

Características Psicométricas de uma Avaliação de Matemática¹

Psychometric Characteristics of a Mathematical Evaluation

Denilson Junio Marques Soares²

Paulo César Emiliano³

Talita Emidio Andrade Soares⁴

RESUMO

O Departamento de Matemática da Universidade Federal de Viçosa tem realizado algumas medidas de prevenção à reprovação dos estudantes na disciplina de Cálculo, cujos índices têm assustado os professores. Uma dessas medidas está na elaboração de avaliações de matemática básica que objetivam identificar estudantes propensos à reprovação, a fim de proporcionar uma intervenção pedagógica capaz de reverter essa situação. Tendo em vista a importância dessa avaliação como parte integrante dos processos de ensino e aprendizagem da disciplina, o presente artigo tem como objetivo oferecer uma análise estatística pautada nas duas vertentes da Psicometria moderna: a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI), além de uma análise pedagógica dos descritores e distratores de alguns itens-chave que compõem uma dessas avaliações escolhida para um estudo de caso. Os resultados apontaram para uma avaliação composta por itens com variados índices de dificuldade e, no geral, com bom poder discriminativo tanto pela análise via TCT, quanto pela TRI, cujos resultados foram obtidos por meio da estimação, pelo método da média a posteriori, de um modelo logístico de dois parâmetros. A consistência

1. Pretende-se, neste artigo, divulgar os resultados obtidos em uma pesquisa de Mestrado que foi desenvolvida no Programa de Pós-graduação em Estatística Aplicada e Biometria (Linha de Pesquisa: Métodos Estatísticos) da Universidade Federal de Viçosa.

2. Doutorando no Programa de Pós-Graduação em Educação da Universidade Federal do Espírito Santo (UFES). Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Viçosa (UFV). Professor EBTT do Instituto Federal de Minas Gerais – *Campus Piumhi*. E-mail: denilson.marques@ifmg.edu.br.

3. Mestranda no Programa de Pós-Graduação em Educação da Universidade Federal do Espírito Santo (UFES). Licenciada em Matemática pela Universidade Federal de Viçosa (UFV). E-mail: talitaeandrade@gmail.com.

4. Doutor em Estatística e Experimentação Agropecuária pela Universidade Federal de Lavras (UFLA). Professor do Departamento de Estatística da Universidade Federal de Viçosa (UFV). E-mail: paulo.emiliano@ufv.br.

interna da avaliação como instrumento de mensuração de habilidades foi verificada pelo coeficiente alpha de Cronbach, assegurando a qualidade e confiabilidade dos resultados desse estudo. Espera-se que esse trabalho sirva como um instrumento de difusão das teorias psicométricas na análise da estrutura de avaliações e aponte para a necessidade de uma maior discussão pedagógica acerca das possíveis lacunas existentes na aprendizagem de matemática básica dos estudantes do ensino superior.

Palavras-chave: *Psicometria; Teoria clássica dos testes; Teoria de resposta ao item; Análise de itens; Avaliação educacional.*

ABSTRACT

The Mathematics Department from the Federal University of Viçosa (Universidade Federal de Viçosa) has carried out some actions to avoid the students' failure in the Calculus course, whose indexes have startled the professors. One of these actions is the elaboration of basic math evaluations which aim to identify students prone to fail, in order to create a pedagogical intervention capable of reversing this situation. Considering the importance of this evaluation as an integral part of the course teaching and learning processes, the present work aims to offer a statistical analysis based on the two modern psychometrics aspects: The Classical Test Theory (CTT) and the Item Response Theory (IRT). It also presents descriptors and distractors pedagogical analysis of some key items that compose one of these evaluations, chosen for a case study. The results pointed to an evaluation composed of items with varying degrees of difficulty and in general with good discriminative power both by the CTT analysis and by the IRT, whose results were obtained by means of estimation, through the posteriori average method, a two parameters logistic model. The evaluation's internal consistency as a measuring skills instrument was verified by the Cronbach alpha coefficient, assuring the quality and reliability of these study results. It is hoped that this work may be an instrument for the diffusion of psychometric theories in the analysis of the structure of evaluations and point to the need for a greater pedagogical discussion about the possible gaps in the basic mathematics learning of students from Higher Education.

Keywords: *Psychometry; Classical Test Theory; Item Response Theory; Items Analysis; Educational Evaluation.*

Introdução

O aumento no índice de reprovação dos estudantes é uma realidade em todos os níveis de ensino e tem sido objeto de reflexões e estudos de pesquisadores e educadores das mais diversas áreas. No ensino superior, esse aumento tem sido um obstáculo enfrentado por diversas instituições por estar diretamente relacionado a retenção e a evasão, gerando grande

desperdício de recursos financeiros, carência de profissionais capacitados no mercado de trabalho e pondo fim ao sonho de muitos estudantes.

Veloso e Almeida (2013), por exemplo, trazem algumas reflexões acerca do problema da evasão que a Universidade Federal de Mato Grosso (UFMT) tem enfrentado, sugerindo que o problema é um fenômeno institucional, reflexo da ausência de uma política de permanência do aluno no curso de sua opção. Piacentini (2012) traz um estudo de caso referente aos índices de reprovação, abandono e evasão de um curso da Universidade Tecnológica Federal do Paraná (UTFPR) e sugere a criação de programas de tutoria e nivelamento, além de auxílios financeiros para estudantes carentes, visando estacionar o aumento destes índices. Vasconcelos, Almeida e Monteiro (2009) fazem esta mesma análise tendo como objeto de pesquisa estudantes dos cursos de engenharia da Universidade do Minho, em Portugal e concluem que, entre outros fatores, a distância familiar e a matrícula em curso de graduação que não corresponde à primeira opção de escolha também corroboram para este aumento. Para Rosa e Ribeiro (2018),

é de suma importância identificar os alunos potenciais de retenção e evasão e dar-lhes um suporte. Isso porque a decisão por prolongar o curso ou abandonar os estudos não é simples e imediata, mas decorre de um conjunto de fatores em um processo longo. Portanto, acredita-se que intervenções por parte do corpo discente e administrativo da universidade podem colaborar para reter o aluno. É preciso, pois, identificar os sinais de que o aluno vai se evadir para que se possa atuar de maneira preventiva, evitando o atraso e a interrupção da trajetória acadêmica (p. 201).

Embora se configure como uma disciplina aplicável em diversas áreas do conhecimento, o Cálculo tem um incômodo índice de reprovação e tem sido um dos grandes desafios dos cursos de graduação que o contêm em sua estrutura curricular. Em particular, na Universidade Federal de Viçosa (UFV), em que mais da metade dos cursos de graduação cursam a disciplina em questão, este índice tem chamado a atenção dos professores da disciplina que estão se mobilizando para identificar os estudantes que tendem a reprovar na disciplina, visando uma intervenção pedagógica que seja capaz de contornar esta situação.

Na UFV, a disciplina de Cálculo se apresenta em duas vertentes: Cálculo I (MAT 140 e MAT 146) e Cálculo Diferencial e Integral I (MAT 141). O que difere as turmas de Cálculo I são os cursos para os

quais elas são ofertadas. Enquanto que a MAT 140 é ofertada para as engenharias agrícola e ambiental, ambiental, de agrimensura e cartográfica, de alimentos, civil, mecânica, de produção e química, além dos cursos de agronegócio, bioquímica, cooperativismo, ciência da computação, ciências econômicas e química, a MAT 146 é ofertada para os cursos de administração, agronomia, arquitetura e urbanismo, ciência e tecnologia de laticínios, ciências contábeis, engenharia florestal, e zootecnia. Já a MAT 141 é ofertada exclusivamente para estudantes dos cursos de engenharia elétrica, física e matemática.

