CE, a maioria apresentou desempenho inferior à média. Por fim, do ponto de vista estatístico, nota-se que é preciso tomar mais cuidado com a elaboração da prova, pois restaram 20 itens na prova dentre os 35 itens respondidos pelos estudantes.

Palavras-chave: Educação Superior. Avaliação. Teoria da Resposta ao Item. Exame Nacional de Desempenho de Estudantes.

Abstract

This study presents the advantages of using Item Response Theory (IRT), together with an analysis of the main issues of ENADE 2014 and the performance of Computer Engineering students, comparing the results obtained with INEP. For that, 26 series items were analyzed, of which eight were of specific training (ST) and 18 of specific knowledge (SK), answered by 2.424 students from Brazil. The three parameter IRT model (3PM) was used, which involves a difficulty, a discrimination and a pseudoguessing to the items. After IRT analysis, six items were eliminated and the item difficulty levels were given orders similar to those found by the facility indexes. It is suggested a series of more difficult items for ST and easier items for SK. There was a positive and strong correlation between the scores of the comparative methodologies. For ST, the majority of the students achieved a performance higher than the average of students, already in SK, the majority presented below average performance. Finally, from the statistical point of view, one can see what is more precise and more careful with the creation of the test, since there were 20 items remaining in the test among the 35 items answered by the students.

Keywords: College education. Evaluation. Item Response Theory. ENADE.

¹ Universidade Federal do Rio Grande, E-mail: dspenassato@furg.br

² Universidade do Estado de Santa Catarina, E-mail: rafaeltezza@udesc.br

1 Introdução

O número de Instituições de Educação Superior (IES) e de cursos oferecidos tem crescido muito nos últimos anos no Brasil. Conforme resultados divulgados no último Censo da Educação Superior 2016, realizado anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), 34.366 cursos de graduação foram ofertados em 2.407 IES para um total de 8.052.254 estudantes matriculados; dessas IES, 87,7% são privadas (INEP, 2017).

Entre 2006 e 2016, houve aumento de 62,8% no número de matrículas na educação superior (graduação e sequencial). Porém, em relação a 2015, o aumento foi de 0,2%. Para este período de 10 anos, o número de cursos cresceu 48,1% (INEP, 2017). Especificamente para o curso de Engenharia da Computação, atualmente constam 129 cursos cadastrados no Sistema e-Mec, sendo que 10 desses estão em extinção.

Neste contexto, fica claro o expressivo aumento no número de IES. Por outro lado, desperta preocupação sobre a qualidade da educação que está sendo ofertada à população. Tal crescimento deveria incentivar a competitividade entre as IES, buscando a excelência de seus cursos. No entanto, esta pode não ser a realidade.

Uma vez que a qualidade da educação serve para fonte de consultas pela sociedade e pode auxiliar no desenvolvimento de políticas públicas (LOPES; VENDRAMINI, 2015), torna-se extremamente importante avalia-las de forma eficiente, identificando cursos em condições de desempenho inaceitáveis para que as devidas providências sejam tomadas.

No Brasil, a avaliação do Ensino Superior é realizada pelo Sistema Nacional de Avaliação da Educação Superior (SINAES), criado pela Lei nº 10.861, de 14 de abril de 2004, com três pilares: o Exame Nacional de Desempenho dos Estudantes (ENADE), os

processos de Avaliação de Cursos de Graduação e de Avaliação Institucional. Esses processos permitem conhecer de forma minuciosa o funcionamento e a qualidade dos cursos e IES de todo o Brasil (INEP, 2016).

Foco desse trabalho, o ENADE busca avaliar o desempenho dos estudantes “em relação aos conteúdos programáticos previstos nas diretrizes curriculares, às habilidades e competências para a atualização permanente e aos conhecimentos sobre a realidade brasileira, mundial e sobre outras Áreas do conhecimento” (INEP, 2016, p.3). É uma prova obrigatória para todos os cursos e acontece a cada três anos. A situação de regularidade do estudante deve constar em seu histórico escolar.

A nota final do curso depende do desempenho dos estudantes concluintes que são selecionados para participar da prova, composta por itens de conhecimento geral e específico. É considerada satisfatória uma nota de 3 a 5 para o curso, sendo esta última, a nota máxima.

O resultado detalhado do ENADE é divulgado em um relatório específico de cada área, que usa a Teoria Clássica de Testes (TCT), sendo esta a metodologia tradicionalmente utilizada para avaliar os itens e mensurar o escore no processo de avaliação e que possui limitações.

Esse estudo visa apresentar as vantagens da utilização da TRI, em conjunto com a análise das questões objetivas do ENADE 2014 para o curso de Engenharia da Computação, comparando a dificuldade com os índices de facilidade dos itens divulgado no relatório da área pelo INEP (2016). Além disso, apresenta-se uma análise do desempenho dos estudantes em cada componente da prova.

A TRI é uma metodologia para quantificar/analisar os resultados de um teste (exame, prova, questionário etc) e surge como uma alternativa à TCT.

Conforme destacado por Lopes e Vendramini (2015), é importante que os instrumentos utilizados nessas avaliações sejam de qualidade, isto é, com propriedades psicométricas que garantam a validade e a confiabilidade de seus resultados, uma vez que os impactos atingem um grande número de IES e, conseqüentemente, os docentes e estudantes.

