

Estudo sobre Construção de Escalas com Base na Teoria da Resposta ao Item: Avaliação de Proficiência em Conteúdos Matemáticos Básicos

Scale Construction Study based on Item Response Theory: Evaluation of Proficiency in Basic Mathematical Contents

Tânia Robaskiewicz Coneglian **Fujii***

 ORCID iD 0000-0001-6341-1235

Aparecida Donizete Pires de **Souza****

 ORCID iD 0000-0001-9533-5804

Monica **Fürkotter*****

 ORCID iD 0000-0003-3479-5289

Adriano Ferreti **Borgatto******

 ORCID iD 0000-0001-6280-2525

Mariana **Cúri*******

 ORCID iD 0000-0002-7651-1064

Resumo

Neste artigo apresenta-se um estudo sobre construção de escalas, com base na Teoria da Resposta ao Item (TRI), para medir proficiência em conteúdos matemáticos básicos, necessários ao acompanhamento das disciplinas de Cálculo e similares, de ingressantes em cursos da área de Ciências Exatas. Adotou-se o modelo logístico unidimensional de três parâmetros, que estabelece média zero e desvio padrão 1, para as proficiências dos indivíduos. As proficiências estimadas foram transformadas em outra escala, optando-se por valores adotados por sistemas de avaliação brasileiros, a saber, 250 e 50. O instrumento de medida consistiu em uma prova com 36

* Mestre em Ciências (Matemática Aplicada e Computacional) pela Universidade Estadual Paulista (Unesp). Professora colaboradora na Escola Estadual Santos Dumont, Paranacity, Paraná, Brasil. E-mail: taniaconeglian@gmail.com.

** Doutora em Ciências (Engenharia de Produção) pela Universidade Federal do Rio de Janeiro (COPPE/UFRJ). Professora do Programa de Pós-graduação em Matemática Aplicada e Computacional, Mestrado, da Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente, São Paulo, Brasil. E-mail: aparecida.dp.souza@unesp.br.

*** Doutora em Ciências (Matemática) pela Universidade de São Paulo (USP). Professora do Programa de Pós-graduação em Educação da Universidade do Oeste Paulista (Unoeste), Presidente Prudente, São Paulo, Brasil. E-mail: monicaf@unoeste.br.

**** Doutor em Agronomia (Estatística e Experimentação Agronômica) pela Escola Superior de Agricultura Luiz de Queiroz (ESALQ/USP). Professor do Programa de Pós-Graduação em Educação Física da Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina, Brasil. E-mail: adriano.borgatto@ufsc.br.

***** Doutora em Estatística pela Universidade de São Paulo (USP). Professora do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional da Universidade de São Paulo (USP), São Carlos, São Paulo, Brasil. E-mail: mcuri@icmc.usp.br.

itens, de cinco alternativas, somente uma correta, elaborados com base em uma matriz de referência, dividida em três temas, “Espaço e Forma”, “Grandezas e Medidas” e “Números e Operações, Álgebra e Funções”. Cada tema é composto por competências, que descrevem as habilidades que se deseja medir. Para a construção da escala foram especificados níveis de proficiência, representando pontos selecionados pelos pesquisadores para serem interpretados pedagogicamente. Estabelecidos os níveis âncora, foram definidos os itens âncora, a partir de critérios, como, por exemplo, o número de acertos, os percentuais de acertos e a diferença entre seus valores, para níveis consecutivos. Com base nestes critérios, comparou-se três métodos de posicionamento dos itens, mostrando as dificuldades de interpretação em pontos da escala. Tais dificuldades oportunizaram a propositura de outro método, segmentando a escala em faixas de proficiência, com base em agrupamentos hierárquicos dos níveis, o que permitiu a interpretação da escala em toda a sua amplitude.

Palavras-chave: Construção e Interpretação de Escalas. Teoria da Resposta ao Item. Proficiência em Conteúdos Matemáticos Básicos.

Abstract

This article presents a study on scale construction, based on the Item Response Theory (IRT), to measure the proficiency in basic mathematical contents, which are key to the follow-up of Calculus and similar subjects, for those entering courses in the Exact Sciences area. The one-dimensional logistic model with three parameters was adopted, which establishes zero as the mean and a standard deviation of 1, for individuals' proficiencies. The estimated proficiencies were transformed in another scale, opting for values adopted by Brazilian evaluation systems: 250 and 50. The measurement instrument consisted of a test with 36 items, with five alternatives each, only one of them correct, that were elaborated based on a reference matrix, divided into three themes, “Space and Form”, “Quantities and Measures”, and “Numbers and Operations, Algebra and Functions”. Each subject is composed of competencies, which describe the skills to be measured. To build the scale, proficiency levels were specified, representing points selected by the researchers to be pedagogically interpreted. Once the anchor levels are established, anchor items were defined based on some criteria, such as the number of correct answers, the percentage of correct answers and the difference between their values, for consecutive levels. Based on these criteria, three methods of items' positioning were compared, showing the difficulties of interpretation in points of the scale. Such difficulties made it possible to propose another method, segmenting the scale into ranges of proficiency, based on hierarchical groupings of levels, which allowed the scale to be interpreted in all its breadth.

Keywords: Scale Construction and Interpretation. Item Response Theory. Proficiency in Basic Mathematical Content.

1 Introdução

A evasão e a retenção de alunos ingressantes nos cursos de graduação da área de Ciências Exatas em disciplinas envolvendo Cálculo Diferencial e Integral, Geometria Analítica e Álgebra Linear tem sido tema de preocupação em diferentes instâncias das Instituições de Ensino Superior.

Considerando que o conhecimento sobre a proficiência dos ingressantes em conteúdos matemáticos básicos pode contribuir para a adoção de metodologias diferenciadas em determinadas disciplinas, bem como para o planejamento de atividades paralelas, visando suprir suas deficiências, tanto do Ensino Fundamental quanto do Ensino Médio, esta pesquisa teve por objetivo a construção de uma escala para avaliação da proficiência desses alunos em tais conteúdos, utilizando a Teoria da Resposta ao Item (TRI), visando fornecer subsídios para o planejamento das disciplinas. Buscando identificar e compreender as dificuldades apresentadas

pelos ingressantes e iniciar a construção desta escala, partiu-se dos resultados e das escalas construídas para dois importantes sistemas de avaliação em larga escala no Brasil: o Sistema de Avaliação da Educação Básica (SAEB) e o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP). Tais sistemas realizam periodicamente um diagnóstico da Educação Básica brasileira e paulista e de fatores que podem interferir no desempenho do estudante. O SAEB é realizado, bienalmente, pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), enquanto o SARESP é realizado anualmente pela Secretaria da Educação do Estado de São Paulo (Seduc/SP).

Os resultados das proficiências obtidas no SARESP para todas as séries/anos avaliados estão na mesma escala do SAEB, com média 250 e desvio padrão 50. Para a interpretação das escalas, os níveis são escolhidos como no SAEB, com espaçamento de 25 pontos, o equivalente a meia unidade de desvio padrão da escala de proficiência. A Escala de Matemática, por sua vez, é interpretada nos pontos 75 a 475. A descrição de cada um dos pontos é feita com base nos resultados de desempenho dos alunos na prova de Matemática e de acordo com as habilidades detalhadas nas matrizes de referência para Avaliação do SARESP ou do SAEB, quando for o caso. Os pontos da escala são agrupados em níveis de proficiência com pontuações distintas para cada etapa da escolaridade. Os resultados das duas avaliações fornecem o percentual de alunos em cada nível de proficiência para cada uma das etapas da escolaridade. No caso específico do 3º ano do Ensino Médio os níveis de proficiência e as respectivas pontuações são sintetizados na Tabela 1, a seguir. No caso do SAEB os resultados são referentes às escolas estaduais, rurais e urbanas, do Estado de São Paulo.

Tabela 1 – Níveis de proficiência em Matemática e percentuais de alunos em cada nível, do SAEB e do SARESP, para alunos da 3ª série do Ensino Médio

Níveis	Faixa de Proficiência	Níveis de Proficiência		Percentual de alunos nos níveis de proficiência				
		SAEB	SARESP	SAEB		SARESP		
				2015	2017	2015	2016	2017
0	≤ 225	Insuficiente	Abaixo do Básico	64,4	60,2	44,4	47,6	46,8
1] 225 a 250]		Abaixo do Básico					
2] 250 a 275]		Abaixo do Básico					
3] 275 a 300]	Básico	Básico	31,2	35,6	49,7	47,3	48,7
4] 300 a 325]		Básico					
5] 325 a 350]		Básico					
6] 350 a 375]	Adequado	Adequado	4,0	3,9	5,6	4,8	4,3
7] 385 a 400]		Adequado					
8] 400 a 425]		Adequado					
9] 425 a 450]	Adequado	Avançado	0,4	0,3	0,3	0,3	0,4
10	≥ 450		Avançado					

Fonte: Elaborada pelos autores, a partir de dados do SAEB e do SARESP (2020)

Tais informações permitem concluir que, considerando qualquer um dos sistemas de

avaliação, no período de 2015 a 2017, aproximadamente 95% dos alunos apresentaram proficiência em Matemática nos níveis básico e abaixo do básico. Somente cerca de 5% deles apresentaram níveis adequado ou avançado, revelando que grande parte dos alunos das escolas públicas concluem o Ensino Médio com defasagem nos conteúdos matemáticos básicos.