Um levantamento inicial dos dados referente ao índice de reprovação dessas turmas entre os anos de 2011 e 2016, nos permitiram identificar que em Cálculo I essa problemática é mais intensa. As Tabelas 1 e 2 apresentam como estes índices se comportaram para as disciplinas MAT 140 e MAT 146, respectivamente. Cabe ressaltar que no ano de 2012 devido a uma greve dos servidores da universidade, ao retornar as atividades acadêmicas o estudante pôde efetuar o trancamento da disciplina, o que pode responder pelo diferente índice de reprovação para este ano.

Tabela 1. Índice de reprovação em Cálculo I na UFV – MAT 140

Ano	Número de estudantes matriculados	Número de estudantes Aprovados	Índice de reprovação
2011	535	233	0,56
2012*	511	291	0,43
2013	594	239	0,60
2014	539	195	0,66
2015	428	169	0,61
2016	516	197	0,62

*Nesse ano, servidores da UFV iniciaram uma greve reivindicando, principalmente, a reestruturação da carreira e o reajuste salarial.

Fonte: Os autores (2019)

Tabela 2. Índice de reprovação em Cálculo I na UFV – MAT 146

Ano	Número de estudantes matriculados	Número de estudantes Aprovados	Índice de reprovação
2011	661	328	0,50
2012*	550	329	0,40
2013	658	313	0,52
2014	607	232	0,62
2015	368	100	0,73
2016	476	177	0,63

*Nesse ano, servidores da UFV iniciaram uma greve reivindicando, principalmente, a reestruturação da carreira e o reajuste salarial.

Fonte: Os autores (2019)

De fato, há diversos estudos acadêmicos que propõem práticas pedagógicas capazes de auxiliar na busca pela solução do problema exposto. Lopes (1999) e Wisland, Freitas e Ishida (2014), por exemplo, apresentam estudos acerca dos crescentes índices de reprovação de Cálculo em diferentes cenários e propõem soluções, como a criação de um curso de nivelamento, cujo objetivo seria o de revisar os principais conteúdos de matemática básica que se fazem necessários nos cursos de Cálculo, além de uma revisão acerca do número de estudantes matriculados por turma e um aumento no número de vagas em turmas de tutorias o que seriam possíveis soluções paliativas.

Medidas como estas já foram tomadas na UFV cujos docentes também criaram turmas especiais voltadas para estudantes propensos à reprovação, que receberam o nome de “estudo dirigido”. As turmas de estudo dirigido contam com uma maior carga horária semanal e com um número reduzido de estudantes. Nelas, o professor tem um tempo maior para dedicar-se à resolução de exercícios e o aluno é, a todo momento, incentivado a estabelecer conexões entre os conceitos trabalhados e a desenvolver métodos para solucionar os problemas propostos.

Para selecionar os estudantes para serem matriculados nas turmas de estudo dirigido de Cálculo, tem-se aplicado desde o primeiro semestre do ano de 2016, no início de cada período letivo, uma avaliação de matemática básica composta por 50 itens de múltipla escolha, cada um com cinco alternativas de resposta, elaborados tomando como referência alguns conteúdos de matemática básica considerados necessários para

um bom desenvolvimento da disciplina. Estas avaliações têm a finalidade de estimar o nível de conhecimento em matemática do estudante antes de cursar a disciplina e utilizar algumas das informações obtidas nesse processo para identificar os estudantes que necessitam de um maior acompanhamento no decorrer da mesma, além de servir como um instrumento de apoio aos professores no desenvolvimento dos seus planos de ensino.

A formulação dos itens destas avaliações tem sido de responsabilidade dos professores do Departamento de Matemática da UFV que pretendem desenvolver, futuramente, um banco de itens para que a formulação das avaliações ocorra de maneira sistemática nos próximos anos, mas para isto é necessário que se faça uma análise aprofundada de cada um dos itens já aplicados, visando aproveitar o máximo de informação possível que eles possam oferecer.

Seguindo esta vertente, o presente artigo tem por objetivo avaliar as propriedades psicométricas de uma destas avaliações, baseando-se na Teoria Clássica dos Testes (TCT) e na Teoria de Resposta ao Item (TRI) e, deste modo, investigar a qualidade métrica dos itens que a compõem. Estas teorias são provenientes da Psicometria, área que relaciona a Psicologia e a Estatística, e cuja proposta está na elaboração de instrumentos de medida para o conhecimento e o comportamento humano.

Em síntese, enquanto a TCT investiga as propriedades do conjunto de itens que constituem um teste, a TRI se preocupa em investigar individualmente as propriedades de cada um deles. Cada uma dessas teorias apresenta vantagens e desvantagens em relação a outra, e juntas oferecem um leque de possibilidades de análises que podem ser realizadas em testes avaliativos.

Dessarte, assumimos como objetivos específicos a análise do nível de dificuldade e do poder discriminativo de cada item, além da precisão do instrumento de medida e da informação que eles fornecem para o cálculo das habilidades dos estudantes em Matemática, visando também contribuir para a difusão tanto da TCT, quanto da TRI, na análise da estrutura das avaliações.

Material e Métodos

Os dados sob estudo são provenientes de uma avaliação com conteúdos de matemática básica considerados essenciais para o bom desenvolvimento da disciplina de Cálculo e cujos tópicos encontram-se na Tabela 3. Esta avaliação, tomada para um estudo de caso, continha 50 itens de múltipla escolha e fora aplicada no segundo semestre de 2016 a 477 estudantes, matriculados na disciplina.

Tabela 3. Tópicos que compõem o teste analisado

Itens	Tópico
1, 6 e 7	Teoria dos conjuntos e conjuntos numéricos
2, 3, 4, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 e 18	Expressões numéricas e algébricas
20, 21 e 27	Equações e inequações
22, 23, 24, 25, 26, 35, 36, 49 e 50	Funções: introdução, domínio e imagem de funções, funções do 1º e 2º graus
19 e 37	Função modular
41, 42, 43 e 44	Função exponencial e logarítmica
28, 38, 39 e 40	Função composta e inversa
29	Transformação de funções
30, 31, 32, 33 e 34	Polinômios e equações polinomiais
45, 46, 47 e 48	Trigonometria

Fonte: Os autores (2019)

Cada item possuía 5 alternativas de resposta, das quais apenas uma era a correta. Todos os itens apresentavam a opção *não sei* entre elas. O objetivo de incluí-la entre as alternativas estava na busca do controle ao acerto casual, que poderia prejudicar análises estatísticas, e também para servir como um auxílio aos professores, visto que tópicos que apresentavam um grande índice de resposta para esta alternativa, indicariam que uma revisão sobre o assunto se faria necessária.

Os métodos foram primeiramente desenvolvidos seguindo a linha bibliográfica, por meio de um referencial teórico que se concentrou nos

principais trabalhos relacionados a TCT e a TRI, tanto a nível nacional, com os trabalhos de Pasquali (2003) e Andrade, Tavares e Valle (2000), quanto a nível internacional, com os trabalhos de Baker (2001) e Embretson e Reise (2013). Ademais, considerou-se trabalhos em que os autores se propuseram a realizar uma análise psicométrica de avaliações externas, dos quais destaca-se Primi, Hutz e Silva (2011), Lopes e Vendramini (2015) e Oliveira (2017).

Na parte prática, como um estudo de caso, pôde-se aplicar os conceitos estudados e a aprendizagem desenvolvida a fim de alcançar os objetivos propostos, cujo centro está na análise de uma avaliação de matemática básica utilizando ambas as teorias. Para isto, selecionou-se a avaliação aplicada no segundo semestre do ano de 2016, que contou com 477 respondentes, número que representa quase o dobro do total de respondentes da edição anterior da avaliação, justificando a seleção desta.