Destaca-se que uma versão mais curta deste artigo foi publicado no evento 8º Conferência Sul em Modelagem Computacional e VIII Seminário e Workshop em Engenharia Oceânica (SPENASSATO; TEZZA, 2018). Para esta revista, acrescentou-se mais detalhes sobre a amostra utilizada, fez-se uma análise dos conteúdos abordados nos itens mais difíceis e mais fáceis, além da análise do desempenho dos estudantes pela TRI, comparando-os com os escores divulgados pelo INEP (2016).

2 Teoria da Resposta ao Item

A TRI compreende um conjunto de modelos matemáticos que representam a probabilidade de um indivíduo dar certa resposta a um item (questão), em função dos parâmetros do item e do traço latente do indivíduo (HAMBLETON; SWAMINATHAN; ROGERS, 1991; ANDRADE; TAVARES; VALLE, 2000).

O traço latente é uma característica que não pode ser mensurada diretamente, mas pode ser estimada por meio de um conjunto de itens desenvolvidos para tal finalidade, sendo representada pelo escore no teste, nota, proficiência, habilidade, desempenho, entre outros termos.

Essa metodologia tem como foco central o estudo individualizado das características dos itens de um teste, por isso, algumas vantagens podem ser obtidas em relação à TCT, em que o resultado do teste dependerá apenas do número de itens respondidos corretamente.

A TCT apresenta características dos itens que são dependentes do grupo de indivíduos utilizados para obtê-los e o traço latente do indivíduo depende das características dos itens que são aplicados (HAMBLETON; SWAMINATHAN; ROGERS, 1991; EMBRETSON; REISE 2000; KLEIN, 2009). Assim, quando os itens do teste forem muito difíceis, os respondentes parecerão ter baixo traço latente e vice-versa.

Vários autores destacam os benefícios da TRI, como Andrade, Tavares e Valle (2000), Hambleton, Swaminathan e Rogers (1991), Embretson e Reise (2000). Dentre os principais, tem-se que as características (parâmetros) dos itens são independentes do grupo utilizado para sua estimação e a estimativa do traço latente não depende dos itens apresentados; o modelo não requer testes paralelos para avaliar a confiabilidade e é possível obter a precisão para cada nível em particular do traço latente.

Também, como a TRI coloca itens e respondentes na mesma escala, é possível fazer uma interpretação detalhada do que o traço latente estimado representa, além de possibilitar a comparação de indivíduos que foram submetidos a testes totalmente ou parcialmente diferentes, bem como efetuar comparações ao longo do tempo.

A TRI dispõe de vários modelos, os quais dependem basicamente do tipo de dados (dicotômicos, policotômicos etc), do número de populações envolvidas e da quantidade de traços latentes que está sendo medida (unidimensionais ou multidimensionais) (ANDRADE; TAVARES; VALLE, 2000). Nesse estudo será utilizado o Modelo Logístico Unidimensional de Três Parâmetros (MLU3P).

O MLU3P (BIRNBAUM, 1968) baseia-se no fato de que os indivíduos com maior traço latente possuem maior probabilidade de acertar o item e que esta relação é não linear.

Assim, a probabilidade p_{ij} de que o indivíduo j responda corretamente ao item i é dada pela função de resposta ao item (FRI), conforme a Equação (1):

$$p_{ij} = P(Y_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (1),$$

onde $i = 1, \dots, I; j = 1, \dots, n$ e θ_j é o traço latente do indivíduo j , dado pelo escore na prova. A representação gráfica da FRI é também chamada de curva característica do item (CCI). O parâmetro de dificuldade do item i (b_i) está na mesma escala do traço latente, $b_i \in (-\infty, +\infty)$. A discriminação do item i (a_i) é proporcional à inclinação da CCI no ponto b_i , $a_i > 0$, e quanto maior o seu valor, maior será o poder do item em discriminar os respondentes. O parâmetro c_i representa a probabilidade de acerto casual do item i , $c_i \in [0,1]$.

O processo de estimação dos parâmetros dos itens é chamado de calibração. Conforme destacam Primi, Hutz e Silva (2011), uma vez determinado o traço latente do indivíduo e os parâmetros dos itens, por meio do modelo pode-se calcular uma expectativa teórica do padrão de resposta aos itens (acertos). Logo, se houver uma boa correspondência entre as observações e o modelo, essas estimativas serão bem acuradas.

Diversos estudos têm sido realizados no âmbito da TRI, nas diferentes áreas do conhecimento, demonstrando sua relevância para o meio científico. Em avaliações educacionais brasileiras em larga escala, podem-se citar as provas do Exame Nacional do Ensino Médio (ENEM), do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP), do Sistema de Avaliação da Educação Básica (SAEB) e da Prova Brasil.

No contexto do ENADE, diversos trabalhos vêm sendo desenvolvidos ao longo dos anos, com foco na avaliação das propriedades

psicométricas da prova de diferentes cursos, seja para formação geral e/ou específica, apenas das questões objetivas e/ou das questões discursivas.

Podem-se citar os trabalhos de Primi, Hutz e Silva (2011) para a prova ENADE 2006 de Psicologia, empregando o modelo Rasch e de créditos parciais (para as questões dissertativas) da TRI. Lopes e Vendramini (2015) avaliaram as propriedades psicométricas da prova de Pedagogia do ENADE 2005 pelo modelo de Rasch. Nogueira (2008) aplicou a TRI para avaliar as questões de FG do ENADE 2004 e 2005 por meio do modelo de Rasch. Scher et al. (2014) apresentam uma análise da prova do ENADE 2009 do Curso de Administração e Camargo et al. (2016) avaliaram o desempenho dos alunos de Ciências Contábeis na prova ENADE 2012, ambos estudos usaram o MLU3P.