Paralelamente a esses resultados, os relatórios para os vestibulares da Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp) de 2015, 2016 e 2017, emitidos pela Fundação para o Vestibular da Universidade Estadual Paulista (VUNESP), informam a partir de dados do questionário socioeconômico, aplicado no momento da inscrição para o vestibular, que 42,4%, 42,9% e 51,6%, respectivamente, dos alunos ingressantes nos cursos da área de Ciências Exatas cursaram o Ensino Médio todo ou em maior parte em escolas públicas.

Esses dados e os altos índices de retenção e evasão em disciplinas dos cursos de graduação na área de Ciências Exatas, motivaram a construção de uma escala de proficiência em conteúdos matemáticos, necessários ao acompanhamento das disciplinas básicas desses cursos.

Tradicionalmente, quando se deseja medir a proficiência de um indivíduo em uma área do conhecimento, aplica-se uma prova (teste) com determinado número de itens (questões) e, com base no número de acertos, determina-se seu escore. A pontuação obtida segue os princípios da Teoria Clássica de Testes (TCT¹). Neste contexto, a proficiência do indivíduo depende fortemente do teste ao qual foi submetido e é difícil realizar comparações entre indivíduos submetidos a provas diferentes. A Teoria da Resposta ao Item (TRI) permite mensurar características que não podem ser medidas diretamente (traço latente), por meio de um conjunto de respostas a itens de um instrumento de avaliação, superando tais dificuldades. Na TRI, o elemento principal é o item e não o instrumento.

Embora a TRI tenha surgido inicialmente no contexto da Teoria Psicométrica em avaliações psicológicas, atualmente, é amplamente utilizada em avaliações educacionais. Com base na TRI, pode-se estabelecer escalas de proficiência interpretáveis que possibilitam, por exemplo, a comparação entre indivíduos e o acompanhamento da evolução dos sistemas de ensino ao longo dos anos. O SAEB, o SARESP e o Exame Nacional do Ensino Médio (ENEM) são exemplos de sistemas brasileiros que utilizam a TRI.

Mais especificamente, a TRI consiste em uma família de modelos matemáticos que relaciona variáveis observáveis (as respostas dadas aos itens de uma prova) e variáveis que não

¹ A TCT tem como foco o comportamento, não o traço latente. Isto é, o escore total em um teste. As aptidões são medidas, em geral, pela soma das pontuações referentes às respostas dadas a uma série de itens, expressa no chamado escore total (PASQUALI, 2003).

podem ser observadas diretamente (a proficiência de um indivíduo em Matemática básica).

Considerando o interesse em medir proficiência em conteúdos matemáticos básicos, para ingressantes em cursos da área de Ciências Exatas, o foco desta pesquisa está no estudo sobre a construção de uma escala para este fim, com base na TRI. As etapas que envolvem esta tarefa consistem na definição da matriz de referência para o traço latente (proficiência), elaboração ou escolha de itens, construção e aplicação da prova, tratamento dos dados, especificação do modelo, estimação dos parâmetros dos itens e da proficiência dos respondentes e, finalmente, a construção e interpretação da escala.

Cabe mencionar que levantamentos realizados no Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) e na *Scientific Electronic Library Online (SciELO)*, utilizando os descritores “Teoria da Resposta ao Item”, “Cálculo Diferencial e Integral”, “matriz de referência”, “proficiência em conteúdos matemáticos” e “construção de escalas” articulados com os operadores booleanos AND e OR, sem filtro, resultaram em apenas três pesquisas acadêmicas, envolvendo a TRI e proficiência em conteúdos matemáticos, desenvolvidas por Rossi (2015), Gomes (2016) e Soares (2018).

Rossi (2015), com o objetivo de construir uma escala para medir proficiência em conteúdos matemáticos básicos, de ingressantes nos cursos da área de Ciências Exatas da Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia (FCT), campus Presidente Prudente, apresentou uma aplicação da TRI, utilizando o modelo logístico unidimensional de três parâmetros. Elaborou uma matriz de referência, a partir das matrizes do SAEB², tendo em vista a inclusão de competências consideradas indispensáveis para a aprendizagem de conteúdos de Cálculo e disciplinas similares. O instrumento base para avaliação foi composto por 32 itens, associados criteriosamente às competências e habilidades descritas na matriz de referência. Utilizando o critério de Beaton e Allen (1992) para a construção da escala, posicionou como âncora 9 dos 32 itens da prova, o que não permitiu interpretá-la em toda a sua extensão.

As inovações implantadas pelo Departamento de Matemática da Universidade de Brasília (UnB), no que se refere ao modelo pedagógico de aulas e ao processo de avaliação de alunos de Cálculo Diferencial e Integral, baseado na concepção de avaliação de competências, com suporte da Psicometria e da Estatística foram apresentadas por Gomes (2016). Foram criados testes de múltipla escolha, como parte do processo de avaliação da aprendizagem,

² Matrizes de referência do 9º ano do Ensino Fundamental e 3º ano do Ensino Médio do SAEB.

recorrendo-se à construção de uma matriz de referência para avaliação do desenvolvimento de habilidades e competências, bem como à TRI e à TCT para calibração dos parâmetros dos itens e obtenção das proficiências dos estudantes. Apresentou, também, a construção inicial de uma escala de proficiência para feedback aos participantes e um esquema simplificado de uma proposta de Avaliação Computacional Adaptada para a disciplina.

Soares (2018) aplicou a TCT e a TRI (modelo unidimensional de dois parâmetros) para analisar as propriedades psicométricas de um teste de avaliação em Matemática Básica, para ingressantes da Universidade Federal de Viçosa (UFV) em cursos que incluem em sua estrutura curricular a disciplina de Cálculo. Baseado em resultados da TCT e da TRI, concluiu que o teste apresentava itens com poder discriminativo em diferentes níveis de dificuldade, revelando que a avaliação proposta é eficiente.

Outra pesquisa relevante nesta linha é a de Monteiro, Afonso e Pires (2017). As autoras analisaram a construção e a aplicação de um teste, que teve por objetivo identificar os conhecimentos em Matemática de ingressantes no Ensino Superior de cursos da área de Ciências e Tecnologias, em Portugal. As características metrológicas do teste foram analisadas também no âmbito da TCT e da TRI, com base no Modelo de Rasch³. Enfatizaram que o teste é apropriado para organizar recomendações e esboçar soluções que visem promover o sucesso acadêmico, incluindo a melhoria efetiva das práticas de ensino nas unidades curriculares de Matemática do 1º ano de cursos superiores.

Considerando similaridades entre os trabalhos descritos e a pesquisa aqui apresentada ressalta-se que todos têm como base avaliar a proficiência em conteúdos matemáticos básicos necessários ao acompanhamento da disciplina de Cálculo e similares, mas somente Rossi (2015) trata da construção de uma escala para este fim.

O trabalho desenvolvido nesta pesquisa dá continuidade ao de Rossi (2015), comparando três métodos de posicionamento de itens e propondo um outro, baseado na segmentação da escala em faixas de proficiência, com base em agrupamentos hierárquicos dos níveis. Agrupados os níveis, os itens âncora são posicionados nas respectivas faixas, permitindo a interpretação da escala em toda a sua amplitude, revelando o ineditismo desta pesquisa.

2 TRI: conceitos, definições e estimação de parâmetros

A TRI consiste em um conjunto de modelos matemáticos que representam a

³ Modelo logístico unidimensional de 1 parâmetro (ML1), proposto por Rasch na década de 1960. Considera somente o parâmetro de dificuldade do item.

probabilidade de resposta correta a um item em função dos parâmetros do item e da proficiência (ou traço latente) do indivíduo. Essa relação é expressa de tal modo que, quanto maior a proficiência, maior será a probabilidade de resposta correta ao item (ANDRADE; TAVARES; VALLE, 2000). O principal foco da TRI é o item e não o teste como um todo, contemplando a análise de itens, através das estimativas de seus parâmetros, e a análise do instrumento de medida como um todo. Neste contexto, a medida da proficiência não depende dos itens apresentados ao indivíduo, e os parâmetros de discriminação e de dificuldade do item não dependem do grupo de indivíduos, que se refere a uma amostra de uma população (ANDRADE; TAVARES; VALLE, 2000). Assim, na TRI, um item mede determinado conhecimento, independentemente de quem o está respondendo e a proficiência de um indivíduo não depende dos itens apresentados a ele (ANJOS; ANDRADE, 2012).

Os diversos modelos propostos na literatura dependem fundamentalmente de três fatores: da natureza do item, dicotômicos ou não; do número de populações envolvidas, uma ou mais de uma; e, da quantidade de traços latentes, um (unidimensional) ou mais de um (multidimensional). Os mais utilizados para itens dicotômicos são os modelos logísticos que, segundo Andrade, Tavares e Valle (2000), podem ser aplicados tanto para análise de itens de múltipla escolha, quanto para a análise de itens abertos, corrigidos como certo ou errado.

Essencialmente, existem três modelos logísticos, que diferem pelo número de parâmetros. São conhecidos como modelo logístico de um parâmetro (ML1), que considera somente o parâmetro de dificuldade do item, de dois parâmetros (ML2), de dificuldade e de discriminação, e o de três parâmetros (ML3), de dificuldade, de discriminação e da probabilidade de acerto ao acaso, apresentado na Subseção 2.1.