Em seguida, coletou-se os dados (gabaritos) que foram transcritos para uma planilha, visando organizá-los para a leitura e análise por meio do *software* R (R, 2019), cuja adoção se deu por ser livre, de código aberto e de fácil manipulação. Os principais pacotes desse *software* utilizados foram o *ltm* (RIZOPOULOS, 2006) e *mirt* (CHALMERS, 2012).

Para as análises propostas, as respostas dos 477 estudantes foram, inicialmente, dicotomizadas, ou seja, transformadas em 0 para respostas incorretas e 1 para respostas corretas. Em seguida, algumas estatísticas descritivas foram obtidas, como o cálculo de medidas de posição e dispersão e análise das frequências de resposta.

Dentro da abordagem clássica, a primeira análise feita diz respeito à consistência interna da avaliação a fim de verificar a qualidade do instrumento de medida na mensuração das habilidades dos respondentes. Para isto, utilizou-se o coeficiente alfa de Cronbach, cujo valor mínimo sugerido por Hair Júnior *et al.* (2010) para considerar o teste como satisfatório é 0,7.

Em seguida, foi analisado para cada item avaliado o percentual de acertos e erros e, conseqüentemente, o índice de dificuldade dos itens, viabilizando a classificação entre fácil (média de acerto superior a 70%), moderado (média de acerto maior que 30% e menor ou igual a 70%) e difícil (média de acerto menor ou igual a 30%), como proposto por Condé (2001).

Para a análise da discriminação, utilizou-se o coeficiente de correlação ponto-bisserial (ρ_{pb}), que representa a correlação entre uma variável categórica dicotômica (acerto e erro) e uma variável contínua (escore do teste). Este índice, varia entre -1 e 1 sendo que, quanto mais próximo de 1 mais discriminativo é o item. No caso em que o valor ρ_{pb} retornado é negativo, considera-se que o item necessita de uma revisão, visto que isto sugeriria que estudantes com menor conhecimento tendem a acertá-lo com maior probabilidade do que estudantes com maior conhecimento (SOARES, 2018).

Sob o enfoque da TRI, primeiramente foi realizada uma análise de componentes principais em busca de evidências de unidimensionalidade, pressuposto para análises via TRI e procedeu-se às estimativas dos coeficientes dos modelos logísticos de um, dois e três parâmetros, que consideram a discriminação do item, a discriminação e a dificuldade do item e a discriminação, a dificuldade e a probabilidade de acerto ao acaso do item, respectivamente, pelo método da média *a posteriori*.

Embora já exista um modelo logístico de quatro parâmetros, que permite controlar fatores aleatórios que fazem com que sujeitos com alta habilidades não respondam corretamente ao item, podendo ocasionar em análises pouco confiáveis do instrumento de medida, “aparentemente não existe nenhuma vantagem dele em relação ao modelo de três parâmetros” (COUTO & PRIMÍ, 2011, p. 6) e, por conta disto, este modelo não foi considerado nas análises aqui realizadas.

Segundo Andrade, Tavares e Valle (2000), a expressão matemática do modelo logístico de três parâmetros é dada por:

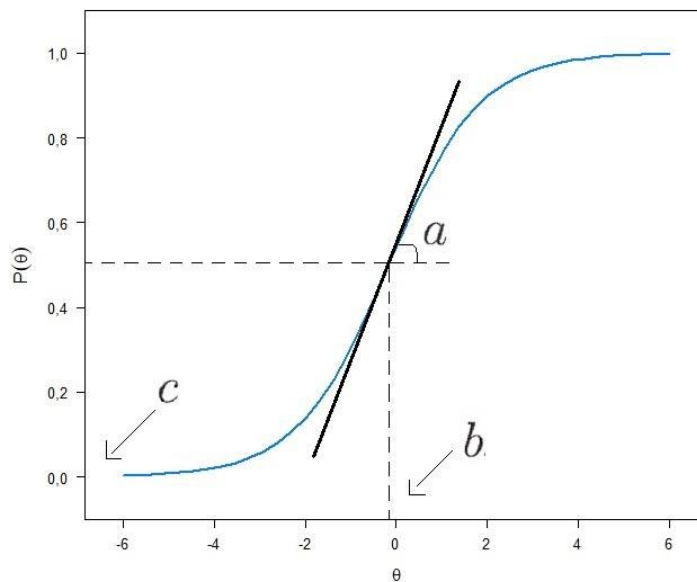
$$P(\theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}} \quad (1)$$

em que $P(\theta_j, a_i, b_i, c_i)$ é a probabilidade do indivíduo j com habilidade θ_j acertar o item i , a_i é o parâmetro de discriminação do item i , com valor proporcional à inclinação da curva característica do item no ponto de abscissa b_i , que representa o parâmetro de dificuldade do item i , medido na mesma escala de habilidade e c_i é o parâmetro do modelo que

representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i , também conhecido como probabilidade do chute certo ou probabilidade de acerto casual.

A fórmula, desenvolvida por Birnbaum (1968), para o modelo logístico de dois parâmetros pode ser encontrada substituindo $c_i = 0$ na Equação 1. Analogamente, para encontrarmos a fórmula do modelo logístico de um parâmetro, criado por Rasch (1960), basta substituírmos $c_i = 0$ e $a_i = 1$ na Equação 1. A Figura 1 representa uma curva característica do item, em que os parâmetros do item encontram-se destacados.

Figura 1. Parâmetros da curva característica do item



Fonte: Os autores (2019)

Através dos critérios de informação de Akaike (AIC) e bayesiano (BIC) em conjunto com o teste de razão de verossimilhanças (TRV), selecionou-se o modelo que melhor se ajustava aos dados. Feito isto, executou-se um teste Qui-quadrado de Bock em busca de verificar a qualidade do ajuste do modelo escolhido.

Para a análise subjetiva do índice de dificuldade de um item sob a ótica da TRI, pode-se construir uma tabela de classificação que leva em consideração os quantis da curva logística sob estudo. Por exemplo, caso queira-se classificar um item em cinco grupos (muito fácil, fácil, moderado, difícil e muito difícil), pode-se utilizar o 1º, 2º, 3º e 4º quintis correspondem ao 20º, 40º, 60º e 80º percentis, respectivamente.

Para a análise do índice de discriminação, tomou-se como base a Tabela 4, proposta por Baker (2001).

Tabela 4. Classificação para o índice de discriminação

Classificação	Parâmetro de discriminação
Nenhuma	0
Muito Baixa	0,01 – 0,34
Baixa	0,35 – 0,64
Moderada	0,65 – 1,34
Alta	1,35 – 1,69
Muito alta	(1,70; $+\infty$)
Perfeita	$+\infty$

Fonte: Adaptado de Baker (2001).

