



Teoria da Resposta ao Item em Ambiente R como Suporte no Ensino e Aprendizagem de Engenharia*

Item Response Theory in R Software as Support in Engineering Teaching and Learning

Olga Maria Formigoni Carvalho Walter¹

Andréa Cristina Konrath²

Elisa Henning³

Antonio Cezar Bornia⁴

Resumo

Este artigo é um estudo exploratório e tem como objetivo investigar o uso dos pacotes *irtoys* e *itm*, específicos para análises de Teoria de Resposta ao Item (TRI) no ambiente R. Os dados utilizados são do Exame Nacional de Desempenho de Estudantes (ENADE) de 2014 do curso de Engenharia de Produção. O ENADE é um componente curricular obrigatório de avaliação dos cursos de graduação, no qual participam estudantes concluintes do curso. Os resultados demonstram que, de maneira geral, os dois pacotes oferecem as mesmas funcionalidades e se apresentam estáveis nas aplicações realizadas. Conclui-se que os pacotes investigados podem auxiliar nas análises e interpretações, constituindo-se como uma ferramenta útil para complementar o ensino e aprendizagem da TRI.

Palavras-chave: Teoria da Resposta ao Item. *itm*. *irtoys*. Ensino. Software R.

*Submetido em 14/02/2019 - Aceito em 10/04/2020

¹Universidade Federal de Santa Catarina – Doutoranda do Programa de Pós-Graduação em Engenharia de Produção., – olgaformigoni@gmail.com

²Universidade Federal de Santa Catarina – Departamento de Informática e Estatística., – andreack@gmail.com

³Universidade do Estado de Santa Catarina – Departamento de Matemática., – elisa.henning@udesc.br

⁴Universidade Federal de Santa Catarina – UFSC - Departamento de Engenharia de Produção e Sistemas., – cezar.bornia@ufsc.br

Abstract

This paper is an exploratory study that aims to investigate the use of `irtoys` and `ltm` packages, specific for Item Response Theory (IRT) analysis in the R software. The data used are from the 2014 National Examination of Student Performance (ENADE) applied to the undergraduate course in Production Engineering. ENADE is a compulsory curricular component of undergraduate courses, in which concluding students participate. The results show that, in general, the two packages offer the same functionalities and are stable in the applications studied. We concluded that both packages can aid in the analysis and interpretation, constituting a useful tool to complement the teaching and learning of IRT.

Keywords: Item Response Theory. `ltm`. `irtoys`. Teaching. R software.

1 INTRODUÇÃO

A sociedade vem passando por uma revolução tecnológica com grande desenvolvimento na área da informação, ampliando possibilidades em diversas áreas do conhecimento e afetando a sociedade como um todo.

O uso de tecnologias na sala de aula não garante mudanças nas formas de ensinar e aprender. A introdução de tecnologias deve permitir enriquecer o ambiente educacional, ampliando a maneira de construção de conhecimentos, buscando atuação crítica e criativa por parte de alunos e professores (MAINART; SANTOS, 2010).

Para otimizar o aprendizado, é necessário que o desempenho do aluno seja avaliado periodicamente, bem como a maneira como são transmitidas as informações. Essa prática permite o aperfeiçoamento no processo de ensino aprendizagem, contribuindo para a qualidade da educação. Nesse sentido, o meio de avaliação (teste, prova, exame), bem como os itens que o compõem, deve ser bem elaborado. Essa premissa serve tanto localmente, em sala de aula, como de forma mais ampla, por meio das avaliações em larga escala. Os resultados das avaliações em larga escala podem servir como norteadores no aprimoramento das políticas educacionais rumo à universalização da qualidade do ensino (SILVA, 2010).

A maneira tradicionalmente utilizada para avaliar o desempenho educacional é a Teoria Clássica de Testes (TCT), pela qual o traço latente é mensurado por meio do somatório dos pontos atribuídos a cada item do questionário, obtendo-se dessa forma, o *score* total do teste. Essa técnica tem como vantagem ser de fácil interpretação, não exigindo muitas suposições sobre os dados. Por outro lado, apresenta como limitação sua dependência do conjunto de itens que compõem o instrumento de medida, o que torna sua aplicação limitada (JUSTINO; ANDRADE, 2007). Porém, há uma técnica que apresenta vantagens sobre a TCT: a Teoria da Resposta ao Item (TRI). A 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 do respondente. A relação é expressa de tal forma que, quanto maior a habilidade do respondente, maior a probabilidade de acertar o item. Um dos aspectos positivos da TRI é permitir a comparação entre populações submetidas a provas diferentes que tenham alguns itens comuns. Isso é possível, pois, na TRI, os elementos centrais são os itens e não a prova como um todo (ANDRADE; TAVARES; VALLE, 2000).

No Brasil, a TRI foi aplicada pela primeira vez em 1995 na análise dos dados do Sistema Nacional do Ensino Básico (SAEB) (ANDRADE; TAVARES; VALLE, 2000). Atualmente, a TRI vem sendo bastante difundida, tanto no Brasil como no mundo todo, nas mais diversas áreas do conhecimento, principalmente na Educação e Psicologia (CASTRO et al., 2013; MOREIRA JUNIOR et al., 2015), sendo também aplicada para avaliação da sustentabilidade ambiental (TRIERWEILLER et al., 2013; VINCENZI et al., 2018).

Embora a maioria das aplicações esteja relacionada com as avaliações educacionais, a TRI também tem sido utilizada em muitas outras áreas, tais como na administração, nas Ciências da Saúde, na Psicologia e nas Engenharias (MOREIRA JUNIOR; SZINVELSKI, 2013).

Trabalhos publicados no Brasil que avaliam traços latentes nos mais diversos temas podem ser consultados em Moreira Junior (2010) e Castro et al. (2013).

Nesse sentido, devido à constante necessidade de melhorias no processo de ensino aprendizagem e na qualidade da educação e, considerando também a TRI como uma técnica em expansão no Brasil, este artigo tem como objetivo explorar a utilização de dois pacotes do ambiente R (irtoys e ltm) específicos para análise da TRI.

2 REFERENCIAL TEÓRICO

2.1 Exame Nacional de Desempenho dos Estudantes (ENADE)

O Exame Nacional de Desempenho dos Estudantes (ENADE) é uma prova escrita, aplicada anualmente, que avalia os concluintes dos cursos de graduação considerando conteúdos programáticos previstos nas diretrizes curriculares dos cursos, o desenvolvimento de competências e habilidades na formação geral e profissional, e o nível de atualização dos estudantes com relação à realidade brasileira e mundial. O objetivo é obter informações para avaliação dos cursos de ensino superior brasileiros (INEP, 2020).

A aplicação da prova é de responsabilidade do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e sua primeira edição ocorreu em 2004.

As áreas de avaliação do conhecimento para os cursos de bacharelado e licenciatura provêm das áreas divulgadas pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Já os eixos tecnológicos tem como base o Catálogo Nacional de Cursos Superiores de Tecnologia (CNCST), do Ministério da Educação (MEC) (INEP, 2020).

2.2 Teoria da Resposta ao Item (TRI)

A TRI é uma técnica estatística que surgiu para preencher algumas lacunas resultantes das limitações da TCT, teoria que usualmente era, e ainda é, utilizada nas avaliações.