2.1 O modelo logístico unidimensional de três parâmetros (ML3)

O ML3 especifica a probabilidade de um indivíduo j , $j = 1, 2, \dots, n$, com proficiência (traço latente) θ_j , responder corretamente ao item i , $i = 1, 2, \dots, I$. Os três parâmetros associados aos itens são: de dificuldade (ou de posição) (b_i), medido na mesma escala do traço latente (θ_j); de discriminação (ou inclinação) (a_i); e, de probabilidade de acerto ao acaso (c_i), probabilidade de um indivíduo com baixa proficiência responder corretamente a este item.

A Curva Característica do Item (CCI), que representa a relação entre a probabilidade do indivíduo j , com proficiência θ_j , responder corretamente ao item i e a escala de proficiência, bem como os parâmetros do modelo associados ao item, e sua forma em “S”, podem ser

visualizadas na Figura 1(a). Nas Figuras 1(b) e 1(c) são exemplificadas curvas características para diferentes valores associados aos parâmetros de discriminação e de dificuldade, para um específico valor da probabilidade de acerto ao acaso. A CCI apresenta menor inclinação, indicativo de que o item tem baixo poder de discriminação, para baixos valores de a_i . Neste caso, indivíduos com baixa ou alta proficiência terão aproximadamente a mesma probabilidade de acertar o item. Para um alto valor a_i obtém-se curvas características “íngremes”, que acabam dividindo os indivíduos em dois grupos: os que possuem proficiência abaixo e os que possuem proficiência acima da localização do item. O parâmetro b_i , na mesma escala de θ , representa a proficiência necessária para a probabilidade de acerto dada por $(1 + c_i)/2$ (ANDRADE; TAVARES; VALLE, 2000). Deste modo, quanto maior o valor b_i , mais difícil é o item.

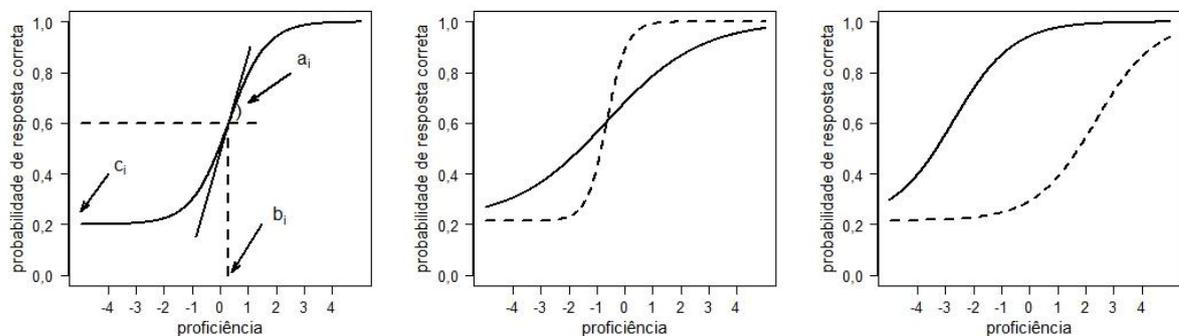


Figura 1 – Exemplos de CCI: (a) em função dos parâmetros; (b) para diferentes valores de a_i (linha contínua $a_i = 0,603$; linha tracejada $a_i = 2,833$), $b_i = -0,638$ e $c_i = 0,215$; (c) para diferentes valores de b_i (linha contínua $b_i = -2,729$; linha tracejada $b_i = 2,351$), $a_i = 0,938$ e $c_i = 0,215$.

Fonte: Elaborada pelos autores ((a) adaptada de Andrade, Tavares e Valle (2000)).

2.2 Estimação dos parâmetros dos itens e das proficiências para o Modelo ML3

Um elemento chave no processo de estimação dos parâmetros é a função de verossimilhança. De modo geral, como o modelo é especificado em função da probabilidade de um indivíduo, com determinada proficiência, responder corretamente a um item, a função de verossimilhança é construída a partir dos produtos das probabilidades associadas a cada item e a cada indivíduo, carregando assim toda informação contida no conjunto de respostas. Na abordagem clássica, em geral, a estimação dos parâmetros é feita pelo Método de Máxima Verossimilhança, que utiliza como estimativas os valores para os parâmetros que maximizam a função de verossimilhança (ANDRADE; TAVARES; VALLE, 2000). Neste trabalho optou-se pela abordagem bayesiana, que consiste em combinar as informações contidas na função de verossimilhança com informações contidas na distribuição *a priori* conjunta especificada para os parâmetros dos itens e das proficiências, culminando na distribuição *a posteriori* conjunta. (MIGON; GAMERMAN; LOUZADA, 2014). Para a estimação dos parâmetros utilizou-se o

amostrador de Gibbs, algoritmo da classe dos Métodos de Monte Carlo via Cadeia de Markov (MCMC), implementado no software Bayesian Inference Using Gibbs Sampler (OpenBUGS) (LUNN; SPIEGELHALTER; BEST, 2009). A ideia básica do MCMC é gerar amostras da distribuição *a posteriori*.

3 Matriz de referência, definição dos itens, construção e aplicação da prova

A matriz de referência adotada nesta pesquisa, elaborada por Rossi (2015), foi construída com base nos descritores das matrizes de referência do 9º ano do Ensino Fundamental e 3º ano do Ensino Médio do SAEB, que apresentam similaridades com as habilidades presentes na matriz de referência do SARESP. Tal matriz não contempla todo o currículo escolar, assim como ocorre com as matrizes de referência dos sistemas de avaliação mencionados. É constituída de 36 descritores, considerados necessários ao acompanhamento das disciplinas básicas de Matemática em cursos na área de Ciências Exatas. Dos 36, quatro não estão nas matrizes de referência do SAEB, a saber, os relacionados a equações e inequações do segundo grau e a equações e inequações modulares. A matriz, apresentada no Anexo 1, Quadro 1, engloba os temas “Espaço e Forma”, “Grandezas e Medidas” e “Números e Operações e Álgebra e Funções”. Em cada um destes temas foram considerados descritores, que especificam o que os itens das provas devem medir, ao associarem conteúdos curriculares com competências e habilidades desejáveis. Ressalta-se que cada um dos descritores, cuidadosamente investigado, tem papel fundamental no desenvolvimento dos conteúdos das disciplinas básicas de interesse.

Dos 32 itens que compuseram a prova aplicada por Rossi (2015) aos ingressantes de 2014, 26 foram utilizados na prova de 2017, por mostrarem-se adequados para avaliar as competências para as quais foram propostos, cinco foram modificados e outros cinco incluídos, com o objetivo de melhor estimar as proficiências nos pontos localizados nas extremidades da escala, sendo um para cada descritor. Os 36 itens foram construídos a partir de exemplos presentes no Plano de Desenvolvimento da Educação (PDE) da Prova Brasil (BRASIL 2008a) e do SAEB (BRASIL, 2008b). A prova foi aplicada no primeiro semestre de 2017, para os alunos matriculados nas disciplinas Cálculo Diferencial e Integral I ou Matemática Elementar I, do primeiro ano, de oito cursos da área de Ciências Exatas da FCT/Unesp.

4 Construção de escalas para avaliação de proficiência com base na TRI

A proficiência de um indivíduo pode assumir teoricamente qualquer valor real entre $-\infty$ e $+\infty$, diferente do que acontece na medida escore, de um teste com I questões corrigidas como certo ou errado, em que a proficiência assume valores de acordo com pontuações propostas inicialmente. De acordo com a TRI, para a definição de uma escala é necessário estabelecer uma origem e uma unidade de medida, escolhidas de modo que representem, respectivamente, o valor médio e o desvio padrão para a proficiência dos indivíduos da população em estudo. Como já mencionado, na especificação do modelo se utiliza a escala com média zero e desvio padrão um. Nesta escala, os valores do parâmetro b_i variam (tipicamente) entre -3 e $+3$ e espera-se que o parâmetro a_i assuma valores maiores do que 0, sendo $a_i \geq 1$ os valores mais apropriados, segundo Andrade, Tavares e Valle (2000).

Após a estimação dos parâmetros dos itens (também conhecida como calibração dos itens) e das proficiências, estimadas em uma escala com média zero e desvio padrão um, escala $(0,1)$, é possível efetuar uma transformação linear, para qualquer outra escala que o pesquisador desejar, trabalhando com números inteiros positivos, tendo como propósito facilitar a construção e a utilização da escala. Por exemplo, um indivíduo com proficiência 1,5 na escala $(0,1)$ está 1,5 desvio padrão acima da média. Na escala utilizada pelo ENEM $(500, 100)$ esse indivíduo teria proficiência 650 e pelo SAEB $(250,50)$ seria 325. É importante ressaltar que, independente da escala, a probabilidade de um indivíduo responder corretamente a um certo item é sempre a mesma, ou seja, a proficiência de um indivíduo é invariante à escala de medidas. Nesta pesquisa optou-se por utilizar a escala do SAEB.

As escalas são construídas a partir da especificação de níveis de proficiência, conhecidos como níveis âncora, que representam pontos na escala selecionados pelo pesquisador (analista) para serem interpretados pedagogicamente. Uma vez estabelecidos os níveis âncora, a partir de algum critério considerado, define-se os itens âncora. É importante ressaltar que cada item pode estar associado a uma ou mais competências da matriz de referência e que sua classificação como âncora, para um particular nível, é que permite a interpretação pedagógica. Dependendo do critério adotado para o posicionamento dos itens, é essencial que os níveis âncora sejam especificados de maneira que não fiquem muito próximos uns dos outros e que a prova contenha um número suficiente de itens, a fim de possibilitar a construção e a interpretação da escala de proficiência, dado que não se pode ter certeza sobre quantos itens serão classificados, nem ao menos se serão classificados itens âncora (no teste aplicado) para todos os níveis fixados.