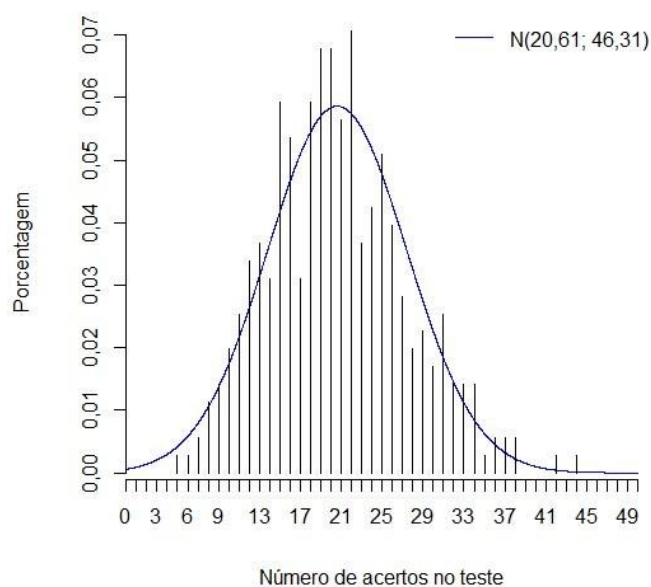
Finalmente, visando uma análise geométrica, foram plotadas as curvas características dos itens, cujas coordenadas foram obtidas pela estimação dos coeficientes já realizada e as curvas de informação, útil por permitir analisar quanto um item traz de informação para a medida da habilidade do indivíduo, sendo um poderoso método para descrever itens, bem como para selecioná-los. O processo de estimação das proficiências dos estudantes pela TRI se deu através do método da média *a posteriori*, proposto por Azevedo (2003), cujas referências se encontram nos trabalhos de Baker (2001) e Andrade, Tavares e Valle (2000).

Resultados e Discussão

Segundo a ótica da TCT, a avaliação apresentou notas que variaram entre 5 e 44 pontos. A média obtida foi de 20,61 pontos, o desvio-padrão de 6,81 pontos e o coeficiente de variação foi 33,04%. Além disso, o primeiro e o terceiro quartil foram de 16 e 25 pontos, respectivamente.

A Figura 2 representa um gráfico de barras da frequência das notas da avaliação. É interessante observar que os extremos inferior e superior possuem uma probabilidade ínfima de ocorrência e que a maioria dos estudantes obteve notas que variavam entre 13 e 27 pontos.

Figura 2. Distribuição do número de acertos dos estudantes



Fonte: Os autores (2019)

O coeficiente Alpha de Cronbach foi superior a 0,7 para todos os itens da avaliação, sendo a média igual a 0,8036, com desvio-padrão igual a 0,004. Dessa forma, conclui-se que nenhum item apresentou problemas de consistência interna.

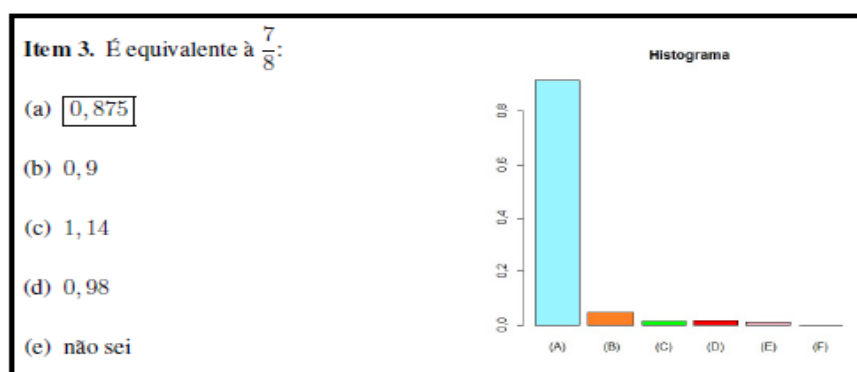
Quanto aos coeficientes de correlação ponto-bisserial destaca-se que os dos itens 4, 18 e 46 foram negativos (-0,2032, -0,0951 e -0,0671), indicando que esses itens são mais acertados por indivíduos de menor habilidade do que os de maior habilidade. Nestes casos, sugere-se que o item seja revisado ou descartado.

Descartando estes itens tem-se que, dos 47 itens restantes, 7 são classificados como fáceis, 24 como moderados e 16 como difíceis. O

item mais fácil foi o tem 3, que obteve cerca de 91,61% de acerto, enquanto que os mais difíceis foram os itens 16 e 17, que obtiveram apenas cerca de 9,64% de acerto, cada.

A Figura 3 traz o item mais fácil pela ótica da TCT e o respectivo gráfico de barras, cujo eixo das abscissas representa a alternativa marcada pelo estudante e eixo das ordenadas representa a proporção com que esta alternativa foi marcada. A alternativa F, representa os estudantes que não marcaram nenhuma das opções. Observe que se trata de um item que exigia um conhecimento simples de expressões numéricas, especificamente era exigido do aluno uma escrita na forma decimal de uma fração irredutível. Dos 477 respondentes, 437 acertaram este item.

Figura 3. Item mais fácil da avaliação pela TCT e gráfico de barras das alternativas marcadas



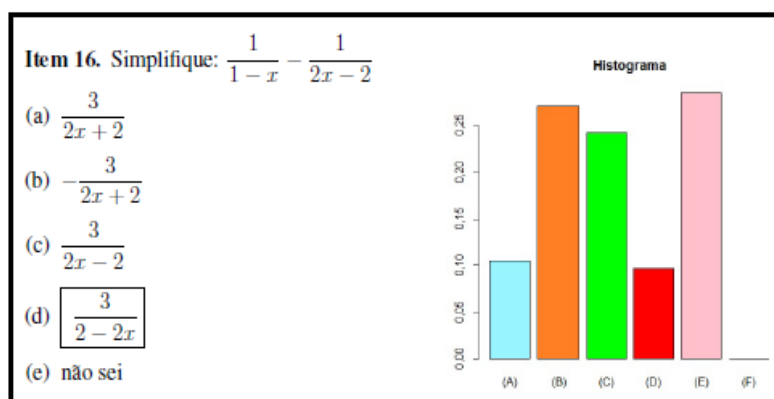
Fonte: Os autores (2019)

Dentre os estudantes que responderam incorretamente ao item 3, mais da metade optou pela alternativa B, cujo valor (0,9) era o que mais se aproximava da resposta correta e que pode ter sido tomado como uma aproximação para a mesma.

As Figuras 4 e 5 trazem os itens mais difíceis da avaliação pela ótica da TCT e os respectivos gráficos de barras. Apenas 19 dos 477 respondentes acertaram estes itens. Em ambos, exigia-se do aluno um conhecimento prévio em fatoração e simplificação de frações algébricas, indi-

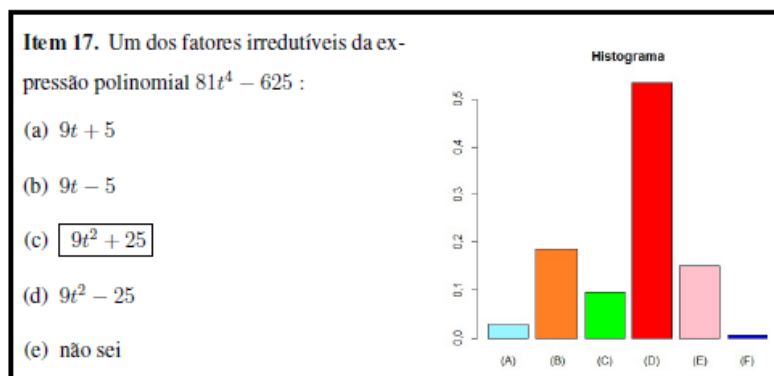
cando que estes tópicos precisam ser retomados e revisados pelos professores da disciplina.

Figura 4. Item mais difícil da avaliação pela TCT e gráfico de barras das alternativas marcadas



Fonte: Os autores (2019)

Figura 5. Item mais difícil da avaliação pela TCT e gráfico de barras das alternativas marcadas



Fonte: Os autores (2019)

O item 16 exigia que o respondente efetuasse o Cálculo do mínimo múltiplo comum (mmc) dos fatores algébricos $1 - x$ e $2x - 2$. Para isto, o estudante poderia perceber que o segundo é um múltiplo do primeiro:

$2x - 2 = -2(1 - x)$, tornando a resolução mais simples de ser executada. Este tipo de raciocínio é comum em exercícios que envolvem o Cálculo de limites de polinômios e deveria ser trabalhado desde os primórdios do ensino da álgebra, ainda no ensino fundamental.