3 ENADE para Engenharia da Computação

O ENADE é formado por dois componentes: (1) Componente de Formação Geral (FG), que configura a parte comum às provas das diferentes áreas e avalia competências, habilidades e conhecimentos gerais dos estudantes, os quais facilitam a compreensão de temas exteriores ao âmbito específico de sua profissão e à realidade brasileira e mundial (com oito questões objetivas e duas discursivas); e (2) Componente de Conhecimento Específico (CE), que contempla a especificidade de cada área, no domínio dos conhecimentos e habilidades esperados para o perfil profissional (com 27 questões objetivas de múltipla escolha e três discursivas) (INEP, 2016). Cada prova possui pesos diferentes, assim como o tipo de questão.

A prova de CE foi composta por um núcleo de conteúdos básicos e outro de conteúdos profissionalizantes, baseados no perfil dos

profissionais da área. Assim, a prova avaliou se o estudante desenvolveu, em seu processo de formação, as seguintes competências e habilidades (INEP, 2016, p.10):

I - antever as implicações humanísticas, sociais, ambientais, éticas, profissionais, legais (inclusive relacionadas à propriedade intelectual) e políticas dos sistemas computacionais;

II - identificar demandas socioeconômicas e ambientais relevantes, planejar, especificar e projetar sistemas de computação, seguindo teorias, princípios, métodos e procedimentos interdisciplinares;

III - construir, testar, verificar e validar sistemas de computação, seguindo métodos, técnicas e procedimentos interdisciplinares;

IV - perceber as necessidades de atualização decorrentes da evolução tecnológica e social;

V - relacionar problemas do mundo real com suas soluções, considerando aspectos de computabilidade e de escalabilidade;

VI - analisar, desenvolver, avaliar e aperfeiçoar software e hardware em arquiteturas de computadores;

VII - analisar, desenvolver, avaliar e aperfeiçoar sistemas de automação e sistemas inteligentes;

VIII - analisar, desenvolver, avaliar e aperfeiçoar sistemas de informação computadorizados;

IX - analisar, desenvolver, avaliar e aperfeiçoar circuitos eletroeletrônicos;

X - gerenciar pessoas e infraestrutura de Sistemas de Computação;

XI - perceber as necessidades de inovação e inserção internacional com atitudes criativas e empreendedoras.

A prova do ENADE 2014 foi aplicada aos estudantes concluintes (que se encontravam

no último ano do curso) da Área de Engenharia da Computação, no dia 23 de novembro de 2014, com duração total de quatro horas.

Conforme divulgado no relatório de área (INEP, 2016), inscreveram-se 2.972 estudantes no Brasil, destes, 14,2% estavam ausentes. Logo, dos 2.550 estudantes do curso de Engenharia da Computação (eComp) inscritos e presentes, 1.613 (63,3%) estudavam em Universidades, 516 (20,2%) em Faculdades e 421 (16,5%) estavam vinculados a Centros Universitários. Tais estudantes eram procedentes de 18 Estados.

4 Método

4.1 Descrição da amostra

Consideraram-se como critérios de inclusão nesse estudo os alunos de cursos presenciais, que estavam presentes e tiveram suas provas validadas. Dessa forma, foram consideradas as respostas de 2.424 estudantes do curso de eComp de todo o Brasil. Houve predominância do sexo Masculino (M), com 88,4%. Além disso, observa-se na Figura 1 que a maioria dos estudantes era da região Sudeste do Brasil, sendo a região Centro-Oeste com menor participação.

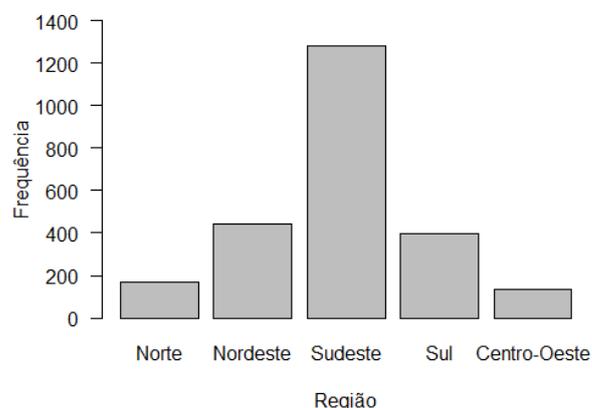


Figura 1: Participação dos estudantes por Região

Os estudantes possuíam, em média, 25,2 anos e desvio padrão de 3,8 anos. As idades variaram entre 20 e 59 anos e a idade mediana foi de 24 anos.

A nota média dos estudantes da eComp no ENADE foi de 47,55 e mediana de 47,35. A nota máxima obtida foi de 93,8 (estudante da UNICAMP-Campinas/SP). Essa nota final leva em consideração os dois componentes avaliados na prova (FG e CE).

4.2 Análise dos dados

O ENADE possui oito questões objetivas de múltipla escolha de FG e 27 de CE, totalizando 35 questões. Destaca-se que nove itens de CE foram excluídos pelo próprio ENADE, logo, não foram computados no escore. Essas questões, ou foram anuladas pela comissão ou apresentaram um coeficiente de correlação ponto bisserial abaixo de 0,2.

O coeficiente de correlação ponto bisserial é uma medida de capacidade de discriminação do item em relação ao resultado do teste, ou seja, medem a correlação do resultado de um item em particular do teste com o resultado do teste (SOARES, 2005). Portanto, um item deve ser mais acertado por estudantes que tiveram bom desempenho do que pelos que tiveram desempenho ruim (INEP, 2016).