Os critérios para posicionamento de itens discutidos neste trabalho envolvem o proposto

por Beaton e Allen (1992), os adotados pelos sistemas de avaliação SARESP e SAEB e o proposto em Barros (2016), que consistiu na segmentação de uma escala para a interpretação de um indicador de nível socioeconômico, formando grupos de níveis, através do método de agrupamento hierárquico de Ward⁴, descritos segundo os itens/as categorias de cada nível.

Para descrever cada um dos três primeiros critérios, considera-se três níveis âncora consecutivos, K, Y e Z, com $K < Y < Z$. De acordo com Beaton e Allen (1992), um determinado item é âncora para o nível Z se, e somente se, satisfizer simultaneamente a três condições: ser respondido corretamente por 65% ou mais dos indivíduos com nível de proficiência Z; por menos de 50% dos indivíduos com nível de proficiência Y; e, a diferença entre as proporções efetivas de indivíduos que responderam corretamente ao item, nesses níveis, ser maior que 30%.

O critério adotado pelo SARESP consiste em posicionar como âncora, no nível Z, os itens respondidos corretamente por 65% ou mais dos indivíduos e por menos de 50% para o nível K, dois espaçamentos anteriores ao nível Z. Por outro lado, o SAEB posiciona como âncora, no nível Z, os itens respondidos corretamente por 65% ou mais dos indivíduos e menor do que este valor para quaisquer níveis anteriores e que tenha sido respondido, corretamente ou não, por pelo menos 50 indivíduos. A distância entre os níveis âncora é definida em unidades de desvio padrão, sendo um para Beaton e Allen (1992) e meio para o SARESP e o SAEB.

Para a descrição do posicionamento de itens a partir da segmentação da escala, considera-se a probabilidade de resposta correta para um item, condicionada a um particular nível de proficiência, apresentada em uma tabela $I \times N$, cujas linhas representam os I itens e as colunas os N níveis de proficiência para o construto a ser medido. Para a segmentação da escala utiliza-se, inicialmente, o agrupamento hierárquico aglomerativo dos níveis de proficiência de acordo com as similaridades entre os vetores coluna da referida tabela. Inicia-se com N grupos (níveis) e, a partir de algum critério de similaridade, determina-se as fusões sucessivas. Isso é feito dois a dois até que um único grupo se forme. Os níveis mais semelhantes são os primeiros agrupados e, sucessivamente, os grupos são mesclados de acordo com suas semelhanças. Conforme a similaridade diminui, todos os subgrupos são fundidos em um único agrupamento.

O método de Ward consiste no agrupamento hierárquico de níveis com base na minimização de uma função de informação. Este método é, geralmente, implementado considerando como critério a minimização da soma dos quadrados dos erros (SQE). Isto é, a minimização da soma dos desvios quadrados de cada nível do grupo em relação à média do grupo (centróide). Em cada etapa da análise, a união de cada par possível de grupos é

⁴ Para mais informações sobre a análise de agrupamentos hierárquicos e outros métodos de agrupamentos consultar Hair *et al.* (2009).

considerada e o menor aumento na SQE resulta na combinação de dois grupos. Uma vez aplicado o método de agrupamento dos níveis define-se o número de grupos (os segmentos da escala a ser considerados) e posiciona-se como âncora, dentro de um particular grupo, o item cuja estimativa para o parâmetro de discriminação é maior ou igual a um e a probabilidade de resposta correta, condicionada aos respectivos níveis, maior ou igual a 0,60.

5 Resultados e discussões

Após a aplicação da prova e tratamento dos dados realizou-se uma análise exploratória envolvendo estatísticas descritivas tais como número mínimo, máximo e médio de acertos por curso, e número de acertos, de não respostas e percentual de respostas por item. Na sequência, analisou-se os resultados para as estimativas dos parâmetros dos itens e das proficiências e efetuou-se o posicionamento dos itens na escala, de acordo com os diferentes critérios mencionados, conforme descrito nas Subseções 5.1 e 5.2.

5.1 Análise dos itens com base nas estimativas para os parâmetros do modelo

A partir da análise das estimativas obtidas para os parâmetros dos itens optou-se por excluir os itens 1 e 30. O primeiro, respondido corretamente por 274 dos 277 participantes, mostrou pouco poder de discriminação e coeficiente de correlação bisserial⁵ negativo, calculado no programa BILOG - MG. O item 30, por problemas no gabarito.

No Anexo 2, Tabela 2, são apresentadas as estimativas para os parâmetros de discriminação (a_i), de dificuldade (b_i) e de acerto ao acaso (c_i) para os 34 itens, bem como seus respectivos intervalos de 95% de credibilidade⁶, obtidos após a exclusão dos itens 1 e 30. Os valores estimados para o parâmetro a_i estão próximos ou são maiores que 1, o que é desejável. Os itens 22 e 25, com valores 0,6713 e 0,3803, respectivamente, foram mantidos por apresentarem coeficiente de correlação bisserial positivo. Por sua vez, os itens 4 e 36 apresentaram o maior poder de discriminação, com valor para o parâmetro a_i estimado em 1,8180 e 1,8250, nessa ordem. Em relação à dificuldade dos itens, o esperado é que estejam no intervalo (-3, 3). O item 10 apresentou o valor -2,5530, sendo considerado o mais fácil dos

⁵ O coeficiente de correlação bisserial, medida do poder discriminativo de um item, é baseado na diferença entre a média no teste dos sujeitos que acertaram o item e a média total do teste, em relação ao desvio padrão do teste, corrigido pelo produto das proporções de sujeitos que acertaram e que erraram o item, em relação a uma medida de correção para normalidade (PASQUALI, 2003).

⁶ Intervalos que concentram 95% dos valores estimados para a respectiva distribuição de probabilidade.

ítems, e o item 5, o valor 2,7540, sendo, portanto, o mais difícil.

O valor esperado para o parâmetro de acerto ao acaso deve estar em torno de 0,2, uma vez que os itens que compõem a prova possuem 5 alternativas e apenas uma correta. Os itens 4, 21, 31, 32 e 33, apresentaram estimativas em torno de 0,25, o que indica que uma das alternativas foi desprezada pelo respondente, aumentando a probabilidade de acerto ao acaso das alternativas restantes.

Na sequência, são apresentadas as CCI, as Curvas de Informação dos Itens (CII) e a Curva de Informação do Teste (CIT) (Figura 2). As CII permitem analisar quanto um item contém de informação para a medida de proficiência e a CIT, construída a partir da soma das informações fornecidas pelos itens que compõem o teste, tem a característica de expressar como o teste estima as proficiências em toda a sua gama de pontuação.

De maneira geral, as CCI para os 34 itens possuem comportamento característico (Figura 2(a)). Itens com valor do parâmetro de discriminação mais alto possuem curvas com maior inclinação, enquanto que aqueles com valores mais baixos possuem curva com inclinação mais suave. As curvas posicionadas mais à direita correspondem aos itens com parâmetros de dificuldade mais altos, para os quais a probabilidade de resposta correta é mais alta para indivíduos que possuem valor de proficiência mais alto. As curvas mais à esquerda correspondem aos itens com valor do parâmetro de dificuldade mais baixos, itens considerados fáceis, para os quais a probabilidade de resposta correta é alta para indivíduos com proficiência baixa. Ainda, a partir da Figura 2(a), cabe ressaltar os itens destacados. O item 22 pode ser considerado fácil, pois a probabilidade de resposta correta é de 0,55 para indivíduos com proficiência -1, na escala (0,1), ou 50, na escala (250,50). O 25 possui baixo poder de discriminação, pois indivíduos com proficiência -1 e 1, correspondente a 200 e 300 na escala (250,50), possuem probabilidade de resposta correta aproximadamente igual a 0,47 e 0,63, respectivamente. Para indivíduos com proficiência 1, 300 na escala (250,50), a probabilidade de resposta correta para o item 3 é de 0,82, enquanto que para o item 36, para esta mesma probabilidade, é necessário proficiência próxima de 3, equivalente a 400.

Na Figura 2(b) destaca-se o item 36, com maior valor para a estimativa do parâmetro de discriminação e o menor valor para a estimativa do parâmetro de acerto ao acaso, que mais contém informação para a medida de proficiência em torno de 2 (350 na escala (250,50)).

Conforme a Figura 2(c), a prova aplicada estima melhor as proficiências entre os valores -2 e 3 na escala (0,1), correspondente aos valores entre 150 e 400 na escala (250,50).