No item 17, mais da metade dos respondentes optou pela alternativa D, cujo valor $9t^2 - 25$ representa, de fato, um dos fatores da expressão polinomial $81t^4 - 625$. Entretanto, tal fator pode ser simplificado como $(9t^2 - 25)(9t^2 + 25)$, por exemplo, não podendo ser classificado como irredutível, como descrito no enunciado do item. Assim, acredita-se que este ponto tenha sido alvo de confusão por parte dos respondentes, acarretando no elevado índice de erro do item.

Para proceder a análise via TRI, primeiramente verificou-se, através da análise de componentes principais, um fator predominante, indicando evidências de unidimensionalidade. É interessante ressaltar que para a adequação dos dados para a análise, obteve-se estatística Kaiser-Meyer-Olkin (KMO) igual a 0,76 e teste de esfericidade de Bartlett significativo, apontando para o uso adequado da análise fatorial.

A Tabela 5 representa as estatísticas AIC, BIC e TRV e o valor- p retornado pelo TRV, obtidos na comparação entre os modelos logísticos de um (modelo de Rasch) e dois parâmetros. Observe que os modelos apresentaram uma diferença significativa à 5% de significância, o que indica que um modelo se sobressai em relação ao ajuste. Uma análise da Tabela 5 permite concluir que, tanto pelos critérios de informação, quanto pelo teste de razão de verossimilhanças, há indícios suficientes para se optar pelo modelo logístico de dois parâmetros.

Tabela 5. Comparação entre os modelos logísticos de 1 e 2 parâmetros

Modelo	AIC	BIC	TRV	Valor-p
Rasch	19386,50	19583,84	-9642,25	
Logístico de dois parâmetros	19090,40	19477,33	-9445,20	0

Fonte: Autores (2019)

A Tabela 6 representa a mesma análise de comparação, mas entre os modelos de dois e três parâmetros logísticos. Observe que os modelos não apresentaram uma diferença significativa à 5% de significância, o que se acredita ter sido consequência do acréscimo da opção “*não sei*”

entre as alternativas de resposta, inibindo o chute certo na avaliação. Assim, optou-se pelo modelo logístico de dois parâmetros devido a menor complexidade do mesmo.

Tabela 6. Comparação entre os modelos logísticos de 2 e 3 parâmetros

Modelo	AIC	BIC	TRV	Valor-p
Logístico de dois parâmetros	19090,40	19477,33	-9445,20	
Logístico de três parâmetros	19129,85	19710,24	-9414,92	0,1457

Fonte: Autores (2019)

Escolhido o modelo, visando analisar a qualidade do ajustamento, realizou-se o teste Qui-quadrado de Bock, cuja hipótese de nulidade postula que o modelo se ajusta bem aos itens. Com exceção do item 4, que já apresentara problemas quanto a discriminação, a avaliação não apresentou problemas quanto ao ajuste e pode-se proceder as estimativas dos coeficientes do modelo.

Ao se estimar as habilidades dos estudantes em matemática pelo método da TRI, observa-se que os valores das estatísticas clássicas obtidas são bem próximos dos valores obtidos via TCT. A menor e a maior nota obtida na avaliação foram 4,555 e 43,039 pontos, respectivamente, a média obtida foi de 20,614 pontos, o desvio-padrão de 6,301 e o coeficiente de variação foi 30,57%. Além disso, o primeiro e o terceiro quartil foram 16,125 e 24,450 pontos, respectivamente.

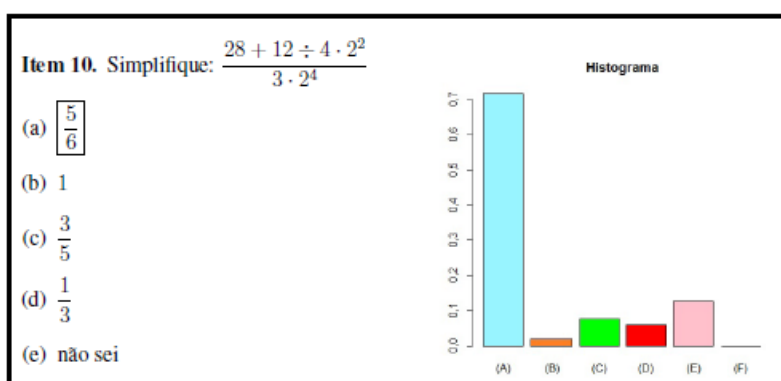
Os itens 4, 18 e 46 também apresentaram discriminação negativa (-0,564, -0,093 e -0,460), pela ótica da TRI, indicando algum erro na elaboração do item ou no gabarito, a mesma conclusão que havíamos obtido ao analisar o coeficiente de correlação ponto bisserial, na TCT.

Descartando estes itens temos que, dos 47 itens restantes, 16 apresentaram discriminação muito baixa ou baixa, 27 apresentaram discriminação média e 4 apresentaram discriminação alta. Em média, obteve-se discriminação igual a 0,7131 com desvio-padrão de 0,4679. O item mais fácil, pela ótica da TRI foi o item 10, também classificado como fácil pela TCT, e o mais difícil, em conformidade com a TCT, foi o item 17. Estes itens apresentaram índices de dificuldade iguais a -3,912 e 3,929, respectivamente. O item 3, que pela TCT foi o mais fácil, obteve

-2,050 para o parâmetro de dificuldade, considerado baixo também pela ótica da TRI, o que confirma as conclusões obtidas.

A Figura 6 traz o item mais fácil pela ótica da TRI e o respectivo gráfico de barras. Observe que este item também exigia do estudante conhecimentos prévios no desenvolvimento de equações numéricas, como o item 3, mais fácil sob o enfoque da TCT. O percentual de acerto para este item foi de 71,70%, sendo classificado também como fácil, tomando como base a classificação proposta por Condé (2001), na TCT. Dos estudantes que responderam incorretamente o item, aproximadamente metade marcaram a opção “não sei”.

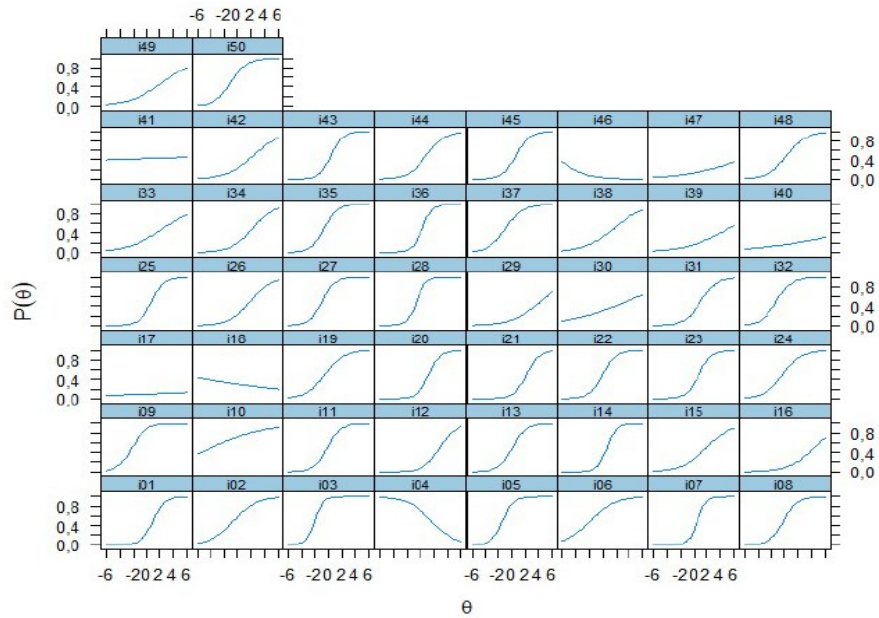
Figura 6. Item mais fácil da avaliação pela TRI e gráfico de barras das alternativas marcadas



Fonte: Os autores (2019)

A Figura 7 mostra as curvas características dos 50 itens da avaliação. Perceba que, os itens 4, 18 e 46 apresentam uma relação inversa entre a habilidade θ do respondente e sua probabilidade de acerto ao item $P(\theta)$, o que já esperava-se pela interpretação feita através do coeficiente de correlação ponto-bisserial, na TCT, e também pela interpretação do índice de discriminação, na TRI. Algumas curvas, como as dos itens 7, 14 e 36 estão bem definidas, indicando qualidade na estrutura dos mesmos.