Devido a este fato, foram analisados pela TRI oito itens para a escala de FG e 18 para a escala de CE, totalizando 26 itens para investigação pela TRI. Manteve-se a mesma numeração dos itens que foi divulgado pelo INEP (2016) para proporcionar a comparação com os índices de facilidade por eles divulgados. A estimação dos parâmetros dos itens foi realizada por meio do modelo MLU3P da TRI com o auxílio do *software* Bilog (ZIMOWSKI et al., 2003). O *software* R (R Core Team, 2018) também foi utilizado para representação gráfica.

No processo de calibração, primeiramente estimam-se os parâmetros dos itens considerando uma distribuição normal para o traço latente e, posteriormente, consideram-se os parâmetros dos itens conhecidos e estima-se o traço latente dos estudantes. O Bilog considera as seguintes restrições para os parâmetros dos itens em termos de distribuições de probabilidade: $a_i \sim \text{log-normal}(1;1,649)$ e $c_i \sim \text{beta}(5,17)$.

4.3 Escala

No momento de calibração dos itens é que se define a escala que servirá de referência para a interpretação de resultados do teste e que torna possível a comparabilidade dos respondentes. No estudo utilizou-se a escala Normal, com média zero e desvio padrão um, $N(0,1)$.

Conforme Klein (2009), a interpretação significa escolher alguns níveis da escala e descrever os conhecimentos e habilidades que os estudantes demonstraram possuir quando situados em torno desses níveis. Esses pontos são interpretados pedagogicamente por especialistas da área. Neste estudo, não fez-se a interpretação da escala, mas, em Andrade, Tavares e Valle (2000) e Klein (2009) pode-se obter mais informações sobre o posicionamento dos itens em níveis âncora da escala.

5 Resultados

Apresenta-se a análise dos itens e do desempenho dos estudantes para os componentes de formação geral e conhecimentos específicos.

5.1 Análise dos itens de conhecimentos gerais

A Tabela 1 apresenta os parâmetros dos itens (a - discriminação, b - dificuldade e c - probabilidade de acerto ao acaso) e erro padrão (EP - identificado com *) de

estimação para a prova de formação geral. Dos oito itens analisados, um deles precisou ser excluído (item 2 da prova objetiva), pois apresentou correlação ponto bisserial abaixo de 0,2 para a amostra de estudantes eComp e, conseqüentemente, baixo poder de discriminação e EP de estimação do parâmetro de dificuldade do item elevado.

Tabela 1: Parâmetros dos itens FG e EP.

Item	<i>a</i>	<i>b</i>	<i>c</i>
I1	0,702	-0,201	0,167
	0,110*	0,290*	0,074*
I3	1,404	-1,175	0,109
	0,154*	0,131*	0,052*
I4	1,346	-1,855	0,150
	0,147*	0,182*	0,070*
I5	0,831	1,169	0,186
	0,212*	0,214*	0,062*
I6	0,934	-0,508	0,186
	0,139*	0,252*	0,078*
I7	1,305	-0,508	0,138
	0,181*	0,153*	0,061*
I8	1,072	-1,720	0,145
	0,116*	0,204*	0,068*

Os sete itens apresentaram bom poder de discriminação (acima de 0,702), sendo que o mais fácil foi o item 4 ($b = -1,855$) que aborda porcentagens e o uso de redes sociais no processo de recrutamento de funcionários. O mais difícil foi o item 5 ($b = 1,169$) que trata de um dispositivo simples para ajudar a salvar vidas de bebês que ficam presos no canal vaginal, durante o parto. Ambas questões exigem interpretação do enunciado.

A Tabela 2 apresenta o resultado divulgado pelo INEP para eComp, utilizando a TCT, que usa como base o índice de facilidade, obtido pela porcentagem de respostas corretas para cada item. Pode-se observar que a ordem dos itens (mais fácil a mais difícil) manteve-se praticamente a mesma quando comparado ao nível de dificuldade de cada item (b) estimado pela TRI.

Tabela 2: Índice de facilidade e classificação das questões objetivas de FG.

Item	Índice de facilidade	
	Valor	Classificação
1	0,61	Fácil
3	0,79	Fácil
4	0,89	Muito fácil
5	0,43	Médio
6	0,67	Fácil
7	0,68	Fácil
8	0,84	Fácil

Fonte: Adaptado de INEP (2016).

A Figura 2 apresenta a CCI e a função de informação do item (FII) para os itens 7 e 8. A FII mostra para quais níveis da escala o item fornece mais informação, isto é, maior precisão de medição. Assim, um item produz a máxima informação sobre o traço latente em torno do nível de dificuldade ($\theta = b$) e quando ele for muito discriminativo e pouco acertado ao acaso (BAKER, 2001).

A FII também mostra quanto o item está contribuindo para a composição do escore final. Assim, percebe-se que o item 7 é mais informativo do que o item 8 (Figuras 2b e 2d). Na CCI é possível observar a representação de cada parâmetro do item na curva, quanto mais inclinada, maior a discriminação.

A Figura 3 apresenta a Função de Informação do Teste (FIT) para a prova de FG (linha contínua). Ela informa a precisão de mensuração do escore dos estudantes para cada nível da escala (linha tracejada), bem como em quais níveis devem ser acrescentados mais itens para melhorar a precisão.

Como são poucos itens para FG, a mensuração da nota acaba não sendo muito precisa, principalmente para estudantes acima da média zero, mostrando que precisam ser adicionadas questões mais difíceis à escala. Além disso, informa que a

prova de FG foi mais eficiente para mensurar a nota de estudantes em torno do pico -1.