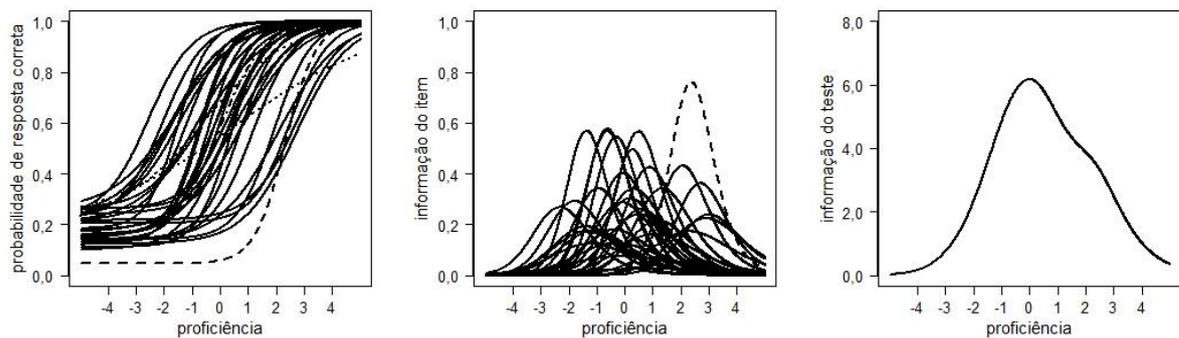


Figura 2 (a) – CCI (linha pontilhada 22 e 25, tracejada e pontilhada 3, tracejada 36); (b) – CII (tracejada 36); (c) – CIT.

Fonte: Elaborada pelos autores (2020)

5.2 Posicionamento dos itens na escala

Os resultados para o posicionamento dos itens na escala, de acordo com as quatro metodologias em estudo, descritos a seguir, são sintetizados na Figura 3.

Para o posicionamento de itens, pelo critério de Beaton e Allen (1992), especificou-se níveis âncora no intervalo de 50 a 450, intercalados por um desvio padrão, e a partir do modelo estimou-se as respectivas probabilidades de resposta correta. As três condições exigidas para posicionamento como âncora só foram satisfeitas para 7 dos 34 itens, ficando níveis âncora sem posicionamento de itens, dificultando a interpretação da escala (Figura 3(a)).

Com base na metodologia adotada pelo SARESP, especificou-se níveis âncora, na escala (250,50), intercalados por meio desvio padrão, em um intervalo de 75 a 425. Conforme Figura 3(b), 22 dos 34 itens foram posicionados, ficando também níveis sem posicionamento de itens.

Considerando o posicionamento de itens âncora pelo SAEB, que adota os mesmos níveis do SARESP, observa-se que somente os itens 25 e 36 não foram posicionados. O item 25 apresenta valor baixo para o parâmetro de discriminação e o item 36 foi respondido corretamente por apenas 27 indivíduos, indicando ser um item muito difícil. Ainda assim, para alguns níveis da escala não se tem itens posicionados, dificultando a interpretação (Figura 3(c)). No Anexo 3, Tabela 3, estão destacados em negrito os itens e as probabilidades de resposta correta, condicionada ao nível em que o item foi posicionado como âncora (critério SAEB).

Visando superar a limitação observada nessas três metodologias, poucos itens posicionados e presença de níveis sem posicionamento de itens, optou-se por classificar os níveis da escala em grupos, a partir da análise de agrupamentos hierárquicos, segmentando a escala em faixas de proficiência (BARROS, 2016). Considerou-se níveis âncora também no intervalo 75 a 425, intercalados por 0,1 unidades de desvio padrão, totalizando 71 valores. Calculou-se, então, as probabilidades de resposta correta para cada item, condicionado ao

respectivo nível, e, na sequência, realizou-se a análise de agrupamentos dos níveis.

A análise de agrupamentos hierárquicos foi realizada no software *Statistical Package for the Social Sciences (SPSS)*, a partir do método de Ward, considerando as variáveis de entrada assim dispostas: os 71 níveis de proficiência em colunas e as respectivas probabilidades de resposta correta associadas aos itens em linhas. A análise do dendrograma sugeriu, inicialmente, segmentar a escala em quatro grupos de níveis, posicionando como âncora, nestes grupos, os itens com estimativa para o parâmetro de discriminação maior ou igual a 1, no nível com probabilidade de resposta correta maior ou igual a 0,60. Esta tarefa posicionou 2, 8, 13 e 5 itens como âncora em cada um dos quatro grupos. Considerando o número de itens âncora do terceiro grupo, a saber, 13, repetiu-se o procedimento de agrupamento de níveis para este grupo, buscando melhorar a interpretação da escala. Tal procedimento resultou em dois novos grupos, com 9 e 4 itens, respectivamente. Para criar uma ideia de hierarquia e ordenamento e facilitar a interpretação, atribuiu-se um nome a cada grupo, conforme Tabela 4 e Figura 3(d).

O exposto revela que 28 dos 34 itens foram posicionados como âncora. Os seis outros, 21, 22, 25, 26, 27 e 35, apresentaram valores para o parâmetro de discriminação abaixo de 1. Verificou-se, então, em que grupo estes itens estariam posicionados, caso tivessem sido classificados como âncora, ficando a competência abordada pelo item 22 na interpretação do grupo 2 e a abordada pelo 35 na interpretação do grupo 4. Apenas os itens 21, 25, 26 e 27 não foram alocados, por não serem condizentes com os demais nos grupos.

O próximo passo após a segmentação consiste na interpretação pedagógica da escala, realizada de acordo com os itens posicionados nos grupos. A partir da Tabela 4, observa-se que 32,85% dos alunos possuem conhecimento abaixo do básico, o que evidencia a importância da interpretação pedagógica da escala na identificação da defasagem de conteúdos matemáticos dos alunos, fundamentando políticas para o enfrentamento das dificuldades dos alunos em relação à aprendizagem matemática, que poderão contribuir para minimizar a retenção e a evasão históricas nas disciplinas básicas de cursos da área de Ciências Exatas.

Tabela 4 – Intervalo da escala, porcentagem de alunos, número de itens âncora posicionados e nomes atribuídos a cada um dos grupos

Grupos	Faixa de Proficiência	% de alunos	No. Itens âncora	Nome Atribuído
1	[75 a 155]	1,08	2	Insuficiente
2]155 a 230]	31,77	8	Abaixo do Básico
3]230 a 275]	34,66	9	Básico
4]275 a 335]	29,60	4	Acima do Básico
5]335 a 425]	2,89	5	Proficiente

Fonte: Elaborada pelos autores (2020)

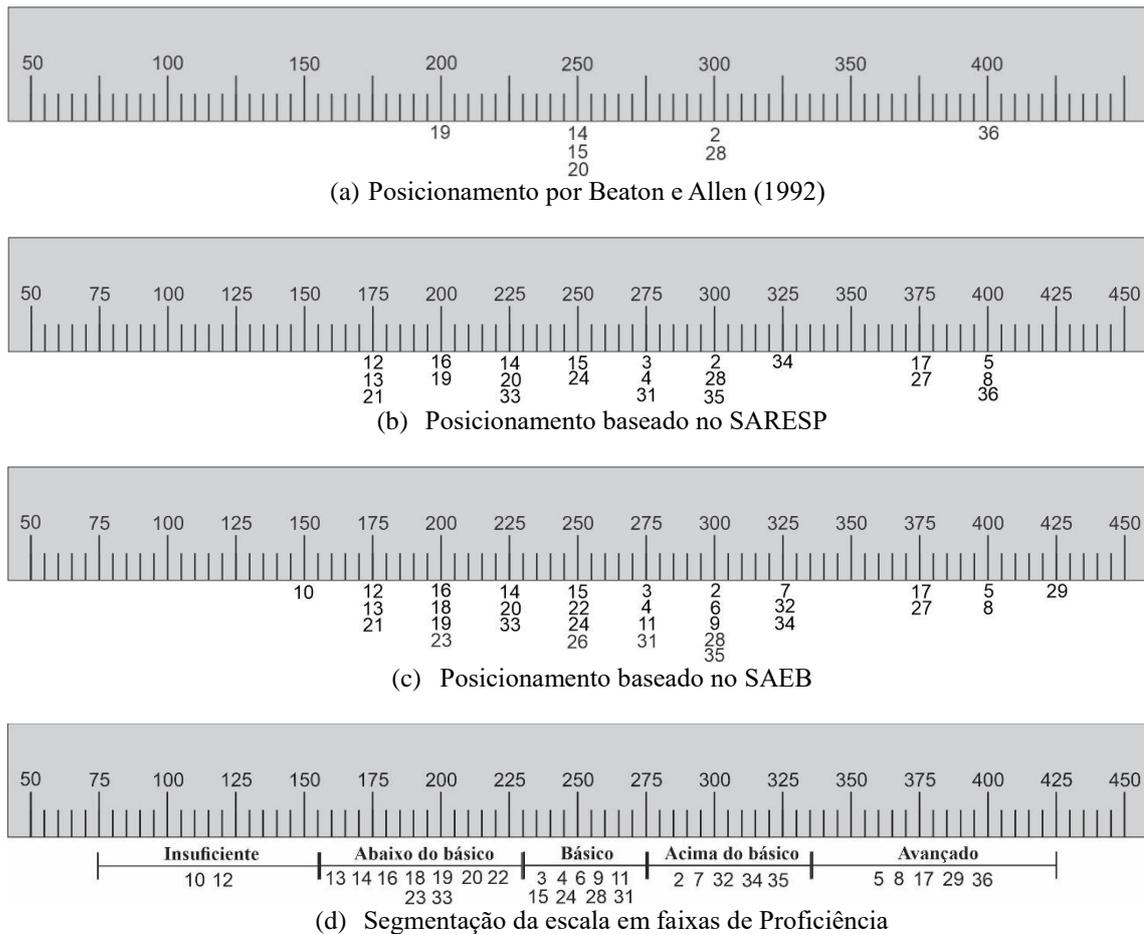


Figura 3 – Posicionamento de itens de acordo com Beaton e Allen (1992), SARESP, SAEB e a segmentação da escala em faixas de proficiências
 Fonte: Elaborada pelos autores (2020)

5.3 Interpretação pedagógica da escala

O Quadro 2 sintetiza a interpretação pedagógica da escala para cada grupo, de acordo com as competências descritas na matriz de Referência, construída com base nas competências presentes nas matrizes do SAEB e SARESP (Seção 3). É importante ressaltar que alunos com proficiência em um grupo dominam as habilidades associadas ao seu grupo e aos anteriores.