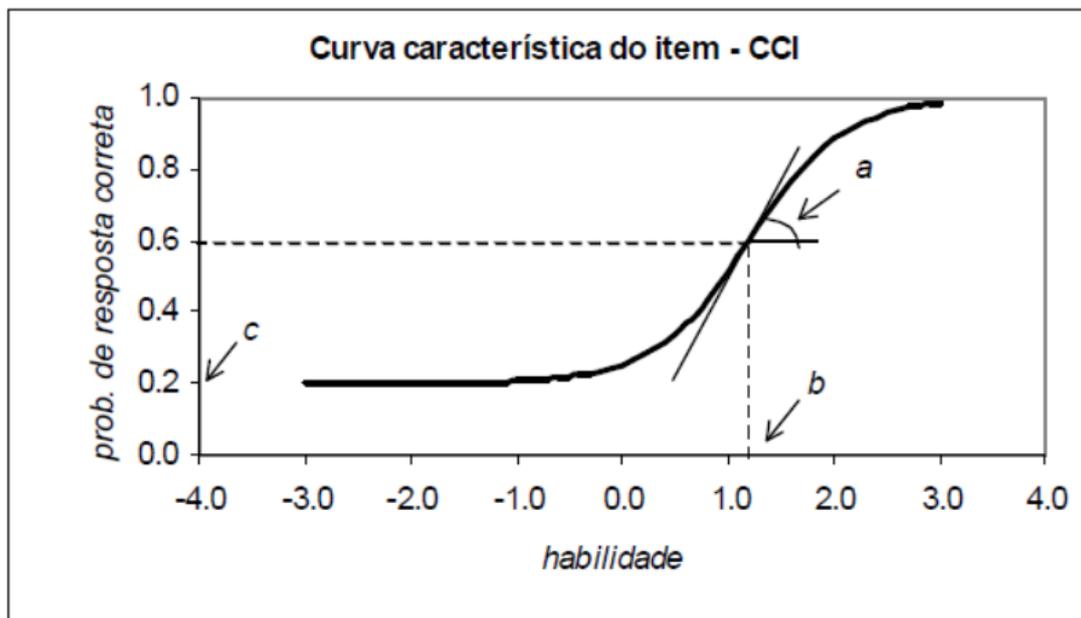
Os modelos da TRI podem ser classificados em função de três fatores: da natureza do item (dicotômico ou não dicotômico), do número de populações envolvidas e da quantidade de traços latentes que está sendo medida (unidimensionais ou multidimensionais) (ANDRADE; TAVARES; VALLE, 2000).

Em relação aos modelos dicotômicos unidimensionais, destacam-se os modelos logísticos de 1, 2 ou 3 parâmetros, ou quais caracterizam a dificuldade do item, a discriminação do item e a probabilidade de acerto ao acaso. O modelo logístico de 3 parâmetros (ML3), que será utilizado no presente estudo, é representado pela Equação (1) (ANDRADE; TAVARES; VALLE, 2000):

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

com $i = 1, 2, 3, \dots, I$ (itens propostos) e $j = 1, 2, \dots, n$ (indivíduos que compõe a amostra). U_{ij} corresponde a variável dicotômica que assume os valores 1 (um), quando o indivíduo j , no estudo em questão, acerta o item, ou assume 0 (zero) quando o indivíduo erra o item; θ_j representa o traço latente do j -ésimo indivíduo; $P(U_{ij} = 1|\theta_j)$ é a probabilidade de um indivíduo j com desempenho θ_j acertar o item i ; a_i representa o parâmetro de discriminação do item i ; b_i é o parâmetro de dificuldade do item i e c_i representa a probabilidade de acerto ao acaso, ou seja, representa a probabilidade de pessoas com baixo desempenho acertarem a resposta do item.

A Figura 1 apresenta uma Curva Característica do Item (CCI). Conforme Andrade, Tavares e Valle (2000), o modelo está baseado no fato de que indivíduos com maior habilidade têm uma probabilidade maior de acertar o item, sendo que a relação não é linear. O gráfico apresentado tem formato em “S”, sendo que os parâmetros dos itens definem a inclinação e deslocamento na escala de habilidade. A respectiva escala é uma escala arbitrária, na qual a importância são as relações de ordem existentes entre seus pontos e não essencialmente sua magnitude (ANDRADE; TAVARES; VALLE, 2000). Em relação à mensuração, o parâmetro b é mensurado na mesma unidade da habilidade, enquanto que o parâmetro c pode assumir valores entre 0 e 1, pois trata-se de uma probabilidade e não depende da escala. O parâmetro b representa a habilidade para uma probabilidade de acerto igual a $(1 + c)/2$. Logo, quanto maior o valor de b , mais difícil é o item (ANDRADE; TAVARES; VALLE, 2000). Conforme os autores, a probabilidade de um aluno com baixa habilidade responder corretamente o item, ou seja, a probabilidade de acerto ao acaso é representada pelo parâmetro c . O parâmetro a é proporcional à derivada da tangente da curva no ponto de inflexão; desse modo, itens com a negativo não são esperados sob esse modelo, pois baixos valores do parâmetro a indicam que o item tem pouco poder de discriminação e valores muito altos apontam itens com curvas características muito “íngremes”, os quais podem ser divididos em dois grupos: os que possuem habilidades abaixo do valor do parâmetro b , e os que possuem habilidades acima do valor do parâmetro b (ANDRADE; TAVARES; VALLE, 2000).

Figura 1 – Exemplo de uma curva característica do Item – CCI

Fonte: Andrade, Tavares e Valle (2000).

O gráfico da CCI permite analisar a relação entre a probabilidade e o desempenho no traço latente que está sendo medido, ou seja, possibilita verificar quanto um item contém de informação para a medida de habilidade (ANDRADE; TAVARES; VALLE, 2000; COUTO; PRIMI, 2011). De acordo com os autores, na TRI, a escala de habilidade pode assumir valores entre $-\infty$ e $+\infty$. Logo, torna-se necessário definir um ponto de partida e também uma unidade de medida para a elaboração da escala.

3 MATERIAIS E MÉTODOS

Este estudo é classificado como exploratório, pois investiga a utilização de dois pacotes do ambiente R (CORE TEAM, 2018): ltm (RIZOPOULOS, 2006) e irtoys (PARTCHEV, 2010), específicos para análises da TRI. Os dados utilizados são do Exame Nacional de Desempenho de Estudantes (ENADE) de 2014 do curso de Engenharia de Produção.

3.1 População e Amostra

Os dados utilizados para análise são informações públicas e estão disponíveis para *download* no portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP, 2018a). Tratam-se das respostas do Exame Nacional do Desempenho dos Estudantes (ENADE) do ano de 2014 dos estudantes concluintes do curso de Engenharia de Produção. Conforme a Portaria Normativa do MEC nº 840, de 24 de agosto de 2018, são considerados estudantes concluintes:

De cursos de bacharelado e licenciatura: aqueles que tenham integralizado 80% (oitenta por cento) ou mais da carga horária mínima do currículo do curso definido pela instituição de educação superior e não tenham colado grau até o último dia do período de retificação de inscrições, ou que tenham previsão de integralização da carga horária do curso até julho do ano subsequente da edição do Enade (MEC, 2018, Art. 44, par. 2º).

Na edição de 2014 do ENADE, a área de Engenharia de Produção contou com a participação de estudantes de 329 cursos, destacando-se a predominância das Instituições de Ensino Superior (IES) privadas, que concentraram 249 dos 329 cursos de Engenharia de Produção, número correspondente a 75,7% dos cursos avaliados. O Sudeste do Brasil foi a região de maior representação, com 205 cursos, ou 62,3% do total nacional. As regiões Sul e Nordeste tiveram representações, respectivamente, de 16,1% e de 13,7% do total de cursos. A região de menor concentração foi a Norte, com 11 cursos ou 3,3% do total, seguida pela região Centro-Oeste com 15 cursos (4,6%) (INEP, 2018b).

Para o curso em análise, inscreveram-se 17.307 estudantes, sendo que 2.225 deles não compareceram a prova. Essas observações, por se tratarem de alunos ausentes, foram eliminadas das análises.

Assim a amostra é composta por 15.082 estudantes, que em sua maior parte (68,8%) são do sexo masculino. O percentual de estudantes que se graduavam em IES públicas e cursaram todo o Ensino Médio em escolas públicas foi de 30,5%. Dos que se graduavam em IES Privadas, 58,6% fizeram o Ensino Médio em escolas públicas (INEP, 2018b).

3.2 Instrumento de Avaliação

O caderno de provas do ENADE 2014 do curso de Engenharia de Produção é composto por questões discursivas e de múltipla escolha (objetivas), de formação geral e do componente específico da área, e questões relativas à percepção do aluno quanto à prova. As questões estão distribuídas de acordo com a Tabela 1.

Tabela 1 – Questões do caderno de provas ENADE/2014 da área Engenharia de Produção

Partes	Número das Questões	Peso das questões no componente	Peso dos componentes no cálculo da nota
Formação Geral/Discursivas	D1 e D2	40%	25%
Formação Geral/Objetivas	1 a 8	60%	
Componente Específico/Discursivas	D3 e D5	15%	75%
Componente Específico/Objetivas	9 a 18	85%	
- Conteúdos básicos	19 a 35		
- Conteúdos profissionalizantes			
Questionário de Percepção da Prova	1 a 9	-	-

Fonte: ENADE (2014).