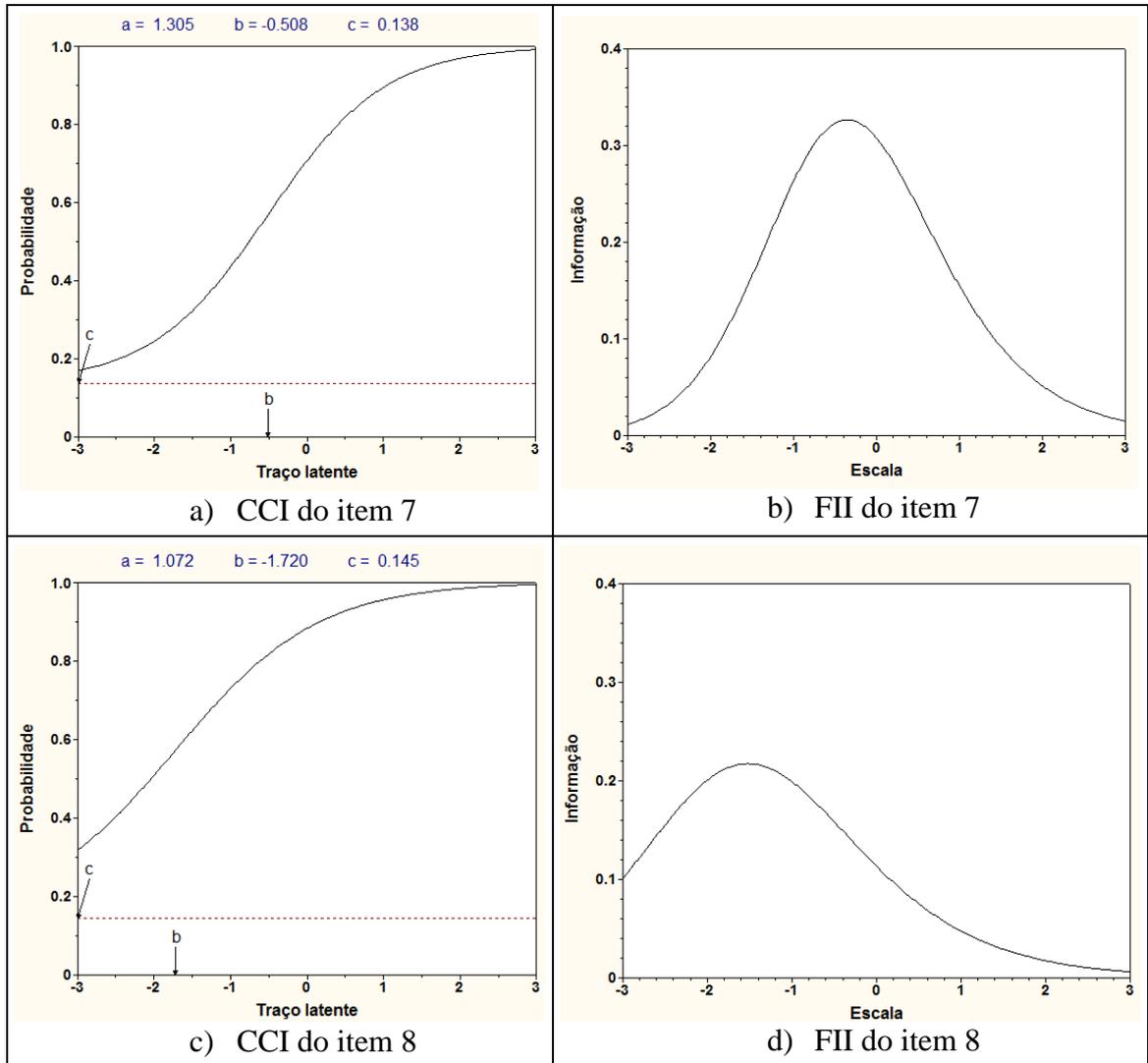


Figura 2: CCI e FII dos itens 7 e 8 de FG

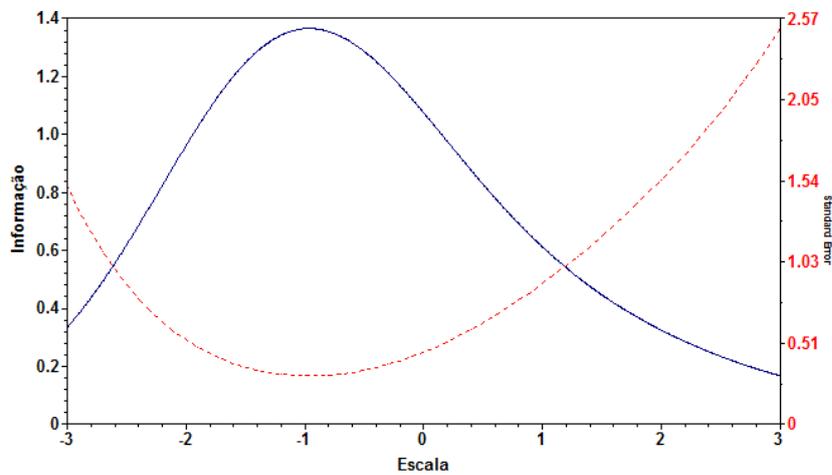


Figura 3: FIT para prova de FG

5.2 Análise do desempenho na prova de conhecimentos gerais

Nesta escala, zero significa o desempenho médio dos estudantes na prova e valores positivos e negativos indicam desempenhos acima ou abaixo da média. A Figura 4 mostra o histograma desses desempenhos. Tem-se que 47,3% dos estudantes tiveram escore inferior à média zero e 52,7% zero ou mais, ou seja, a maioria teve desempenho acima da média. Do total, 15 estudantes erraram todos os itens e 319 acertaram todos os itens.