Descrição das faixas de proficiência da escala
<p>Grupo 1 - Insuficiente [75 a 155] Os alunos com proficiência estimada nesta faixa, certamente:</p> <ul style="list-style-type: none"> • Resolvem problemas com números naturais, envolvendo operações de adição e multiplicação. • Reconhecem representações de um número racional. Reconhecem a equivalência entre números racionais, apresentados na forma fracionária e efetuam a divisão do numerador pelo denominador, obtendo a representação na forma decimal. Entendem que este decimal representa a forma percentual do número racional, quando multiplicado por 100.
<p>Grupo 2 - Abaixo do Básico [155 a 230] Os alunos com proficiência estimada nesta faixa, certamente:</p> <ul style="list-style-type: none"> • Identificam fração como representação de uma situação problema. Compreendem a correspondência entre uma situação relatada e a fração que representa uma razão. Identificam o complementar de uma fração.

- Identificam frações equivalentes. Reconhecem a representação de frações equivalentes através da simplificação dos numeradores e denominadores.
- Resolvem problemas com números racionais, envolvendo as operações de adição, multiplicação e divisão. Reconhecem a terça parte de um objeto, utilizando a operação de divisão e identificam a quantidade equivalente a uma fração desse objeto, utilizando a multiplicação ou a soma das partes calculadas.
- Substituem as variáveis de uma expressão por números inteiros e calculam seu valor numérico. Efetuam, também, o cálculo de potência e multiplicação entre números negativos.
- Reconhecem regularidade em uma sequência e a representam por meio de uma expressão algébrica. Identificam a correspondência entre duas variáveis e reconhecem a expressão algébrica que a representa.
- Reconhecem a expressão algébrica que representa uma função a partir de uma tabela. Identificam a relação expressa pelos dados fornecidos, observando a partir da tabela uma regularidade a cada linha.
- Identificam o gráfico que representa uma situação descrita em um texto.
- Resolvem problemas envolvendo modelagem, por meio de uma equação do 1º grau. Transpõem os dados numéricos apresentados para uma situação geral, adaptada uma equação do tipo $y = ax + b$, ou ainda, raciocinam aritmeticamente sobre os dados numéricos, realizando operações inversas para obter a resposta.
- Identificam a representação algébrica de uma função exponencial crescente. Reconhecem uma função exponencial simples (genericamente dada por $y = a^x$ sendo $a > 0$) e identificam que para valores de a maiores que 1 a função é crescente.

Grupo 3 -Básico [230 a 275]

Os alunos com proficiência estimada nesta faixa, certamente:

- Identificam a equação de uma reta a partir de dois pontos dados. Calculam o coeficiente angular da reta e determinam a equação a partir da substituição de um dos pontos dados e do coeficiente encontrado utilizando a equação da reta tangente ou ainda, determinam a equação através da resolução de um sistema de equações.
- Identificam que o ponto de interseção de duas retas pode ser determinado algebricamente a partir da montagem de um sistema linear de duas equações.
- Resolvem problemas envolvendo cálculo de perímetro de figuras planas a partir de uma situação problema. Identificam e calculam o perímetro de uma figura.
- Identificam a localização de números reais na reta numérica. Fornecidos dois pontos em uma reta numérica, reconhecem que as abscissas desses dois pontos dados estão localizadas entre 0 e 1 e identificam que o produto de dois números menores que um é igual a um número menor que o menor deles.
- Identificam a localização de números racionais na reta numérica. Reconhecem a ordem de crescimento dos números racionais na reta e identificam corretamente os valores solicitados.
- Efetuam cálculos que envolvem operações de adição, subtração, multiplicação e potenciação com números racionais (em uma expressão algébrica do tipo $[(1 + 1/n)(1 - 1/n)]^2$, em que n é um número natural qualquer $\neq 0$). Operam adequadamente soma e subtração de números racionais, assim como o cálculo da potência e do produto.
- Reconhecem a representação algébrica de uma função de primeiro grau dado o seu gráfico. Reconhecem a equação da reta dada na forma implícita e a determinam através da análise do coeficiente angular e linear.
- Resolvem problemas que envolvem o ponto de máximo de uma função quadrática. Identificam o ponto de máximo desta função, a partir do cálculo das coordenadas do vértice, apresentado o seu gráfico e a equação da parábola. Reconhecem as coordenadas do vértice como ponto de máximo.
- Reconhecem o gráfico de uma função polinomial de primeiro grau. Identificam que o gráfico de uma função polinomial de primeiro grau é dado por uma reta crescente ou decrescente.

Grupo 4 - Acima do Básico [275 a 335]

Os alunos com proficiência estimada nesta faixa, certamente:

- Interpretam geometricamente os coeficientes da equação de uma reta. Reconhecem as propriedades da reta (dada na sua forma genérica $y = ax + b$) manipulando a equação dada, apresentada de forma implícita.
- Resolvem problemas envolvendo cálculo de área de figuras planas. Identificam figuras geométricas formadas em uma figura fornecida. Efetuam o cálculo da área de triângulos, círculos e do quadrado, realizando operações de subtração entre as áreas encontradas e chegam ao resultado solicitado.
- Relacionam as raízes de um polinômio com a sua decomposição em fatores do primeiro grau.
- Reconhecem as propriedades da função logarítmica (do tipo $\log_a x$ com $a > 1$) e identificam a função a partir de sua representação gráfica.
- Resolvem problemas que envolvam função exponencial. Identificam a função exponencial que expressa uma situação descrita no texto. Estabelecem corretamente a relação entre o valor da imagem de uma função com seu valor correspondente no domínio dessa função. Utilizam as propriedades de potenciação necessárias para a resolução.

Grupo 5 - Proficiente [335 a 425]

Os alunos com proficiência estimada nesta faixa, certamente:

- Identificam, dentre as equações do segundo grau com duas incógnitas, a que representa uma circunferência. Reconhecem as propriedades de uma circunferência (genericamente, essa figura geométrica possui centro (x_c, y_c) e raio R). Reconhecem que como qualquer de seus pontos é equidistante do centro, a distância deles ao centro é o raio R. A partir dessas propriedades, identificam a equação reduzida da circunferência utilizando-se o teorema de Pitágoras, dada por $(x - x_c)^2 + (y - y_c)^2 = R^2$. A partir desta, chegam a equação geral da circunferência, que é dada por $x^2 + y^2 - 2xx_c - 2yy_c + x_c^2 + y_c^2 - R^2 = 0$.
- Resolvem problemas envolvendo o volume de sólidos. Calculam o volume total de um cilindro e o volume de esferas efetuando subtração e divisão entre esses volumes em uma situação problema.
- Efetuam cálculos com valores aproximados de radicais. Realizam cálculos que exigem técnicas de substituição, transformação de um número racional em uma fração decimal, propriedades de radiciação e potência.
- Resolvem equações modulares e encontram o conjunto solução de uma equação modular (do tipo $|ax - b| = cx - d$ onde a, b, c e d são números naturais quaisquer). Identificam para quais valores de x a equação possui solução.
- Identificam a representação gráfica das funções trigonométricas e reconhecem suas propriedades. Distinguem corretamente a representação gráfica de funções (do tipo $y = a \sin b x$ com a e b positivos).

Quadro 2 - Descrição da Interpretação da Escala de Proficiência em Cálculo

Fonte: Elaborado pelos autores (2020)

6 Considerações finais e perspectivas futuras

Este trabalho apresenta um estudo sobre construção de escalas de proficiência com base na TRI, culminando na construção e interpretação pedagógica de uma escala para medir proficiência em conteúdos matemáticos, de ingressantes em cursos da área de Ciências Exatas.

A base para a definição do instrumento de avaliação foi construída por Rossi (2015), que elaborou uma matriz de referência e um banco de itens, comprovando a qualidade de uns e sugerindo mudanças em outros. Este instrumento, composto por 36 itens, um para cada uma das competências que compõem a referida matriz, foi aplicado em 2017 para todos os alunos matriculados no primeiro semestre do primeiro ano dos cursos da área de Ciências Exatas da FCT/Unesp. Utilizando-se do modelo ML3 os itens foram calibrados e os resultados confirmaram a eficiência para estimar a proficiência dos respondentes.

Em busca de uma adequada interpretação para escala optou-se, inicialmente, por realizar o posicionamento dos itens como âncora, nos níveis selecionados para interpretação, de acordo com os procedimentos adotados pelo SAEB, pelo SARESP e por Beaton e Allen (1992). Para os três procedimentos, em alguns dos níveis não foram posicionados itens âncora, o que dificultou a interpretação da escala. Diante disso, conforme Barros (2016), optou-se por segmentar a escala em grupos, em faixas de proficiência, utilizando a análise de agrupamentos hierárquicos. Assim, de acordo com o posicionamento dos itens âncora, caracterizou-se cada uma das faixas, possibilitando a interpretação da escala em toda a sua amplitude.