No Componente de Formação Geral, as 8 (oito) questões objetivas de múltipla escolha e as 2 (duas) discursivas tiveram pesos, respectivamente, iguais a 60% e 40%. As discursivas de Formação Geral foram corrigidas levando em consideração o conteúdo, com peso igual a 80%, e aspectos referentes à Língua Portuguesa (ortográficos, textuais, morfosintáticos e vocabulares), com peso igual a 20%. No Componente de Conhecimento Específico da Área de Engenharia de Produção, as 27 (vinte e sete) questões objetivas de múltipla escolha e as 3 (três) discursivas tiveram pesos iguais a, respectivamente, 85% e 15%. As notas dos dois Componentes, de Formação Geral e de Conhecimento Específico, foram então arredondadas à primeira casa decimal. Para a obtenção da nota final do estudante, as notas dos dois componentes foram ponderadas por pesos proporcionais ao número de questões: 25,0% para o Componente de Formação Geral e 75,0% para o Componente de Conhecimento Específico (INEP, 2018a).

Os itens que serão utilizados nas análises nesta pesquisa são as questões objetivas de formação geral e específica, com exceção do item 26, do conteúdo profissionalizante que foi anulado da avaliação, ou seja, serão analisados 34 (trinta e quatro) itens.

3.3 Análise dos dados

Para análise da TRI é necessário fazer uso de algum programa computacional. As análises foram realizadas com o *software* R (Core Team, 2017) utilizando o pacote ltm (RIZOPOULOS, 2006) e pacote irtoys (PARTCHEV, 2010).

Inicialmente os microdados coletados foram dicotomizados, ou seja, como as respostas dos microdados disponibilizados pelo INEP são apresentadas na forma de alternativas A, B, C, D ou E, essas foram comparadas ao gabarito. Assim as respostas corretas receberam o valor um e as respostas erradas foram substituídas por zero. Dessa forma, as respostas foram analisadas

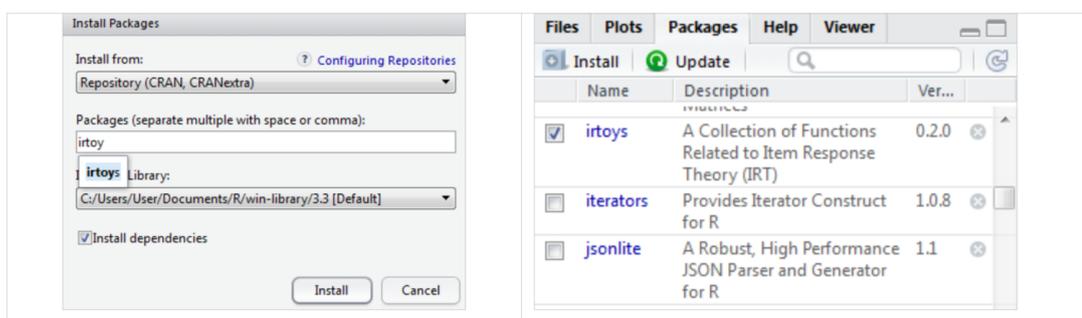
pelo modelo logístico de três parâmetros (ML3) já detalhado na seção 2.

Como já mencionado, a amostra é composta por 15.082 alunos que compareceram a prova, porém 12 deles entregaram a prova em branco. Assim, nas análises, as respostas desses estudantes foram consideradas como erradas.

4 DISCUSSÃO DA APLICAÇÃO DA TEORIA DA RESPOSTA AO ITEM

Essa seção tem por finalidade explorar a aplicação dos pacotes `irtoys` e `ltm` em análises na TRI. No R Studio, uma das maneiras de instalar um pacote é por meio do menu em “Tools / Install Packages...”. Conforme apresentado na Figura 2 (à esquerda), deverão ser digitado o nome do pacote desejado no campo “Packages” e pressionar o botão “Install”.

Figura 2 – Instalação e ativação do pacote `irtoys`



Fonte: Os autores (2018).

Se selecionada a opção “Install dependencies” (Figura 2 à esquerda), serão instalados pacotes adicionais dos quais o pacote `irtoys` depende. Depois de instalado, a ativação do pacote pode ser feita selecionando-o na guia “Packages” do R Studio, conforme Figura 2 (à direita).

4.1 Aplicação com o pacote `irtoys`

A calibração dos itens pode ser feita por meio da função “`est()`”. Essa função estima os parâmetros dos itens apenas na métrica logística (PARTCHEV, 2017). A Figura 3 apresenta os parâmetros dos itens obtidos para o ML3, que é o modelo adequado à prova analisada.

Figura 3 – Parâmetros dos itens estimados pelo pacote irtoys

```

Console C:/
> par.irtoys<-est(dados.enade, model = "3PL", nqp = 20, engine="ltm")
> par.irtoys
$est
      Dscrmn      Dffc1t      Gussng      Dscrmn      Dffc1t      Gussng
i1  0.4974104  0.12779270  0.1518863696  i18  0.8981111 -1.98198014  0.1359456271
i2  0.3262043  1.31553473  0.1264936256  i19  1.5194519 -0.05348985  0.1427958333
i3  1.3988171 -0.38141872  0.2930153726  i20  0.9734322  1.79559940  0.1398652993
i4  1.1103009 -1.40645613  0.3172810350  i21  1.1797452 -0.27727519  0.2381716648
i5  0.6572442  2.53063471  0.2574695916  i22  0.5151497  3.95215961  0.2006201666
i6  0.5205480 -0.82846096  0.0022659761  i23  0.9463370  1.71380588  0.1619443610
i7  0.9185791 -0.12910736  0.1575450432  i24  2.1536435  2.35929177  0.0812024621
i8  0.5449311 -2.96999525  0.0018137445  i25  1.0779910 -0.31420781  0.0157972938
i9  1.4450583  1.45333375  0.2003137016  i27 -1.2689766 -3.71527084  0.0535252042
i10 0.1106699 20.71548447  0.0562643345  i28  0.9621981 -1.22879809  0.0018987518
i11 1.1583399  0.18364757  0.1450115089  i29  0.9008472  1.55566491  0.1626555831
i12 3.0115256  2.51962865  0.3027845250  i30  1.0598069  1.26691806  0.2437284324
i13 0.9893926 -0.45183020  0.0254048990  i31  1.4325040  1.56994360  0.1873595749
i14 1.2020598  1.35264558  0.2113252644  i32  0.7767253  0.12273288  0.0006915424
i15 1.1952258  2.93930526  0.1439829445  i33  0.5968567  0.50075319  0.0099744651
i16 1.4931573  1.99278699  0.2953199195  i34 -0.3803472 -5.72794773  0.0300079930
i17 1.1072184 -0.48776669  0.0866845230  i35  1.8928114  3.05656597  0.1550032546

```

Fonte: Os autores (2018).

Nessa função “est()”, é necessário informar o conjunto de dados dicotomizados. No argumento “model=3PL” é informado que é um modelo logístico de três parâmetros; caso não for informado, assume-se como padrão um modelo logístico de dois parâmetros. O argumento “nqp=20” refere-se ao número de pontos de quadratura utilizados na estimação dos parâmetros. E o argumento “engine='ltm'” indica que a função utilizará recursos do pacote ltm. Outras opções são “BILOG” e “ICL”. No sistema operacional Windows, os arquivos executáveis (icl.exe para o ICL, blm1.exe, blm2.exe, e blm3.exe para o BILOG) devem estar no mesmo diretório dos objetos analisados (PARTCHEV, 2017).