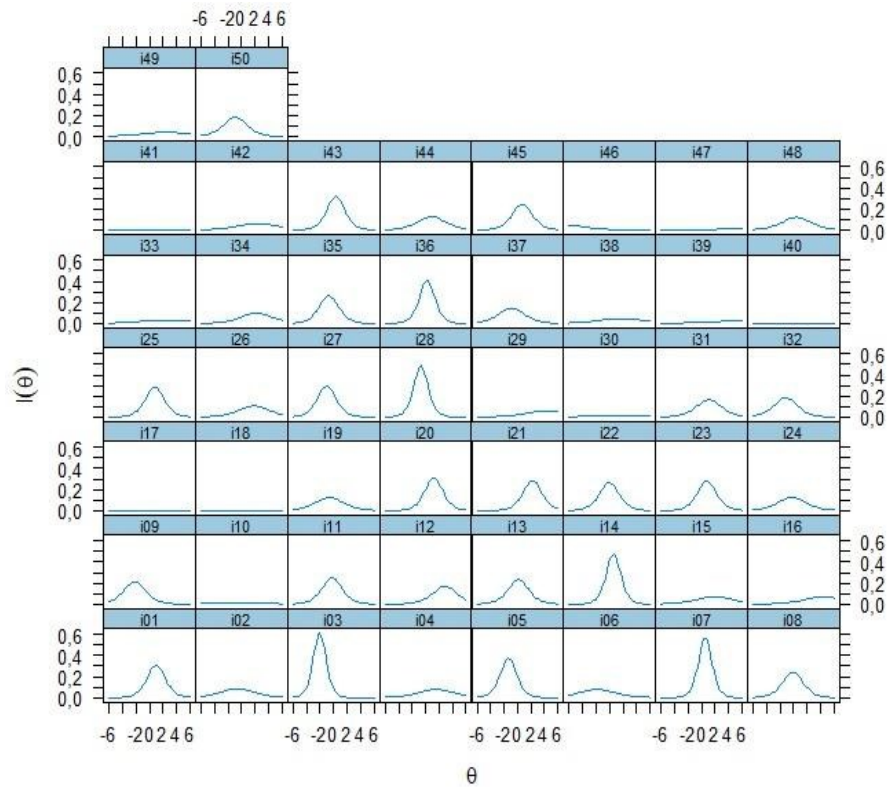
Figura 7. Curvas características dos itens



Fonte: Os autores (2019)

A Figura 8 traz as curvas de informação dos itens. Nelas, pode-se perceber que alguns itens trazem mais informações para o cálculo das habilidades do que outros. Por exemplo, os itens 3 e 7 são bastante informativos. O primeiro, para sujeitos com habilidades negativas. O segundo, para sujeitos com habilidades próximas da média. Entretanto, itens como o 10, 17, 18, 30, 41 e 47 são pouco informativos. Assim, se o objetivo da avaliação é o de avaliar habilidades e pretende-se reduzi-lo, por algum motivo, estes itens também podem ser descartados.

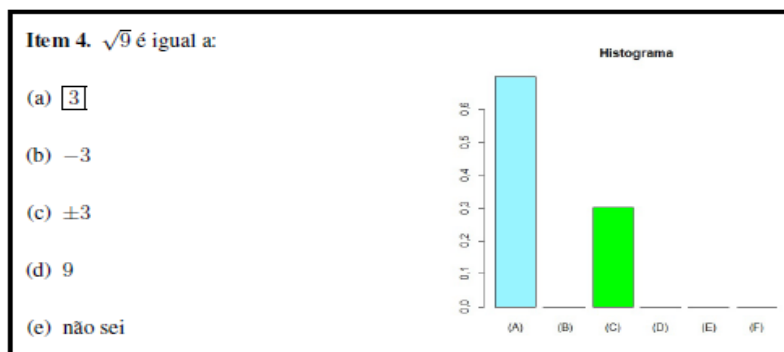
Figura 8. Curvas de informação dos itens



Fonte: Os autores (2019)

As Figuras 9, 10 e 11 trazem os itens passíveis de reformulação ou descarte, pelas análises de discriminação apresentadas.

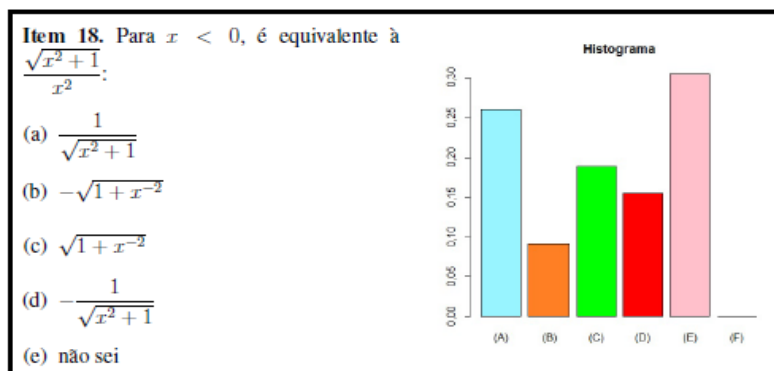
Figura 9. Item com problema de discriminação e gráfico de barras das alternativas marcadas



Fonte: Os autores (2019)

Acredita-se que o problema evidenciado no item 4, de expressões numéricas, tenha relação com uma confusão com o sinal de suas alternativas que pode ocorrer entre os estudantes com um maior grau de conhecimento e que entendem que existe uma diferença entre as soluções da equação $x^2 = 9$ e da radiciação $\sqrt{9}$, em que a equação apresenta como raízes os dois valores expressos pela alternativa C. Observe que as demais alternativas não foram marcadas.

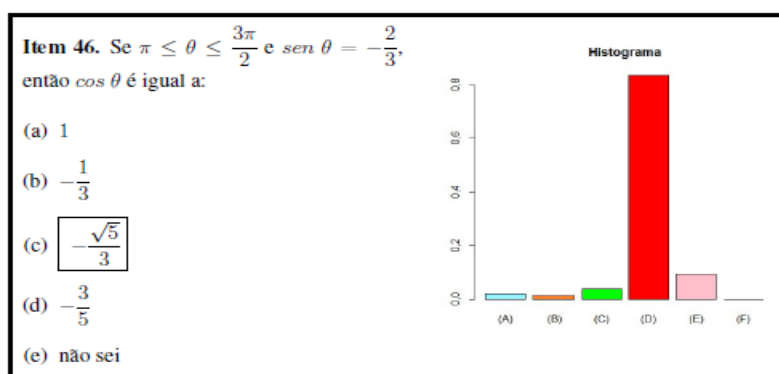
Figura 10. Item com problema de discriminação e gráfico de barras das alternativas marcadas



Fonte: Os autores (2019)

O item 18 exigia conhecimentos de fatoração e simplificação de frações algébricas, mas não apresenta alternativa correta. Neste caso, tanto o coeficiente de correlação ponto-bisserial, da análise via TCT, quanto o índice de discriminação, da análise via TRI, além da curva característica do item, apontaram para este problema, mostrando que estas teorias psicométricas também podem auxiliar na busca por itens com problemas estruturais de formulação.