A proposta de segmentação para a construção da escala de proficiência em conteúdos matemáticos básicos e sua interpretação são contribuições consideradas relevantes neste

trabalho. Embora os resultados apresentados sejam consistentes, a curva de informação do teste mostrou que novos itens precisam ser incluídos para se obter mais informação nas extremidades da escala, o que evidencia a necessidade de um processo contínuo de aprimoramento do banco de questões, e conseqüentemente, desta escala de proficiência.

Os resultados mostraram que aproximadamente um terço dos ingressantes foram classificados nos grupos insuficiente e abaixo do básico, evidenciando a importância de uma escala, associada aos descritores da matriz de referência, e de sua interpretação para a identificação dos conteúdos matemáticos que não são dominados pelos alunos, o que pode comprometer seu desempenho em disciplinas básicas na área de Ciências Exatas.

Como perspectivas futuras considera-se o teste de novos itens para ampliar o banco de questões tendo em vista o aprimoramento da escala de proficiência, com a aplicação da prova nos anos seguintes, e a investigação da influência do meio utilizado para aplicação da prova (impresa ou via *web*) na proficiência do respondente. Considera-se, também, usar o banco de itens calibrados para transformar o instrumento de avaliação em um teste adaptativo, mais enxuto, em que os indivíduos respondem menos itens para a estimativa da proficiência.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo incentivo a esta pesquisa.

Referências

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

ANJOS, A.; ANDRADE, D. F. **Teoria de resposta ao item com o uso do R**. João Pessoa: ABE, 2012.

BARROS, G. T. F. **Procedimentos para a construção de indicadores por meio da teoria de resposta ao item: a criação de uma medida de nível socioeconômico familiar**. 2016. Dissertação (Mestrado em Métodos e Gestão em Avaliação) – Universidade Federal de Santa Catarina, Florianópolis, 2016.

BEATON, A. E.; ALLEN, N. L. Interpreting scales through scale anchoring. **Journal of Educational Statistics**, Los Angeles, v. 17, n. 2, p. 191-204, 1992.

BRASIL. Ministério da Educação. **PDE: Plano de Desenvolvimento da Educação: Prova Brasil: ensino fundamental: matrizes de referência, tópicos e descritores**. Brasília: MEC, SEB, Inep, 2008a. Disponível em: http://portal.mec.gov.br/dmdocuments/prova%20brasil_matriz2.pdf. Acesso em: 21 ago. 2020.

BRASIL. Ministério da Educação. **PDE: Plano de Desenvolvimento da Educação: SAEB: ensino médio: matrizes de referência, tópicos e descritores.** Brasília: MEC, SEB; Inep, 2008b. Disponível em: http://portal.mec.gov.br/dmdocuments/saeb_matriz2.pdf. Acesso em: 21 ago. 2020.

GOMES, F. H. **Uma proposta de exame de proficiência em cálculo diferencial e integral.** 2016. Dissertação (Mestrado Profissional em Matemática) – Universidade de Brasília, Brasília, 2016.

HAIR, J. F. *et al.* **Análise multivariada de dados.** 6. ed. Porto Alegre: Bookman, 2009.

LUNN, D.; SPIEGELHALTER, D.; BEST, N. The bugs project: evolution, critique and future directions. **Statistics in Medicine**, [s. l.], v. 28, n. 25, p. 3049-3067, 2009.

MIGON, H. S.; GAMERMAN, D.; LOUZADA, F. **Statistical inference: an integrated approach.** 2. ed. Boca Raton: CRC Press, 2014.

MONTEIRO, M. H.; AFONSO, M. J.; PIRES, M. A avaliação objetiva dos conhecimentos de Matemática à entrada do Ensino Superior de ciências e tecnologias: construção e resultados de um teste estandardizado de conhecimentos – PMAT. *In*: CONGRESSO NACIONAL DE PRÁTICAS PEDAGÓGICAS NO ENSINO SUPERIOR - CNaPPES, 3., 2016, Lisboa. **Anais eletrônicos [...]**. Lisboa: Universidade de Lisboa, 2017. p. 13-18. Disponível em: <https://cnappes.org/cnappes-2016/files/2014/03/Livro-de-Atas-do-CNaPPES-2016-3.pdf>. Acesso em: 26 ago. 2020.

PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação.** 2. ed. Petrópolis: Vozes, 2003.

ROSSI, P. R. **Construção de uma escala para avaliação da proficiência em conteúdos matemáticos básicos.** 2015. Dissertação (Mestrado em Matemática Aplicada e Computacional) – Universidade Estadual Paulista, Presidente Prudente, 2015.

SOARES, D. J. M. **Teoria clássica dos testes e Teoria de Resposta ao Item aplicadas em uma avaliação de matemática básica.** 2018. 133 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa, Viçosa, 2018.

**Submetido em 08 de Outubro de 2020.
Aprovado em 09 de Julho de 2021.**

Anexo 1

Temas	Descritores
Espaço e Forma	D1 - Identificar a localização de pontos no plano cartesiano; D2 - Interpretar geometricamente os coeficientes da equação de uma reta; D3 - Identificar a equação de uma reta apresentada a partir de dois pontos dados ou de um ponto e sua inclinação; D4 - Relacionar a determinação do ponto de interseção de duas ou mais retas; D5 - Reconhecer, dentre as equações do segundo grau com duas incógnitas, as que representam circunferências.
Grandezas e Medidas	D6 - Resolver problemas envolvendo cálculo de perímetro de figura planas; D7 - Resolver problemas envolvendo cálculo de área de figuras planas; D8 - Resolver problemas envolvendo a área total e/ou volume de um sólido (prisma, pirâmide, cilindro, cone, esfera).
Números e Operações, Álgebra e Funções	D9 - Identificar a localização de números reais na reta numérica; D10 - Resolver problemas com números naturais e inteiros envolvendo diferentes significados das operações (adição, subtração, multiplicação, divisão e potenciação); D11 - Identificar a localização de números racionais na reta numérica; D12 - Reconhecer diferentes representações de um número racional; D13 - Identificar fração como representação que pode estar associada a diferentes significados; D14 - Identificar frações equivalentes; D15 - Efetuar cálculos que envolvam operações (adição, subtração, multiplicação, divisão e potenciação) com números racionais; D16 - Resolver problemas com números racionais que envolvam as operações: adição, subtração, multiplicação, divisão e potenciação; D17 - Efetuar cálculos com valores aproximados de radicais; D18 - Calcular o valor numérico de uma expressão algébrica; D19 - Identificar a expressão algébrica que expressa uma regularidade observada em seqüências de números ou figuras (padrões); D20 - Reconhecer expressão algébrica que representa uma função a partir de uma tabela; D21 - Analisar crescimento/decrescimento, zeros de funções reais apresentadas em gráficos; D22 - Identificar o gráfico que representa uma situação descrita no texto; D23 - Identificar e resolver problemas que envolvam funções de primeiro grau; D24 - Reconhecer a representação algébrica de uma função do primeiro grau, dado o seu gráfico; D25 - Identificar e resolver problemas de inequações de primeiro grau; D26 - Identificar e resolver problemas que envolvam equações de segundo grau; D27 - Identificar e resolver problemas de inequações de segundo grau; D28 - Resolver problemas que envolvam os pontos de máximo ou de mínimo no gráfico de uma função polinomial de segundo grau; D29 - Identificar e resolver equações modulares; D30 - Identificar e resolver inequações modulares; D31 - Reconhecer o gráfico de uma função polinomial de primeiro grau por meio de seus coeficientes; D32 - Relacionar as raízes de um polinômio com sua decomposição em fatores do primeiro grau; D33 - Identificar a representação gráfica e/ou algébrica de uma função exponencial; D34 - Identificar a representação gráfica e/ou algébrica de uma função logarítmica, reconhecendo-a como inversa da função exponencial; D35 - Resolver problemas que envolvam função exponencial D36 - Identificar a representação gráfica e/ou algébrica de funções trigonométricas (seno, cosseno, tangente), reconhecendo suas propriedades.

Quadro 1 – Matriz de Referência

Fonte: Rossi (2015)

Anexo 2
Tabela 2 – Estimativas para os parâmetros a_i de discriminação, b_i de dificuldade e c_i de acerto ao acaso e seus respectivos intervalos de credibilidade (IC) de 95%.