Na coluna “Dscrmn”, “Dffc1t” e “Guessng” da Figura 3, são apresentados o parâmetro de discriminação a , o parâmetro de dificuldade b e o parâmetro de acerto ao acaso c . O item 12 apresenta o maior parâmetro de discriminação ($a_{12} = 3,01$) e o item 27 é o que menos discrimina ($a_{27} = -1,27$). O item 12 trata-se de uma questão do conteúdo básico, que avalia o conteúdo de transferência de calor (condução), equação geral da condução - Lei de Fourier e condição unidimensional em paredes planas. Esses conteúdos são abordados nos primeiros anos do curso. Por outro lado, o item 27 é do conteúdo profissionalizante, e o conteúdo avaliado compreende conceitos sobre o cálculo do lote econômico de compra. Apesar de não terem sido apresentados no enunciado da questão os parâmetros envolvidos no cálculo do lote econômico de compra, os estudantes, conseguiram resolver facilmente a questão, o que mostra que os estudantes fizeram a interpretação desse item adequadamente, ou seja, têm bons níveis de interpretação.

Os itens 27 e 34 apresentaram parâmetro a negativo. O item 10 possui $a_{10} = 0,11$ e $b_{10} = 20,72$, parâmetros que indicam que o item tem baixo poder de discriminação e é considerado difícil. Os itens 10, 27 e 34 devem ser eliminados de provas futuras por apresentar os comportamentos descritos anteriormente. Os itens devem ser elaborados de modo que seja possível diferenciar as pessoas que acertam das que erram o item. A elaboração de itens, além de outros fatores, pode ser considerada um dos principais agentes para a qualidade das avaliações, no en-

tanto, elaborar itens eficientes no processo avaliativo não é tarefa simples e nem sempre todos conseguem elaborar itens com qualidade pedagógica que garantam funcionalidade instrumental às avaliações (LAGE et al., 2020).

Desconsiderando os itens 27 e 34, o item mais fácil é o 8, com parâmetro de dificuldade $b_8 = -2,97$. O item 8 aborda, mediante um contexto histórico, as transformações que ocorreram no meio rural e urbano, a partir do século XX, decorrentes da aplicação de novas tecnologias às atividades industriais e agrícolas. Esse é um item que é facilmente respondido, pois está muito bem definido no contexto histórico da evolução da indústria e seu papel na transformação e no consumo, nas alterações e novas alternativas no modo de produzir, muito enfatizado durante todo o curso de Engenharia de Produção. O item 22 ($a_{22} = 0,5151$ e $b_{22} = 3,9522$) trata de logística reversa. Apesar de o conteúdo ser teoricamente simples para concluintes de cursos de Engenharia de Produção, aparentemente, tornou-se difícil devido à formulação da questão.

Com exceção dos itens 3, 4, 5, 12, 14, 16, 21 e 30, todos os demais apresentam o parâmetro de acerto casual c dentro do esperado, ou seja, valores inferiores a 20% para cinco alternativas de resposta. Se o valor for muito superior, pode ser um indicativo de que a opção correta se diferencia de algum modo das demais, atraindo estudantes de baixo desempenho (RABELO, 2013, p. 132).

Na Figura 4 estão os erros padrão dos parâmetros a (coluna 1), b (coluna 2) e c (coluna 3). Observa-se que para o item 10, não houve estimação pelo motivo do alto valor do parâmetro de dificuldade b , já discutido anteriormente.

Figura 4 – Erro padrão dos parâmetros dos itens

	[,1]	[,2]	[,3]		[,1]	[,2]	[,3]
[1,]	0.14447156	1.07246413	0.213529793	[18,]	0.09449035	0.61466747	0.265904782
[2,]	0.15165611	1.77778542	0.259210667	[19,]	0.08367260	0.05930021	0.026525608
[3,]	0.10414659	0.11824421	0.044770325	[20,]	0.09448758	0.06380209	0.016860396
[4,]	0.10146750	0.33630792	0.126464703	[21,]	0.08905246	0.13320038	0.047844532
[5,]	0.15075833	0.15786143	0.036709678	[22,]	0.20031854	0.57087191	0.045612323
[6,]	0.02290391	0.07417904	0.014642924	[23,]	0.09409731	0.06500225	0.018625972
[7,]	0.07967705	0.18165147	0.059600032	[24,]	0.21406470	0.07428745	0.003583702
[8,]	0.02785453	0.15288924	0.019811979	[25,]	0.06047696	0.10165393	0.043522007
[9,]	0.12674903	0.04208656	0.012398029	[26,]	0.22209962	0.34524002	0.003951815
[10,]	NaN	NaN	NaN	[27,]	0.03146456	0.05451123	0.019121584
[11,]	0.08031270	0.08864422	0.033762055	[28,]	0.09192715	0.06814698	0.022469627
[12,]	0.59794747	0.10588110	0.004421856	[29,]	0.10491737	0.06494521	0.021450561
[13,]	0.11292897	0.26818670	0.110105406	[30,]	0.11534111	0.04486150	0.010482247
[14,]	0.09935298	0.04933472	0.015222341	[31,]	0.02493764	0.02743570	0.004206709
[15,]	0.18353035	0.19089590	0.008257396	[32,]	0.03892121	0.13484832	0.036073936
[16,]	0.19515830	0.07493433	0.010389856	[33,]	0.26799608	1.83409114	0.079322224
[17,]	0.06966818	0.12830487	0.053625167	[34,]	0.33358277	0.19997433	0.004152307

Fonte: Os autores (2018).

Cabe lembrar que, na Figura 4, não consta o item 26 (item do conteúdo profissionalizante que foi anulado da avaliação). O item numerado como 26 na Figura 4 é na realidade o item 27. O item 27 é o item 28 e, assim, sucessivamente, até o item 34, que é, na realidade, o item 35 da prova analisada.

Com relação aos erros padrão do parâmetro a , verificam-se valores baixos, como nos itens: 6, 7, 8, 11, 14, 17, 18, 19, 20, 21, 23, 25, 28, 29, 32 e 33. Outros itens possuem valores

que fogem ao padrão dos citados anteriormente. É o caso, por exemplo, dos itens 12, 27, 34 e 35, que apresentaram os maiores valores para o erro padrão deste parâmetro. Erros padrão altos significam que as estimativas dos parâmetros não estão boas.

Os itens 4, 18, 22 e 27 apresentam valores de erro padrão do parâmetro b (coluna [,2]) entre 0,30 e 0,70. Já os itens 1, 2 e 34 apresentam erro padrão maior que 1 para esse parâmetro.

Avaliando os erros padrão do parâmetro de acerto casual c , os itens que apresentaram erro padrão com comportamento diferente dos demais são os itens 1, 2, 4, 13 e 18.

A estimativa da habilidade dos respondentes pode ser obtida por meio da função “`eap()`”. A Figura 5 apresenta as trinta primeiras estimativas das habilidades.