Figura 11. Item com problema de discriminação e gráfico de barras das alternativas marcadas

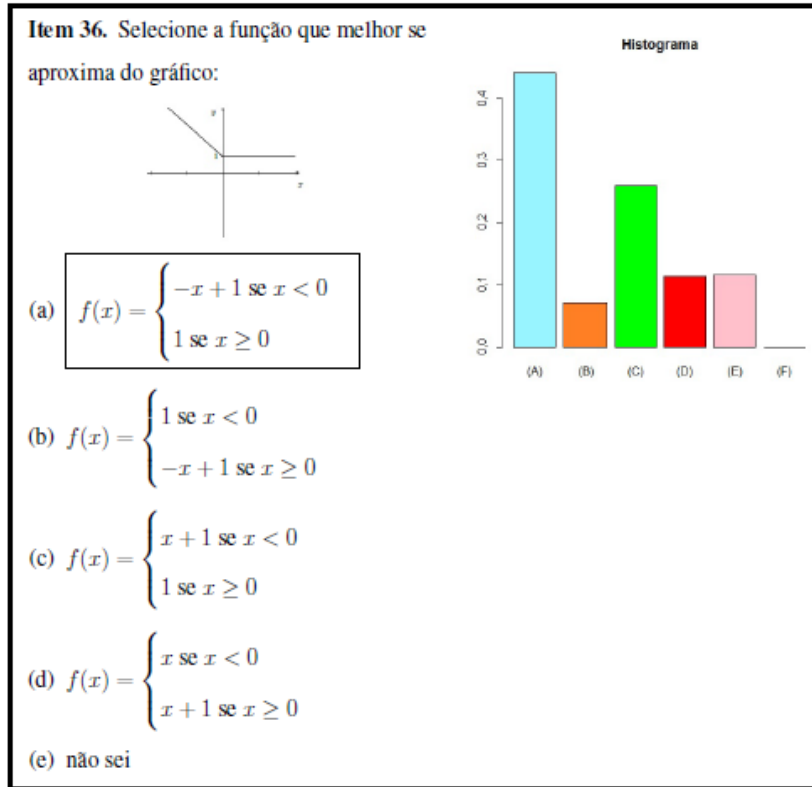


Fonte: Os autores (2019)

O item 46, relacionado ao ciclo trigonométrico, teve um baixo índice de acerto, como pode ser visualizado no gráfico de barras. Apenas 18 dos 477 estudantes apresentaram resposta correta a este item e acredita-se que este fato tenha sido o responsável pelos índices de discriminação terem sido negativos, pois pouca inferência pode ser realizada entre os respondentes que o acertaram.

Analisemos agora o item 36, que apresenta uma curva característica bem definida, e traz informações precisas da habilidade da maioria dos respondentes, visto a simetria de sua curva de informação. A Figura 12 traz este item e seu respectivo gráfico de barras.

Figura 12. Item com boas características psicométricas



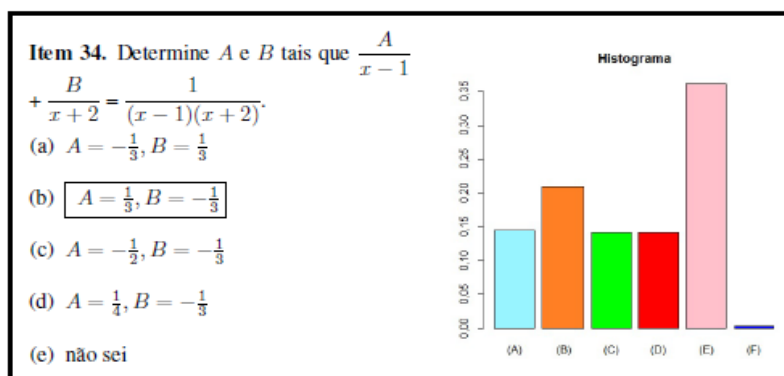
Fonte: Os autores (2019)

O item 36 refere-se às funções definidas por várias sentenças, tópico essencial na aprendizagem dos conceitos de continuidade e diferenciabilidade de funções reais, apresentados na disciplina de Cálculo. O índice de dificuldade para este item foi moderado, tanto pela ótica da TCT quanto pela TRI. A análise da discriminação também classificou o item como moderado, cujos valores para o coeficiente de correlação ponto-bisserial foi de 0,435 e para o índice de discriminação da TRI foi de 1,276. O coeficiente Alpha de Cronbach obtido foi de 0,798.

A Figura 13 traz o item 34 que apresentou os maiores índices para a alternativa “não sei”. Trata-se de item que envolvia conceitos algébricos úteis no processo de decomposição em frações parciais, tópico bastante trabalhado no ensino de integrais, indicando que o professor deve tra-

balhar estes conceitos com bastante cautela para obter ganhos no processo ensino-aprendizagem do conteúdo.

Figura 13. Item com maiores índices para a alternativa não sei



Fonte: Os autores (2019)

Observe que apenas cerca de 20% responderam este item corretamente, sendo que mais do que um terço dos respondentes afirmaram não saber como resolvê-lo. Observe também que há uma similaridade entre os índices de respostas às demais alternativas. A análise clássica retornou valores de 0,209 para o índice de dificuldade e 0,224 para o coeficiente de correlação ponto-bisserial do item. A análise pela TRI indicou 2,259 para o índice de dificuldade e 0,638 para o índice de discriminação. Em ambas, o item foi classificado como difícil e pouco discriminativo. Estes fatos corroboraram para que o item trouxesse pouca informação sobre os respondentes.

Como pôde-se perceber, a TCT e a TRI caminham juntas e muitas das interpretações obtidas através de uma análise sob enfoque de uma destas teorias acabam sendo confirmadas pela outra. Por exemplo, os mesmos itens que apresentaram problemas na discriminação em uma teoria, apresentaram o mesmo problema na outra e a classificação quanto ao nível de dificuldade em ambas também foram harmoniosas. Isto nos leva a perceber que caso queira-se utilizar apenas uma teoria na análise de uma informação, tanto a TCT, quanto a TRI, oferecerá parâmetros confiáveis e significativos. Cabe ao pesquisador definir o tipo de análise específica a ser feita e a complexidade das interpretações.

Considerações Finais

Não é de hoje, e nem é um caso específico da UFV, que a disciplina de Cálculo tem sido um dos grandes desafios dos cursos de graduação que a contêm em sua estrutura curricular. Tal disciplina é responsável por uma altíssima taxa de reprovação e diversas pesquisas em Educação Matemática apontam que pouca relação os estudantes têm conseguido estabelecer entre a sua formação específica e os conteúdos abordados por ela.

Há estudiosos que defendam saídas a partir do uso de estratégias pedagógicas, dentre as quais destacamos o uso de tecnologias e *softwares* matemáticos. Há também aqueles que optam por explorar o potencial da modelagem matemática, da resolução de problemas, dentre outras. A criação de uma disciplina “introdutória”, que abrange os conceitos básicos necessários para um bom rendimento na disciplina, também tem sido uma das alternativas em algumas instituições.