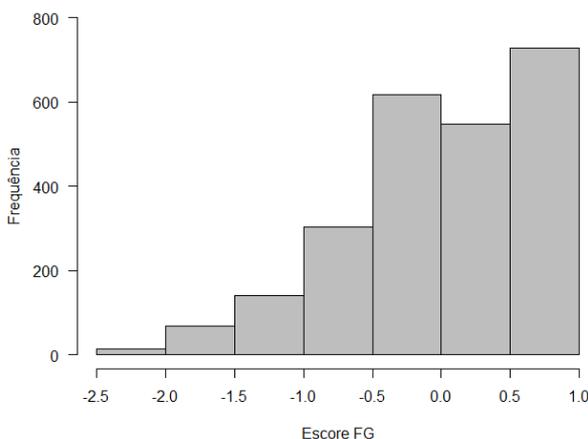


Figura 4: Desempenho dos alunos em conhecimentos gerais

A Figura 5 representa a relação entre os escores na prova objetiva de FG divulgado pelo INEP e os escores obtidos por meio da TRI. Conforme esperado esta relação é positiva e forte, com coeficiente de correlação de Pearson de 0,926. Essa diferença pode ter sido gerada pelo número de itens usados em cada situação (8 pela TCT e 7 pela TRI) e pelas metodologias usadas para se obter o escore.

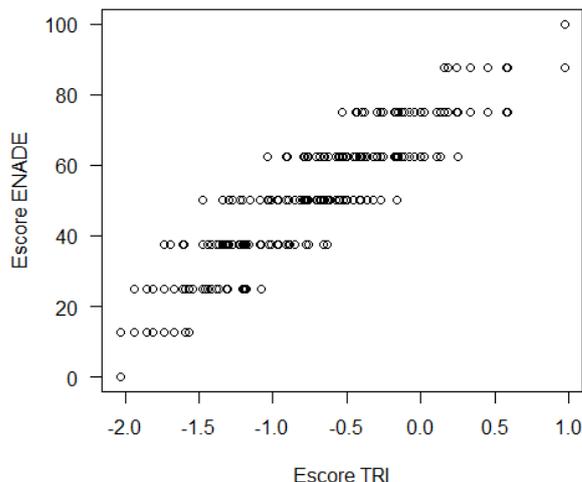


Figura 5: Dispersão entre o escore ENADE em FG e escore pela TRI

5.3 Análise dos itens da prova de conhecimentos específicos

A Tabela 3 apresenta os parâmetros dos itens da prova de CE e o EP (*) de estimação pela TRI. Ressalta-se que, além das nove questões que já haviam sido excluídas pelo próprio INEP, mais cinco questões foram excluídas pela análise da TRI [três itens apresentaram EP elevado – 20,26,33; um item apresentou parâmetro $a < 0,7$ (item 19) e um item apresentou correlação ponto bisserial $< 0,2$ (item 12)]. Ao final, manteve-se 13 itens na prova de CE.

O Item 18 é o mais fácil, que aborda conhecimento sobre o alumínio aeronáutico e suas propriedades, seguido pelos itens 17 que envolve interpretação e representação gráfica em eleições.

O item 10 foi o mais difícil, que trata de conhecimentos sobre circuito elétrico e passagem de corrente elétrica, seguido pelo item 22 que envolve conceitos de grafos, circuito euleriano e ciclo hamiltoniano, para resolver problemas de pesquisa operacional como do carteiro chinês e do caixeiro viajante.

Essa sequência entre os mais fáceis e mais difíceis também se mantiveram pela TCT. Houve algumas alterações para itens com níveis intermediários de dificuldade.

Tabela 3: Parâmetros dos itens e EP para CE.

Item	<i>a</i>	<i>b</i>	<i>c</i>
I9	1,571	0,984	0,207
	0,317*	0,096*	0,036*
I10	0,816	2,538	0,115
	0,210*	0,299*	0,035*
I11	1,156	-0,405	0,143
	0,131*	0,164*	0,063*
I13	1,274	-0,661	0,118
	0,123*	0,131*	0,054*
I14	1,245	-0,196	0,141
	0,154*	0,151*	0,059*
I15	0,926	0,931	0,135
	0,152*	0,160*	0,050*
I16	1,041	1,587	0,238
	0,255*	0,167*	0,043*
I17	1,295	-0,776	0,115
	0,121*	0,127*	0,054*
I18	0,974	-1,724	0,167
	0,099*	0,225*	0,077*
I21	1,406	1,433	0,274
	0,360*	0,128*	0,034*
I22	1,1	1,796	0,178
	0,270*	0,166*	0,036*
I23	1,494	0,936	0,233
	0,300*	0,106*	0,038*
I29	1,338	1,111	0,252
	0,300*	0,124*	0,041*

A Tabela 4 apresenta os resultados usando a TCT para os índices de facilidade. As nove questões excluídas por eles da nota final (24, 25, 27, 28, 30, 31, 32, 34 e 35), conforme

motivos já citados na metodologia, haviam sido classificadas como difícil ou muito difícil. Além desses, os itens excluídos após análise pela TRI, que são 5, haviam sido classificadas como difícil ou médio.