Figura 5 – Estimativa da habilidade dos respondentes

```

> score.irtoys<-eap(dados.enade, par.irtoys, qu=normal.qu())
> score.irtoys[1:30,]
  est      sem n      est      sem n      est      sem n
[1,] 1.9223193 0.4352129 34 [11,] 0.3376044 0.4812770 34 [21,] 0.7789778 0.4716150 34
[2,] 1.6519205 0.5052359 34 [12,] 0.7004165 0.4629984 34 [22,] 2.1320552 0.4765114 34
[3,] 0.2498560 0.4807250 34 [13,] 1.4214240 0.5154478 34 [23,] 1.8825842 0.4580464 34
[4,] 1.1059989 0.4530839 34 [14,] 0.8068479 0.4605202 34 [24,] 1.3697184 0.5158274 34
[5,] 0.6128637 0.4777144 34 [15,] 1.3030384 0.5164096 34 [25,] 2.0254426 0.4278252 34
[6,] 2.1031516 0.5256751 34 [16,] 1.4722671 0.4388820 34 [26,] 1.9203322 0.4579817 34
[7,] 1.9981929 0.5179791 34 [17,] 0.4383670 0.4757753 34 [27,] 0.2479591 0.4983486 34
[8,] 2.0163382 0.4331466 34 [18,] 1.2502160 0.4814954 34 [28,] 0.1802927 0.4804076 34
[9,] 1.4212865 0.4360252 34 [19,] 1.7249528 0.5295131 34 [29,] 0.5676990 0.4635924 34
[10,] 1.7108835 0.4578463 34 [20,] 1.9640372 0.5194168 34 [30,] 0.2634656 0.4934216 34

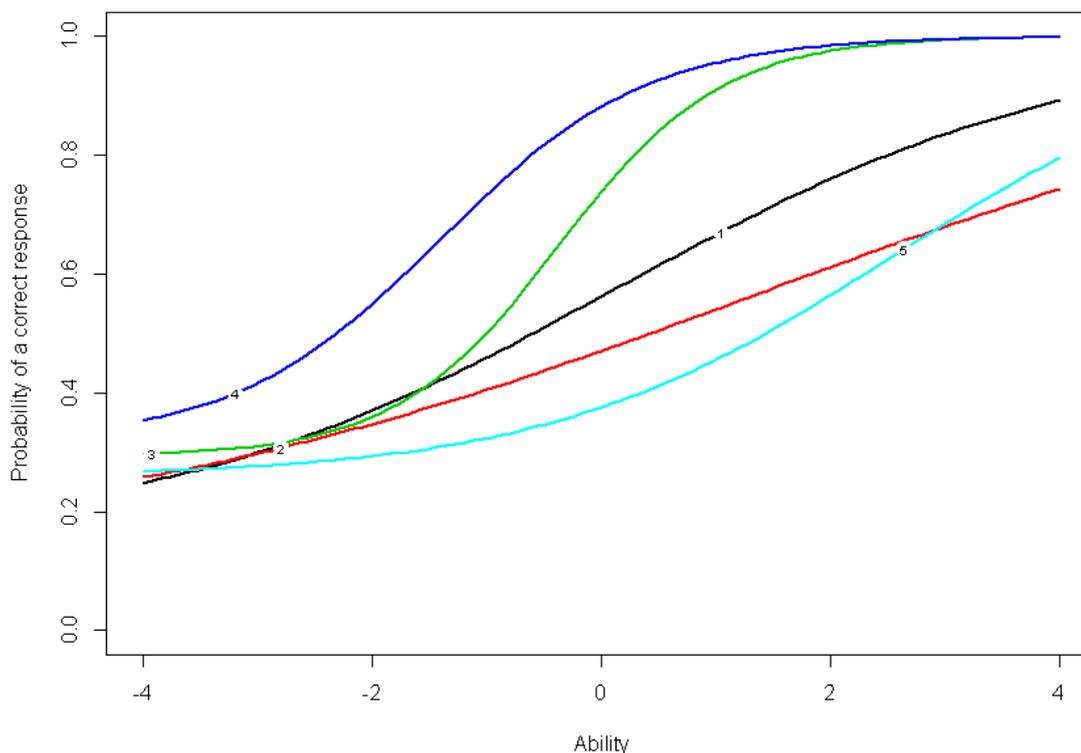
```

Fonte: Os autores (2018).

Na Figura 5, a primeira coluna é a estimativa da habilidade de cada respondente; a segunda, é o erro padrão; e a terceira coluna, o número de respostas de cada respondente. Verificam-se diferentes proficiências entre as trinta primeiras estimativas de habilidade dos estudantes. Por exemplo, o vigésimo segundo estudante apresentou proficiência $\theta_{22} = 2,13$, sendo considerado estudante com maior proficiência dentre os apresentados na Figura 5. Por outro lado, o terceiro estudante e o vigésimo sétimo estudante, apresentaram praticamente a mesma proficiência, com a habilidade θ em torno de 0,25. Isso significa pedagogicamente que o vigésimo sétimo estudante apresenta conhecimentos e habilidades superiores, quando comparado com o conhecimento e habilidades avaliados apresentados pelos dois outros estudantes que se posicionaram em $\theta = 0,25$.

O argumento “`qu=normal.qu()`” controla o número de pontos de quadratura. Na função “`eap()`”, é necessário informar o objeto que contém os parâmetros dos itens. As estimativas dos parâmetros estão no objeto “`par.irtoys`” e as estimativas das habilidades estão no objeto “`score.irtoys`”. Gráficos auxiliam nas análises do comportamento dos itens.

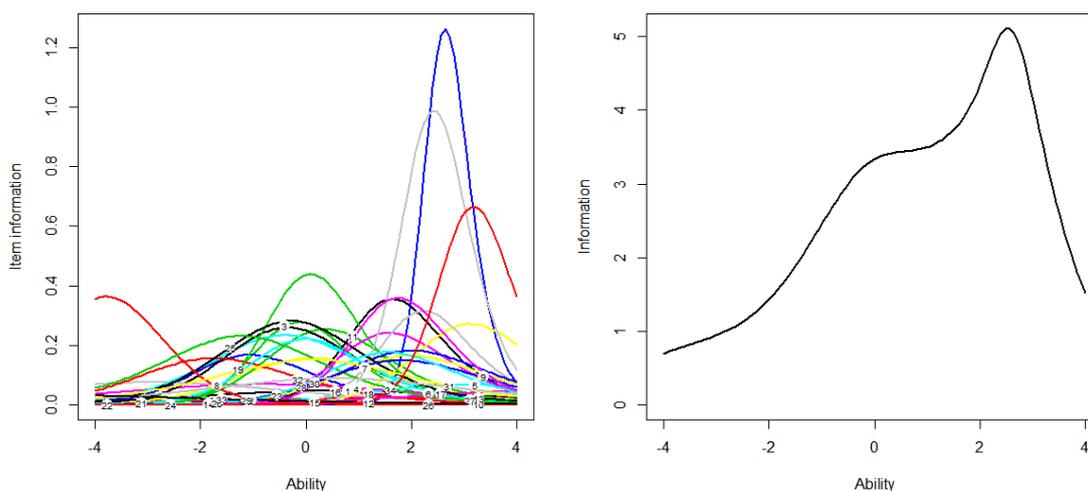
A Figura 6 apresenta a curva característica para os itens 1 até 5 obtidos por meio dos comandos “`plot(irf(par.irtoys, item=1:5),co=c(1:5), label=T)`”. No argumento “`item`”, serão informados os itens que terão a curva plotada no gráfico e, no argumento “`co`”, são informadas as cores das linhas das curvas.

Figura 6 – Curva característica dos cinco primeiros itens

Fonte: Os autores (2018).

O item 3 (curva verde) da Figura 6 é o item que mais discrimina e sua curva apresenta-se na forma de S. De fato, na Figura 3, dentre os itens 1 a 5, o item 3 é o que possui maior valor para o parâmetro de discriminação, com $a_3 = 1,3988$. O item 2 é que menos discrimina, possui $a_1 = 0,3262$ e apresenta-se com pouca inclinação (curva vermelha). Um baixo índice de discriminação indica que o item pode ter sido mal formulado, ou não ser adequado ao grupo em que foi aplicado, pois um item de avaliação educacional deve apresentar capacidade de diferenciar (discriminar) alunos com diferentes níveis de conhecimento. Dessa forma, os itens 3 e 4 mostraram essa capacidade, ao contrário dos itens 1, 2 e 5.

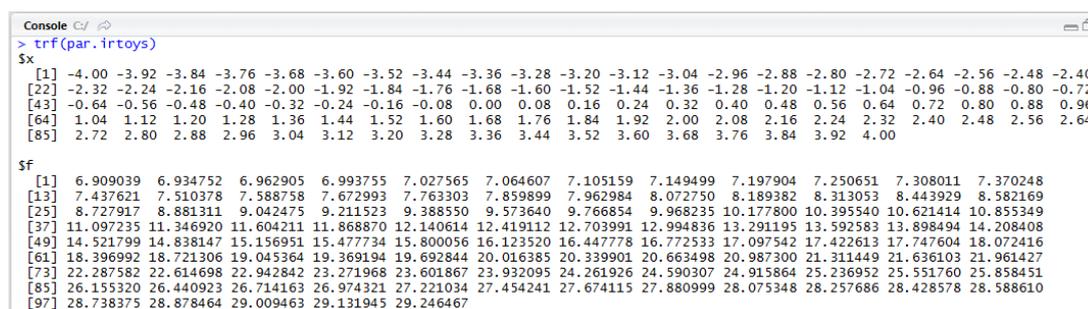
A Figura 7 exibe as curvas de informação dos itens e do instrumento obtidas pelas funções “iif()” e “tif()”, geradas por meio dos comandos “plot(iif(par.irtoys), label = TRUE, co=NA)” e “plot(tif(par.irtoys))”, respectivamente. É necessário informar, como argumento, os parâmetros dos itens, em ambas as funções. O argumento “label = TRUE” mostra a legenda dos itens e “co=NA” gera as linhas do gráfico coloridas.