Todas estas práticas têm sido utilizadas no âmbito da UFV na busca por soluções ao problema da reprovação em Cálculo na Universidade e se baseiam em uma avaliação previamente aplicada aos estudantes matriculados na disciplina, cuja análise através da TCT e da TRI, foi o foco deste trabalho.

Estas teorias podem contribuir com o trabalho dos docentes ao elaborarem avaliações para que possam identificar possíveis lacunas na aprendizagem de matemática dos estudantes as quais podem refletir negativamente em seu desempenho na disciplina. A partir daí, pode-se criar um plano de ações, almejando um trabalho minucioso com o objetivo de diminuir ou pelo menos estacionar o crescente índice de reprovação na disciplina, por meio de um maior acompanhamento frente aos estudantes que apresentarem maior dificuldade.

O domínio dos tópicos de matemática básica que compuseram a avaliação são essenciais para um bom desenvolvimento na disciplina de Cálculo Diferencial e Integral, embora muitos estudantes apresentem falhas de aprendizagem oriundas da má formação proveniente da educação básica. Sendo assim, o baixo rendimento pode ser indício de que uma intervenção pedagógica deve ser urgentemente realizada, como o encaminhamento às turmas especiais de acompanhamento individuali-

zado ou outras ações que já são realizadas pelo departamento de matemática da UFV. Dessa forma, a avaliação pode também ser utilizada na seleção de estudantes para participarem destas atividades especiais.

Ações como estas podem contribuir, consideravelmente, para a qualidade das aulas e para uma melhor formação dos envolvidos no processo ensino-aprendizagem, promovendo ganhos de conhecimento e melhorando a produtividade tanto do aluno, quanto do professor.

Outros estudos futuros podem ser realizados a partir deste. Por exemplo, analisar a relação do curso do estudante com os saberes básicos necessários para um bom desenvolvimento da disciplina, avaliar a matriz curricular e debater a criação de uma disciplina de Pré-Cálculo obrigatória, inserir a TRI como meio alternativo para o processo de avaliação dentro da universidade, dentre outros.

Dessa forma, espera-se que esta pesquisa, ainda que de caráter introdutório e representando um estudo específico na UFV, sirva como um instrumento de difusão dos conceitos e teorias que envolvem a Psicometria na análise da estrutura de avaliações e aponte para a necessidade de uma maior discussão pedagógica acerca das possíveis lacunas existentes na aprendizagem de matemática básica dos estudantes do ensino superior.

Recebido em: 22/10/2019

Aprovado em: 10/11/2020

Referências

- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. Teoria da resposta ao item: conceitos e aplicações. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA (SINAPE), 14., 2000, Caxambu. **Anais...** Caxambu: Associação Brasileira de Estatística, 2000. 164 p.
- AZEVEDO, C. L. N. **Métodos de estimação na teoria de resposta ao item**. 2003. 133 p. Dissertação (Mestrado em Estatística) - Universidade de São Paulo, São Paulo. 2003.
- BAKER, F. **The basics of item response theory**. Annapolis: University of Maryland, 2001.
- BIRNBAUM, A. Some latent trait models and their use in inferring an examinee's ability. In: LORD, Frederic e NOVICK, Melvin. **Statistical**

- Theories of Mental Test Scores.** Reading, Boston, MA: Addison-Wesley, 1968. p. 397-479.
- CHALMERS, P. Mirt: A multidimensional item response theory package for the R environment. **Journal of Statistical Software**, Innsbruck, Austria, v. 48, n. 6, p. 1-29, maio, 2012.
- CONDÉ, F. N. **Análise empírica de itens.** Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais-DAEB/INEP/MEC, Brasília, 2001.
- COUTO, G.; PRIMI, R. Teoria de resposta ao item (TRI): conceitos elementares dos modelos para itens dicotômicos. **Boletim de Psicologia**, São Paulo, SP, v. 61, n. 134, p. 1-15, dezembro 2011.
- EMBRETSON, S.; REISE, S. **Item response theory.** Psychology Press, Hove, United Kingdom, 2013. 384 p.
- HAIR JÚNIOR, J.; BLACK, W.; BABIN, B.; ANDERSON, R. **Multivariate data analysis.** 7a ed. Upper Saddle River, New Jersey: Prentice Hall, 2010.
- LOPES, A. Algumas reflexões sobre a questão do alto índice de reprovação nos cursos de Cálculo da UFRGS. **Matemática Universitária.** Rio de Janeiro, RJ, n.2, v. 26/27, p. 123-146, junho, 1999.
- LOPES, F. L.; VENDRAMINI, C. M. M. Propriedades psicométricas das provas de pedagogia do ENADE via TRI. **Avaliação: Revista da Avaliação da Educação Superior.** Sorocaba, v. 20, n. 1, p. 27-47, 2015.
- OLIVEIRA, A. L. S. **Avaliação psicométrica da medida do componente de formação geral da prova do exame nacional de desempenho de estudantes (ENADE) de 2010, 2011 e 2012.** 2017. 107 p. Dissertação (Mestrado em Métodos e Gestão em Avaliação.) - Universidade Federal de Santa Catarina, Florianópolis. 2017.
- PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação.** Petrópolis: Vozes, 2003. 397 p.
- PIACENTINI, C. C. **Reprovação, abandono evasão: um estudo de caso no curso de bacharelado em zootecnia da Universidade Tecnológica Federal do Paraná–Campus Dois Vizinhos.** 2012. 115 p. Dissertação (Mestrado em Ciências) - Universidade Federal Rural do Rio de Janeiro, Seropédica. 2012.
- PRIMI, R.; HUTZ, C.; SILVA, M. C. R. A prova do ENADE de psicologia 2006: concepção, construção e análise psicométrica da prova. **Avaliação Psicológica**, Campinas, v. 10, n. 3, p. 271-294, 2011.

- R DEVELOPMENT CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation on Statistical Computing, 2019. Disponível em: <https://www.r-project.org>. Acesso: 02/03/2019.
- RASCH, G. **Probabilistic models for some intelligence and achievement tests**. Copenhagen: Danish Institute for Education Research, 1960.
- RIZOPOULOS, D. Ltm: An R package for latent variable modeling and item response analysis. **Journal of statistical software**, Innsbruck, Austria, v. 17, n. 5, p. 125, novembro, 2006.
- ROSA, C. M.; RIBEIRO, R. Acesso, retenção e evasão: os contornos da exclusão na Universidade Federal de Goiás. **Cadernos de Pesquisa**, v. 25, n. 3, p. 185-203, 2018.
- SOARES, D. J. M. **Teoria clássica dos testes e teoria de resposta ao item aplicadas em uma avaliação de matemática básica**. 2018. 121 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa, 2018.
- VASCONCELOS, R.; ALMEIDA, L.; MONTEIRO, S. O insucesso e abandono acadêmico na universidade: Uma análise sobre os cursos de engenharia. In: International Conference on Engineering and Computer Education. **Anais...** Buenos Aires: ICECE, cop. 2009. p. 457-461.
- VELOSO, T. C.; ALMEIDA, E. P. Evasão nos cursos de graduação da Universidade Federal de Mato Grosso, Campus Universitário de Cuiabá—um processo de exclusão. **Série-Estudos-Periódico do Programa de Pós-Graduação em Educação da UCDB**, v. 1, n. 13, p. 133-148, 2013.
- WISLAND, B.; FREITAS, M. C. D.; ISHIDA, C. Y. Desempenho acadêmico dos alunos em curso de engenharia e licenciatura na disciplina de Cálculo I. **Iberoamerican Journal of Industrial Engineering**, Florianópolis, SC, v. 6, n. 11, p. 94-112, julho, 2014.