Tabela 4: Índices de facilidade divulgados pelo INEP para as questões objetivas de CE eComp.

Item	Índice de facilidade	
	Valor	Classificação
9	0,39	Difícil
10	0,22	Difícil
11	0,62	Fácil
13	0,66	Fácil
14	0,58	Médio
15	0,39	Difícil
16	0,37	Difícil
17	0,68	Fácil
18	0,80	Fácil
21	0,39	Difícil
22	0,30	Difícil
23	0,42	Médio
29	0,42	Médio

Fonte: Adaptado de INEP (2016).

A Figura 6 apresenta a FIT para a prova de CE. Nota-se que a escala mede com maior precisão os respondentes que estão acima do desempenho médio (zero), tendo seu pico em aproximadamente 1. Além disso, como a curva está deslocada à direita da média, sugere-se que sejam adicionados mais itens fáceis à escala, pois somente cinco itens dentre os 13 apresentaram nível de dificuldade abaixo da média.

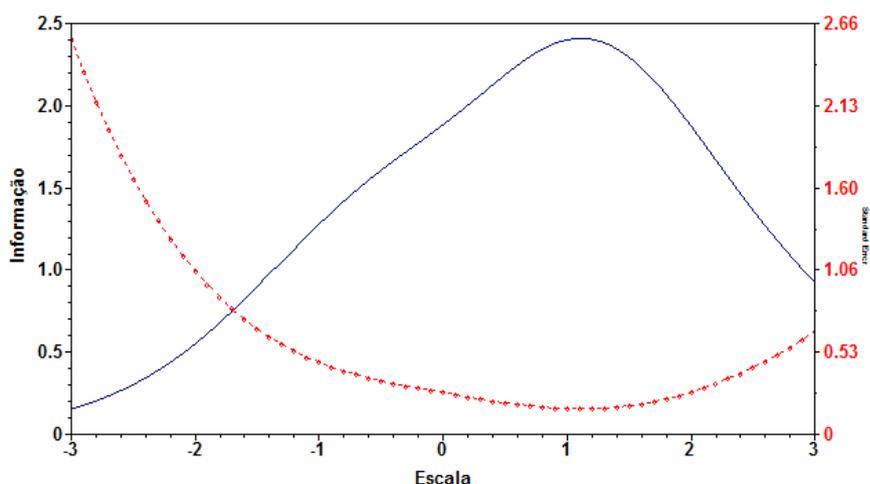


Figura 6: FIT para conhecimento específico

5.4 Análise do desempenho na prova de conhecimentos específicos

Quanto ao desempenho dos alunos em CE (Figura 7), 51,4% dos estudantes tiveram desempenho inferior à média zero. Do total, 8 estudantes erraram todos os itens e apenas 10 acertaram todos os itens.

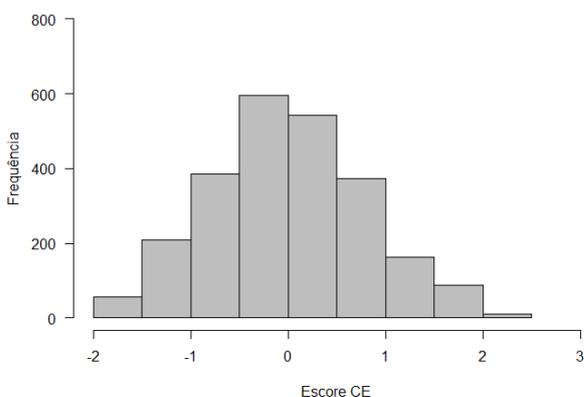


Figura 7: Desempenho dos alunos em conhecimentos específicos.

A Figura 8 representa a relação entre os escores na prova objetiva de CE divulgado pelo INEP e os escores obtidos por meio da TRI. Esta relação é positiva e forte, com coeficiente de correlação de Pearson de 0,914. Neste caso, tem-se 13 itens analisados pela TRI e 18 pela TCT.

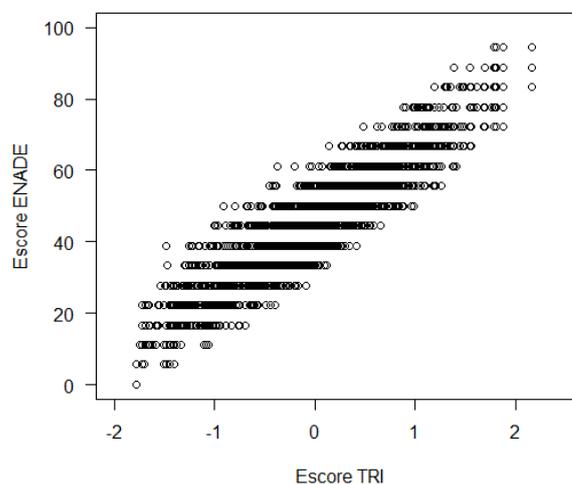


Figura 8: Dispersão entre o escore ENADE em CE e escore pela TRI

A correlação entre os escores FG e escores CE foi de 0,375, o que indica uma relação positiva fraca, ou seja, nem todos os estudantes que apresentaram um bom desempenho em FG, também apresentaram bom desempenho em CE, ou vice-versa.

6 Conclusão

Esse estudo visou apresentar as vantagens da utilização da Teoria de Resposta ao Item (TRI), em conjunto com a análise das questões objetivas do ENADE 2014 e do desempenho dos estudantes de Engenharia da Computação, comparando esses resultados com aqueles divulgados pelo INEP (2016).