Figura 7 – Curva de informação dos itens e do instrumento

Fonte: Os autores (2018).

A Figura 7 (à direita) revela que o instrumento possui mais informação no intervalo de habilidade em aproximadamente 0 a 3. Ou seja, o instrumento é adequado para avaliar a proficiência de alunos de Engenharia de Produção que possuem habilidades no intervalo de 0 a 3.

A função “trf ()” pode ser utilizada para visualizar a relação entre a habilidade dos respondentes e o *score*. Esta função tem como argumentos os parâmetros dos itens e os *scores*. Se não for informado o argumento *scores*, como na Figura 8, serão mostrados apenas cem valores uniformemente distribuídos no intervalo de habilidade -4,00 à 4,00. Na Figura 8, “x” representa a habilidade estimada e “f”, o *score* correspondente esperado.

Figura 8 – Habilidade dos respondentes e *score*

Fonte: Os autores (2018).

De acordo com a Figura 8, o aluno com *score* 6,9 tem menor habilidade (-4,00) e o aluno com maior habilidade (4,00) tem *score* 29,2.

4.2 Aplicação com o pacote ltm

O pacote ltm utiliza funções diferentes para estimar os parâmetros dos itens, dependendo do tipo de modelo. A função “rasch()” é utilizada para o ML1, ou seja, modelos Rasch que possuem somente o parâmetro de dificuldade c . A função “ltm()” é usada para ML2 parâmetros. A função “gpcm()” estima parâmetros do modelo de créditos parciais. A função “grm()” é empregada para ajustar o modelo de resposta gradual. A estimativa dos parâmetros dos itens para o ML3 é obtida com a função “tpm()”, conforme demonstra a Figura 9. O argumento “type= rasch” nessa função também permite estimar modelos de 1 parâmetro.

Figura 9 – Estimativa dos parâmetros dos itens por meio do pacote ltm

```

Console C:/
> par.ltm<-tpm(dados.enade)
> par.ltm

Call:
tpm(data = dados.enade)

Coefficients:
      Gussng  Dffc1t  Dscrmn      Gussng  Dffc1t  Dscrmn      Gussng  Dffc1t  Dscrmn
i1    0.148    0.109    0.495   i13    0.032   -0.436    0.996   i25    0.019   -0.308    1.082
i2    0.136    1.383    0.332   i14    0.211    1.352    1.201   i27    0.054   -3.712   -1.271
i3    0.294   -0.380    1.401   i15    0.144    2.937    1.197   i28    0.002   -1.229    0.962
i4    0.320   -1.399    1.112   i16    0.295    1.992    1.490   i29    0.162    1.554    0.899
i5    0.262    2.528    0.675   i17    0.087   -0.487    1.108   i30    0.244    1.267    1.061
i6    0.002   -0.828    0.521   i18    0.113   -2.034    0.891   i31    0.187    1.569    1.433
i7    0.157   -0.130    0.919   i19    0.143   -0.053    1.520   i32    0.001    0.123    0.777
i8    0.002   -2.968    0.545   i20    0.140    1.795    0.972   i33    0.007    0.489    0.594
i9    0.201    1.453    1.448   i21    0.237   -0.280    1.179   i34    0.027   -5.796   -0.370
i10   0.064   19.777    0.121   i22    0.204    3.915    0.531   i35    0.155    3.061    1.884
i11   0.144    0.182    1.157   i23    0.162    1.713    0.947
i12   0.303    2.522    2.951   i24    0.081    2.356    2.160
Log.Lik: -278866.2

```

Fonte: Os autores (2018).

Na Figura 9, o objeto “par.ltm” contém os parâmetros dos itens. Na coluna “Gussng”, é exibido o parâmetro de acerto casual c . O parâmetro de dificuldade b está na coluna “Dffc1t” e, na coluna “Dscrmn”, é apresentado o parâmetro de discriminação a .

Por meio da função “names()”, é possível verificar as estatísticas calculadas pela função “ltm()”, conforme Figura 10.

Figura 10 – Estatísticas da função “ltm()”

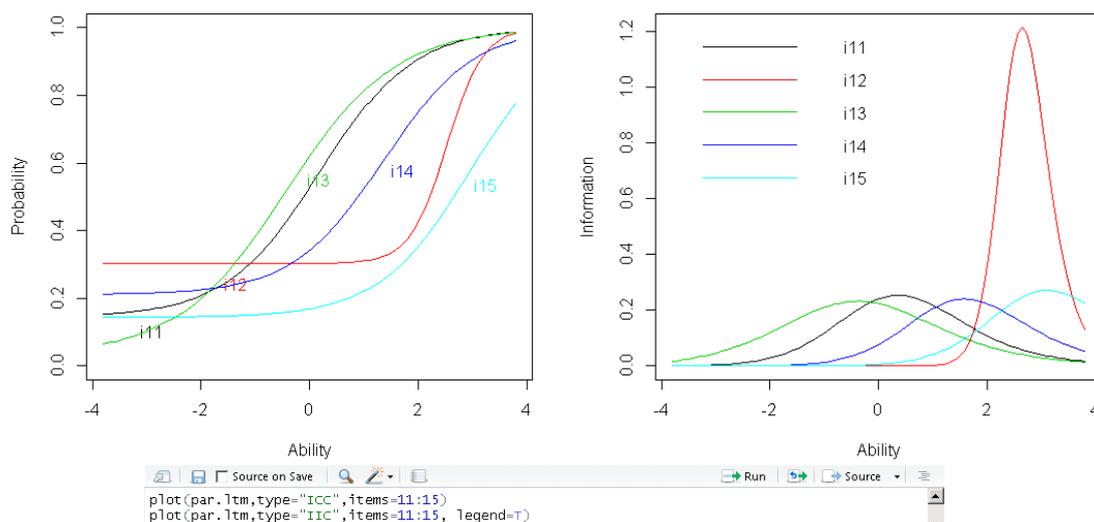
```

Console C:/
> names(par.ltm)
[1] "coefficients" "log.Lik"      "convergence"  "hessian"      "counts"       "patterns"     "GH"           "max.sc"
[9] "type"         "constraint"   "max.guessing" "IRT.param"    "x"            "control"      "na.action"    "call"

```

Fonte: Os autores (2018).

Na Figura 11, são apresentadas algumas estatísticas do objeto “par.ltm” obtidas por meio da função “summary()”. Além das estimativas dos parâmetros e respectivos erros padrão, estatísticas relativas aos parâmetros (z-values) e informações para avaliar o ajuste do modelo (log.lik, AIC e BIC) são apresentadas. Características dos métodos numéricos empregados na estimação e informações sobre o processo de otimização também podem ser visualizadas.

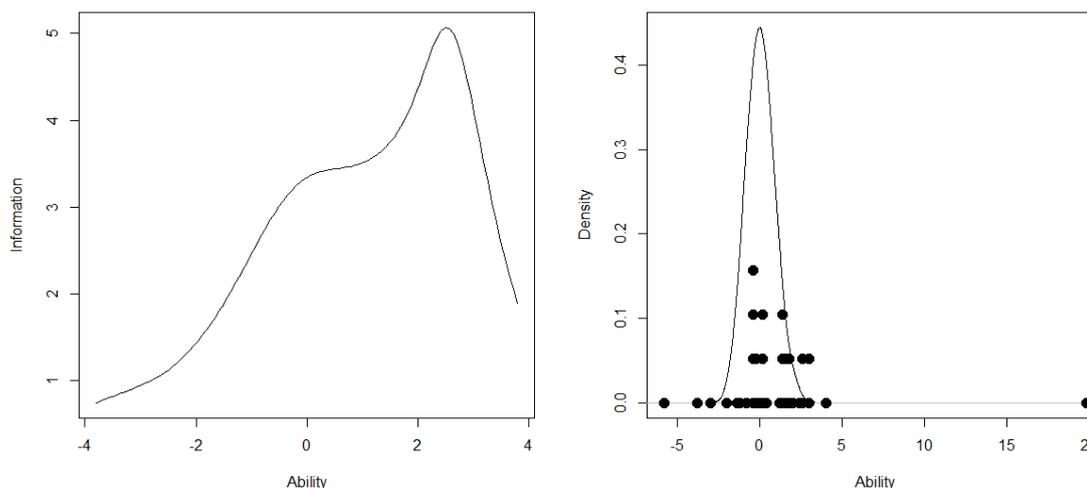
Figura 13 – Curva característica e de informação dos itens 11 à 15

Fonte: Os autores (2018).