Item	\hat{a}_i	IC	\hat{b}_i	IC	\hat{c}_i	IC
2	1,49600	[0,83670; 2,55000]	0,72680	[0,32780; 1,1500]	0,14180	[0,02925; 0,26020]
3	1,28500	[0,79120; 2,03800]	-0,00850	[-0,47250; 0,5110]	0,16120	[0,02844; 0,33450]
4	1,81800	[0,95660; 3,03200]	0,08940	[-0,45180; 0,5212]	0,26570	[0,07930; 0,41570]
5	1,20600	[0,33360; 2,62400]	2,75400	[1,82300; 4,1660]	0,21740	[0,04753; 0,31440]
6	1,13200	[0,66130; 1,84100]	0,30040	[-0,20000; 0,8530]	0,15130	[0,02504; 0,31570]
7	0,95690	[0,44940; 1,82400]	0,79820	[0,13840; 1,5620]	0,17550	[0,03178; 0,34280]
8	1,46700	[0,44120; 2,87100]	2,50200	[1,74200; 3,7260]	0,20540	[0,07119; 0,28640]
9	1,29000	[0,84990; 1,82500]	0,18030	[-0,15780; 0,5720]	0,10170	[0,01690; 0,22870]
10	1,29600	[0,77980; 1,97500]	-2,55300	[-3,63100; -1,7200]	0,23340	[0,03271; 0,53380]
11	1,23300	[0,78030; 1,83300]	-0,26950	[-0,69990; 0,2179]	0,14600	[0,02108; 0,31060]
12	1,33900	[0,82930; 1,99200]	-2,02100	[-2,84100; -1,3180]	0,22130	[0,03428; 0,50350]
13	1,04900	[0,63030; 1,60300]	-1,68400	[-2,67800; -0,8086]	0,23830	[0,03922; 0,52110]
14	1,76500	[1,17800; 2,55400]	-0,75550	[-1,14800; -0,3545]	0,15650	[0,02926; 0,33730]
15	1,69200	[1,15400; 2,42600]	-0,43550	[-0,80090; -0,0571]	0,14170	[0,02132; 0,31200]
16	1,44600	[0,89740; 2,14200]	-1,15300	[-1,80300; -0,5252]	0,21860	[0,03341; 0,46920]
17	1,50600	[0,64420; 2,78800]	1,91900	[1,38300; 2,7310]	0,14170	[0,04770; 0,22000]
18	1,04600	[0,60910; 1,59300]	-1,39900	[-2,37400; -0,5261]	0,22900	[0,03600; 0,49590]
19	1,80000	[1,14800; 2,60300]	-1,51100	[-2,05300; -1,0290]	0,18350	[0,03074; 0,41650]
20	1,69500	[1,19700; 2,32600]	-0,73400	[-1,07300; -0,3828]	0,12200	[0,01755; 0,28150]
21	0,90310	[0,50640; 1,44700]	-1,61400	[-2,78500; -0,4926]	0,25800	[0,03671; 0,55520]
22	0,67130	[0,31740; 1,13500]	-0,58280	[-1,67000; 0,6112]	0,21350	[0,03616; 0,45220]
23	1,07900	[0,64140; 1,62300]	-1,66700	[-2,64600; -0,8875]	0,21580	[0,03287; 0,47290]
24	1,48300	[0,93470; 2,20100]	-0,22880	[-0,65160; 0,2051]	0,16170	[0,03033; 0,32670]
25	0,38030	[0,05865; 1,02000]	0,55930	[-1,08500; 2,3960]	0,18230	[0,03040; 0,43620]
26	0,86730	[0,47210; 1,40400]	-0,62760	[-1,48600; 0,2729]	0,20400	[0,03051; 0,43700]
27	0,94600	[0,43440; 1,84900]	2,08300	[1,34500; 3,1650]	0,13230	[0,02423; 0,24490]
28	1,67600	[1,07300; 2,49800]	0,38140	[0,08397; 0,7155]	0,11140	[0,02370; 0,21800]
29	1,06800	[0,46950; 2,06700]	2,71700	[1,85300; 4,0130]	0,12370	[0,03404; 0,20500]
31	1,31200	[0,60740; 2,48000]	0,32820	[-0,36680; 0,9240]	0,26060	[0,06184; 0,43040]
32	1,17400	[0,43650; 2,33400]	0,91170	[0,17950; 1,6590]	0,24220	[0,05517; 0,39040]
33	1,09000	[0,58790; 1,87600]	-0,67380	[-1,54500; 0,1823]	0,25970	[0,04141; 0,50060]
34	1,34200	[0,70770; 2,30500]	1,18700	[0,74410; 1,7380]	0,14100	[0,03386; 0,24900]
35	0,94950	[0,50090; 1,66700]	0,47590	[-0,12730; 1,1880]	0,16820	[0,03189; 0,34160]
36	1,82500	[0,98140; 3,01800]	2,33000	[1,80400; 3,1000]	0,04620	[0,01182; 0,08515]

Fonte: Elaborada pelos autores.

Anexo 3

Tabela 3 – Probabilidade de resposta correta, para cada um dos itens considerados, condicionada ao nível de proficiência na escala (250,50) intercalados por meio desvio padrão.

Níveis na escala (250,50) intercalados por meio desvio padrão (*)															
Item	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425
2	0,14	0,15	0,15	0,16	0,17	0,20	0,26	0,36	0,50	0,66	0,79	0,89	0,94	0,97	0,99
3	0,17	0,18	0,19	0,22	0,27	0,34	0,45	0,58	0,71	0,82	0,89	0,94	0,97	0,98	0,99
4	0,27	0,27	0,27	0,28	0,30	0,35	0,45	0,60	0,76	0,88	0,95	0,98	0,99	1,00	1,00
5	0,22	0,22	0,22	0,22	0,22	0,23	0,23	0,24	0,27	0,30	0,36	0,44	0,55	0,67	0,77
6	0,16	0,17	0,19	0,21	0,25	0,31	0,40	0,50	0,62	0,74	0,83	0,89	0,94	0,96	0,98
7	0,19	0,20	0,21	0,23	0,26	0,30	0,36	0,44	0,53	0,63	0,72	0,80	0,86	0,91	0,94
8	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,23	0,25	0,28	0,35	0,46	0,60	0,74	0,85
9	0,11	0,12	0,13	0,15	0,19	0,26	0,37	0,50	0,64	0,77	0,86	0,92	0,96	0,98	0,99
10	0,41	0,51	0,63	0,75	0,84	0,91	0,95	0,97	0,99	0,99	1,00	1,00	1,00	1,00	1,00
11	0,16	0,17	0,20	0,24	0,30	0,39	0,51	0,64	0,76	0,85	0,91	0,95	0,97	0,99	0,99
12	0,32	0,39	0,49	0,62	0,74	0,84	0,91	0,95	0,97	0,99	0,99	1,00	1,00	1,00	1,00
13	0,34	0,39	0,47	0,56	0,66	0,75	0,83	0,89	0,93	0,96	0,97	0,98	0,99	0,99	1,00
14	0,16	0,17	0,19	0,24	0,34	0,49	0,67	0,82	0,92	0,96	0,98	0,99	1,00	1,00	1,00
15	0,15	0,15	0,17	0,20	0,26	0,38	0,55	0,72	0,85	0,93	0,97	0,99	0,99	1,00	1,00
16	0,24	0,27	0,32	0,40	0,51	0,65	0,78	0,88	0,93	0,97	0,98	0,99	1,00	1,00	1,00
17	0,14	0,14	0,14	0,14	0,15	0,15	0,16	0,19	0,23	0,31	0,44	0,60	0,75	0,86	0,93
18	0,31	0,35	0,41	0,50	0,59	0,69	0,78	0,86	0,91	0,94	0,96	0,98	0,99	0,99	1,00
19	0,21	0,24	0,30	0,42	0,60	0,77	0,89	0,95	0,98	0,99	1,00	1,00	1,00	1,00	1,00
20	0,13	0,14	0,16	0,21	0,31	0,46	0,65	0,80	0,90	0,96	0,98	0,99	1,00	1,00	1,00
21	0,37	0,42	0,49	0,56	0,65	0,73	0,80	0,86	0,90	0,94	0,96	0,97	0,98	0,99	0,99
22	0,31	0,34	0,38	0,43	0,49	0,55	0,62	0,68	0,74	0,80	0,84	0,88	0,91	0,93	0,95
23	0,31	0,37	0,44	0,54	0,64	0,74	0,83	0,89	0,93	0,96	0,98	0,99	0,99	0,99	1,00
24	0,17	0,18	0,19	0,22	0,27	0,36	0,50	0,65	0,79	0,88	0,94	0,97	0,99	0,99	1,00
25	0,33	0,35	0,38	0,41	0,44	0,47	0,51	0,55	0,59	0,63	0,66	0,70	0,74	0,77	0,80
26	0,26	0,29	0,34	0,39	0,46	0,54	0,62	0,71	0,78	0,84	0,89	0,93	0,95	0,97	0,98
27	0,14	0,14	0,14	0,15	0,16	0,18	0,20	0,24	0,29	0,36	0,45	0,55	0,65	0,74	0,82
28	0,11	0,11	0,12	0,13	0,15	0,19	0,28	0,42	0,60	0,77	0,88	0,94	0,98	0,99	1,00
29	0,12	0,13	0,13	0,13	0,13	0,14	0,15	0,17	0,20	0,24	0,31	0,40	0,51	0,63	0,74
31	0,27	0,27	0,28	0,29	0,32	0,37	0,45	0,55	0,67	0,78	0,87	0,93	0,96	0,98	0,99
32	0,25	0,25	0,26	0,27	0,28	0,31	0,36	0,44	0,53	0,64	0,75	0,83	0,90	0,94	0,97
33	0,29	0,31	0,35	0,40	0,47	0,56	0,66	0,76	0,84	0,90	0,94	0,96	0,98	0,99	0,99
34	0,14	0,14	0,15	0,15	0,16	0,18	0,22	0,29	0,39	0,52	0,66	0,78	0,87	0,93	0,96
35	0,19	0,20	0,21	0,24	0,28	0,33	0,40	0,49	0,59	0,69	0,77	0,84	0,89	0,93	0,96
36	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,06	0,08	0,12	0,22	0,38	0,60	0,78	0,90

(*) Destacados em negrito os itens e as probabilidades nos níveis em que foram posicionados como âncora (SAEB).

Fonte: Elaborada pelos autores.