Sendo assim, foi preciso excluir um item de conhecimentos gerais e cinco itens de conhecimentos específicos por apresentarem correlação ponto bisserial baixa ou não se ajustarem adequadamente ao modelo logístico de três parâmetros da TRI.

Quando isso acontece, gera um parâmetro com erro padrão elevado e precisa ser eliminado do teste ou nem gera um valor para sua estimativa. Possíveis motivos para este problema são má formulação do item, multidimensionalidade do teste e dependência entre os itens do teste.

Dos itens restantes (total de 20), o índice de facilidade e nível de dificuldade pela TRI mantiveram-se, na maioria dos casos, na mesma ordem. Porém, a vantagem de obter os parâmetros é que eles não se alteram devido à propriedade de invariância, desde que eles sejam adequadamente estimados. Pela TCT esses índices se alteram conforme se alteram os respondentes e não permite a comparação de estudantes ao longo do tempo, exceto se forem submetidos a mesma prova.

Além disso, uma vez criado um grande banco de itens pela TRI, pode-se aplicar os testes adaptativos computadorizados (CATs). Os CATs fornecem provas que são adaptadas ao nível de conhecimento dos estudantes, o que permite obter maior precisão em provas mais curtas, reduzindo a fadiga dos estudantes.

Com a escala construída é possível fazer sua interpretação, uma vez que o nível de dificuldade está na mesma escala do desempenho dos estudantes, evidenciando o que eles dominam de conteúdos e o que eles ainda precisam aprender. Para as escalas, sugere-se adicionar mais itens difíceis de FG e itens mais fáceis para CE.

Em relação ao desempenho dos estudantes, a maioria apresentou desempenho superior à média dos estudantes para FG, enquanto para

CE, a maioria apresentou desempenho inferior à média, evidenciando ser uma prova com maior nível de dificuldade para os respondentes. Além disso, há uma correlação positiva e forte entre os escores obtidos pela TRI e pela TCT (acima de 0,9), mas há uma relação fraca entre os escores em FG e CE pela TRI, isto é, ter um bom desempenho em FG, não significa ter um bom em CE, e vice-versa.

Por fim, os resultados evidenciam a importância de uma análise crítica na elaboração dos itens da prova, visto que é uma avaliação tão importante a nível nacional e serve para tomadas de decisão quanto à qualidade do Ensino Superior, uma vez que, dos 35 itens respondidos pelos estudantes, nove já haviam sido eliminados pelo INEP e mais seis itens foram eliminados pela análise TRI, restando 20 itens na prova.

Referências

ANDRADE, D.F.; TAVARES, H.R.; VALLE, R.C. *Teoria de Resposta ao Item: conceitos e aplicações*. São Paulo: ABE, 2000.

BAKER, F.B. *The basics of item response theory*. Washington, DC: ERIC Clearinghouse on Assessment and Evaluation, 2001.

BIRNBAUM, A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In LORD, F.M.; NOVICK, M.R. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley, 395-479, 1968.

CAMARGO, R. V. W.; CAMARGO, R. C. C.P.; ANDRADE, D. F.; BORNIA, A. C. Desempenho dos alunos de ciências contábeis na prova ENADE/2012: uma aplicação da Teoria da Resposta ao Item. REPeC, v. 10, n. 3, p. 332-355, 2016.

- EMBRETSON, S.; REISE, S. P. *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates, Inc., Publishers, 2000.
- HAMBLETON, R.K.; SWAMINATHAN, H.; ROGERS, J. *Fundamentals of Item Response Theory*. Measurement Methods for the Social Science. Newbury Park: SAGE Publications, Inc., 1991.
- INEP. ENADE 2014 – Exame Nacional de Desempenho dos Estudantes. Relatório de área. Engenharia de Computação, Versão: 07/04/2016. (2016). SINAES, 364p. Disponível em: http://download.inep.gov.br/educacao_superior/enade/relatorio_sintese/2014/2014_rel_engenharia_de_computacao.pdf. Acesso: ago. 2018.
- INEP. Censo da Educação Superior 2016. (2017). Disponível em: http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/mec-e-inep-divulgam-dados-do-censo-da-educacao-superior-2016/21206. Acesso em: ago. 2018.
- KLEIN, R. Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB). *Revista Meta: Avaliação*, [S.l.], v. 1, n. 2, p. 125-140, sep. 2009.
- LOPES, F. L.; VENDRAMINI, C. M. M. Propriedades psicométricas das provas de pedagogia do ENADE via TRI. *Avaliação*, v. 20, n. 1, p. 27-47, 2015.
- NOGUEIRA, S. O. ENADE: Análise de itens de formação geral e de estatística pela TRI. Dissertação de Mestrado, Programa de Pós-Graduação em Psicologia, Universidade São Francisco, Itatiba, 2008.
- 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.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Austria. <https://www.R-project.org>
- SOARES, T. M. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. *Pesquisa Operacional*, v. 25, n. 1, p. 83-112, 2005.
- 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, 2014.
- SPENASSATO, D.; TEZZA, R. Aplicação da Teoria da Resposta ao Item na análise da prova do ENADE 2014 - Engenharia da Computação. In: 8º MCSul/VIII SEMENGO, 2018, Rio Grande - RS. Anais do 8º MCSul / VIII SEMENGO, 2018. p. 610-624.
- ZIMOWSKI, M., MURAKI, E., MISLEVY, R., BOCK, D. Software BILOG-MG V3.0. Scientific Software International, Inc., 2003.