Comandos para gerar os gráficos também são apresentados. Na Figura 13, a função “plot()” tem como argumentos os parâmetros dos itens (objeto “par.ltm”) e os itens que devem ser plotados.

Dentre os itens da Figura 13 à direita, o item 12 (curva vermelha) é o que possui maior quantidade de informação no intervalo de habilidade 2 a 3. Demais itens possuem quantidades praticamente iguais de informação, porém em intervalos de habilidades diferentes. Por exemplo, o item 13 (curva verde) e o item 15 (curva azul clara) possuem maior quantidade de informação entre os intervalos de habilidade -2 a 1 e 2 a 4, respectivamente. A função de informação do item mostra o intervalo no qual o item mais contribui para a estimativa do nível de habilidade dos estudantes.

Para gerar a curva de informação do instrumento (Figura 14 à esquerda), o argumento “items” da função “plot()” deve ser igual a zero. Nota-se maior quantidade de informação do instrumento aproximadamente no intervalo de habilidade de 1 a 3. É possível verificar o posicionamento dos itens na escala da prova (Figura 14 à direita) por meio do comando “plot(score.ltm, include.items=T)”. Na função “plot()”, utilizou-se como argumento os *scores* gerados pela função “factor.scores()”. Na Figura 14, é possível identificar claramente os itens 10 e 34 que na Figura 9 (estimação de parâmetros pelo pacote ltm), apresentaram os seguintes valores para o parâmetro de dificuldade b , 19,7768 e -5,7958, respectivamente.

Figura 14 – Curva de informação do instrumento e posicionamento dos itens na escala

Fonte: Os autores (2018).

A função “information()” calcula a quantidade de informação do teste ou item (área sob as curvas de informação do item ou instrumento), em uma faixa especificada. Na Figura 15, é dada a quantidade de informação para o item 12.

Figura 15 – Quantidade de informação do item 12

```

Console
> information(par.ltm, range = c(-4, 0), item=12)
Call:
tpm(data = dados.enade)
Total Information = 1.42
Information in (-4, 0) = 0 (0%)
Based on items 12

> information(par.ltm, range = c(-4, 4), item=12)
Call:
tpm(data = dados.enade)
Total Information = 1.42
Information in (-4, 4) = 1.39 (98.18%)
Based on items 12

```

Fonte: Os autores (2018).

Na Figura 15, à esquerda no intervalo de -4 a 0 (argumento “range”), o item não possui informação e, na mesma figura à direita, no intervalo de -4 a 4, o item apresenta 98,18% de informação. De fato, conforme a Figura 13, que apresenta de maneira gráfica o que foi calculado na Figura 14, o item 12 não possui informação no intervalo de -4 a 0.

4.3 Proficiência dos alunos no ENADE/2014

Para avaliar a proficiência dos estudantes, foi elaborada uma escala para avaliação dos níveis de desempenho (proficiência). As estimativas dos parâmetros dos itens e dos *scores* dos estudantes foram estimadas na mesma métrica, na escala de média zero e desvio padrão igual a um ($\mu = 0$; $\sigma = 1$). O objetivo da escala de proficiência é gerar uma interpretação qualitativa dos parâmetros encontrados pelo modelo, a fim de gerar um significado prático aos valores obtidos.

Para construir a escala, determinam-se os níveis âncoras, que são os pontos da escala a serem interpretados pedagogicamente. Os níveis âncoras caracterizam-se por um conjunto de

itens âncoras, que possuem propriedades matemáticas relacionadas ao índice de discriminação (a) e de dificuldade (b) (ANDRADE; TAVARES; VALLE, 2000).

Um item é considerado “âncora” quando é respondido corretamente pela maior parte dos respondentes (65%) com a proficiência indicada no nível da escala e por no máximo, 50% daqueles que possuem um nível de proficiência inferior. Ainda deve haver uma diferença de 30% entre respondentes com diferentes níveis de habilidade que acertaram o item (ANDRADE; TAVARES; VALLE, 2000). Os itens que atenderam simultaneamente às três condições anteriores foram considerados como itens “âncora”, e como “tipo âncora” aqueles que atenderam a pelo menos duas das condições.

A identificação de itens “âncora” ou “tipo âncora” não é realizada por funções específicas do R, mas é possível desenvolver rotinas específicas para tal finalidade. Para encontrar os níveis de proficiência é necessária a intervenção dos pesquisadores e profissionais envolvidos em razão dos aspectos qualitativos inerentes.

A análise resultou na identificação de três níveis de proficiência representados pelos itens do Quadro 1. O nível 1 compreende aspectos como interpretação de texto e gráficos para solucionar problemas de baixa complexidade. Além das habilidades do nível anterior, no nível 2, são exploradas habilidades voltadas a aspectos regulatórios das atividades produtivas, além de acrescentar aspectos estatísticos e lógicos para solução de questões que exigem raciocínio quantitativo. Já no nível 3, a escala agrega conhecimentos dos níveis anteriores e o aluno tem a habilidade global, com capacidade para entender, aplicar e desenvolver métodos para incrementar a performance e eficiência da cadeia produtiva de produtos e serviços de uma organização, acrescentando também, a competência de visão sistêmica.

Quadro 1 – Encontros e temáticas da unidade de aprendizagem sobre EA realizada no CC

Nível	Proficiência	Interpretação dos Níveis da Escala	Itens	% de alunos
1	$-4 \leq \theta < 0$	Capacidade de explorar aspectos como interpretação de texto e gráficos que permitem solucionar problemas de baixa complexidade.	3, 4, 6, 7, 8, 13, 17*, 18, 19, 21*, 25, 27, 28, 34	50,38
2	$0 \leq \theta < 4$	Habilidades são voltadas para aspectos regulatórios das atividades produtivas, além de agregar aspectos estatísticos e lógicos para solução de problemas com raciocínio quantitativo.	1, 2, 5, 9*, 11, 12, 14, 15, 16*, 20, 22, 23, 24**, 29, 30, 31*, 32, 33, 35	49,61
3	$3 \leq \theta < 5$	Aptidão de desenvolver uma visão global, com competência para entender, aplicar e desenvolver métodos para melhorar o desempenho ao longo de toda a cadeia produtiva de produtos e serviços, ou seja, possui visão sistêmica.	10	0,01

(*): Item tipo âncora; (**): Item âncora.

Fonte: Os autores (2018).

Aproximadamente metade dos estudantes estão concentrados na proficiência em que $\theta < 0$, ou seja, 50,48% dos estudantes encontram-se abaixo da proficiência média e posicionados no nível 1 da escala. Se a metade dos estudantes situam-se no nível 1, significa que possuem proficiências baixas e sabem apenas o mínimo necessário. Assim, metade dos estudantes que realizaram a prova do ENADE de 2014 do curso de Engenharia de Produção não possuem competências orientadas para resolução de problemas complexos, e não têm aptidão para o desenvolvimento e análise de processos que requerem visão sistêmica.

5 CONCLUSÕES E CONSIDERAÇÕES FINAIS

A Teoria da Resposta ao Item (TRI) é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da habilidade do respondente.

A TRI pode ser aplicada nas diferentes áreas de conhecimento; todavia exige o auxílio de recursos computacionais. Dessa forma, o presente artigo buscou explorar a utilização de dois pacotes do ambiente R (irtoys e ltm) específicos para análises de TRI aplicadas aos dados do ENADE de 2014 do curso de Engenharia de Produção.

Conforme demonstrado, os dois pacotes oferecem as mesmas funcionalidades, e se apresentaram estáveis nas aplicações realizadas.

A TRI representa um avanço para todas as áreas do conhecimento e algumas de suas vantagens são fundamentais para facilitar o controle e gestão de recursos e processos (TEZZA; BORNIA, 2009), principalmente no que tange à avaliação educacional. A aplicação da TRI pode auxiliar educadores e gestores na busca da melhoria contínua da qualidade dos cursos.

Os parâmetros dos itens indicam características desejáveis dos mesmos. Particularmente, o parâmetro de discriminação indica a "qualidade" do item, considerando que, para haver uma boa estimativa do conhecimento do estudante, a probabilidade de acerto deve crescer consistentemente com o aumento do conhecimento.

A curva de informação do instrumento mostra o intervalo no qual as estimativas apresentam maior precisão. Isso possibilita visualizar quais intervalos de habilidades (em termos de dificuldade) não estão sendo contemplados pelo instrumento, ou seja, que necessitam mais itens.

A interpretação da escala é possível com o uso de itens âncoras, dando significado pedagógico aos *scores* obtidos pelos estudantes. A escala revela o desempenho de cada aluno em intervalos que apontam o nível de proficiência dos conhecimentos aprendidos. Essas informações são úteis, por exemplo, para que docentes e universidades possam debater ações que melhorem a aprendizagem dos alunos, além de poder focar nos conteúdos que não foram suficientemente compreendidos pelos estudantes.

Os pacotes investigados auxiliam nas análises e interpretações, constituindo-se como uma ferramenta útil para complementar o ensino e aprendizagem da TRI.

O uso de tecnologias como o *software* R, no ensino de disciplinas como a TRI, é importante para incrementar o interesse e estimular os alunos, tornando a aprendizagem mais significativa. Além disso, é possível tornar a aula mais dinâmica, facilitando a explicação dos conteúdos. É importante também que docentes estejam se aperfeiçoando quanto à atualização de pacotes no R, bem como explorando novos pacotes desenvolvidos além dos aqui mencionados. Entende-se que o simples uso das tecnologias não significa contribuição pedagógica, se for empregada apenas como suporte para ilustrar a aula. A tecnologia deve ser aplicada como intermediação no processo de aprendizagem, impulsionando a construção do conhecimento pelo aluno, para obter-se aperfeiçoamento no processo ensino-aprendizagem.

A aplicação da TRI na análise da prova do ENADE de 2014 mostra algumas fragilidades na construção do instrumento de avaliação. É essencial que os itens sejam cuidadosamente elaborados de modo que consigam atingir os objetivos propostos. Ainda, a análise realizada permitiu reconhecer algumas deficiências no que tange às habilidades dos acadêmicos. Profissionais da educação sejam gestores ou professores, podem a partir dos resultados elaborar

ações para melhorar o processo ensino aprendizagem.

Como continuidade desta pesquisa, sugere-se fazer a análise de dimensionalidade do instrumento e ampliar as análises com outros pacotes como, por exemplo, o mirt (CHALMERS, 2012).

REFERÊNCIAS

- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. Teoria da resposta ao item: conceitos e aplicações. **São Paulo: ABE - Associação Brasileira de Estatística**, 2000.
- CASTRO, R. B. de et al. Teoria da resposta ao item: levantamento exploratório da produção científica. In: CONGRESSO BRASILEIRO DE TEORIA DA RESPOSTA AO ITEM. 3., 2013, Belém. **Anais [...]**. Belém: ABAVE, 2013. p. 129–142.
- CHALMERS, R. P. Mirt: A multidimensional item response theory package for the r environment. **Journal of Statistical Software**, v. 48, n. 6, p. 1–29, 2012.
- CORE TEAM, R. **R: A language and environment for statistical computing**. Vienna, Austria. Disponível em: <<https://www.R-project.org>>. Acesso em: 02 abr. 2018.
- COUTO, G.; PRIMI, R. Teoria de resposta ao item (tri): conceitos elementares dos modelos para itens dicotômicos. **Boletim de Psicologia**, v. 61, n. 134, p. 1–15, 2011.
- INEP. - **INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA**. 2018a. Disponível em: <<http://portal.inep.gov.br/microdados>>. Acesso em : 02 abr. 2018a.
- INEP. - **INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA**. ENADE. Versão 07/04/2016- Relatório de Área - Engenharia de Produção, 2018b. Disponível em: <http://download.inep.gov.br/educacao_superior/enade/relatorio_sintese/2014/2014_rel_engenharia_de_producao.pdf>. Acesso em : 02 abr. 2018b.
- INEP. - **INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA**. Exame Nacional de Desempenho dos Estudantes (Enade), 2020. Disponível em: <<http://inep.gov.br/enade>>. Acesso em : 10 jan. 2020.
- JUSTINO, G.; ANDRADE, D. F. de. Software para avaliação de aprendizagem utilizando a teoria da resposta ao item. In: CONGRESSO DA SBC, 27., 2007. Rio de Janeiro. **Anais [...]**. Rio de Janeiro: SBC, 2007. p. 248–255.
- LAGE, A. L. D. et al. **Guia de elaboração e revisão de questões e itens de múltipla escolha**. 2020. Disponível em: <[http://www.adventista.edu.br/_imagens/area_academica/files/guia-de-elaboracao-de-itens-120804112623-phpapp01\(3\).pdf](http://www.adventista.edu.br/_imagens/area_academica/files/guia-de-elaboracao-de-itens-120804112623-phpapp01(3).pdf)>. Acesso em: 10 jan. 2020.
- MAINART, D. de A.; SANTOS, C. M. A importância da tecnologia no processo ensino-aprendizagem. In: CONGRESSO VIRTUAL BRASILEIRO DE ADMINISTRAÇÃO. 7., 2010. **Anais [...]**, 2010. p. 129–142. Disponível em: <http://www.convibra.com.br/upload/paper/adm/adm_1201.pdf>. Acesso em: 23 abr. 2017.
- MEC. - **MINISTÉRIO DA EDUCAÇÃO**. Portaria Normativa MEC nº 840/2018. Disponível em: <http://download.inep.gov.br/educacao_superior/avaliacao_institucional/legislacao_normas/2018/portaria_normativa_GM-MEC_n840_de_24082018.pdf>. Acesso em: 10 jan. 2020.
- MOREIRA JUNIOR, F. de J. Aplicações da teoria da resposta ao item (tri) no brasil. **Revista Brasileira de Biometria**, v. 28, n. 4, p. 137–70, 2010.

MOREIRA JUNIOR, F. de J.; SZINVELSKI, C. Ro. P. Utilização da teoria da resposta ao item na análise dos dados da autoavaliação. Institucional da Universidade Federal de Santa Maria: Vantagens e Oportunidades, 2013. Disponível em: <http://download.inep.gov.br/educacao_superior/avaliacao_institucional/seminarios_regionais/trabalhos_regiao/2013/sul/eixo_2/utilizacao_teorica_resposta_item.pdf>. Acesso em: 23 mar. 2018.

MOREIRA JUNIOR, F. de J. et al. Avaliação da satisfação de alunos por meio do modelo de resposta gradual da teoria da resposta ao item. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 23, n. 86, p. 129–158, 2015.

PARTCHEV, I. **Package ‘irtoys’. Version 0.2.0**. Disponível em: <<https://cran.r-project.org/web/packages/irtoys/irtoys.pdf>>. Acesso em: 23 jan. 2018.

RABELO, M. Avaliação educacional: fundamentos, metodologia e aplicações no contexto brasileiro. **Rio de Janeiro: SBM**, 2013.

RIZOPOULOS, D. ltm: An r package for latent variable modeling and item response theory analyses. **Journal of statistical software**, v. 17, n. 5, p. 1–25, 2006.

SILVA, I. F. O sistema nacional de avaliação: características, dispositivos legais e resultados. **Estudos em avaliação educacional**, v. 21, n. 47, p. 427–448, 2010.

TEZZA, R.; BORNIA, A. C. Teoria da resposta ao item: vantagens e oportunidades para a engenharia de produção. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO. 29., 2009, Salvador. **Anais [...]**. Salvador: ABEPRO, 2009. p. 1–13.

TRIERWEILLER, A. C. et al. Measuring environmental management disclosure in industries in brazil with item response theory. **Journal of Cleaner Production**, v. 47, p. 298–305, 2013.

VINCENZI, S. L. et al. Assessment of environmental sustainability perception through item response theory: A case study in brazil. **Journal of Cleaner Production**, v. 170, p. 1369–1386, 2018.