



UNIVERSIDADE FEDERAL DA GRANDE DOURADOS - UFGD

Edvair Alves Ferreira

TEORIA DE RESPOSTA AO ITEM - TRI
Análise de algumas questões do ENEM - habilidades 24 a 30

Dourados - MS

2018

Edvair Alves Ferreira

TEORIA DE RESPOSTA AO ITEM - TRI
Análise de algumas questões do ENEM - habilidades 24 a 30

Dissertação apresentada ao Mestrado Profissional em Matemática da Faculdade de Ciências Exatas e Tecnologia da Universidade Federal da Grande Dourados, como requisito parcial para obtenção do título de Mestre.

Universidade Federal da Grande Dourados - UFGD
Mestrado Profissional em Matemática - PROFMAT

Orientador: Prof. Dr. Alexandre Pitangui Calixto

Dourados - MS
2018

Dados Internacionais de Catalogação na Publicação (CIP).

F383t	<p>Ferreira, Edvair Alves Teoria de resposta ao item – TRI : análise de algumas questões do ENEM – habilidades 24 a 30 / Edvair Alves Ferreira. – 2018. 87f.</p> <p>Orientador: Prof. Dr. Alexandre Pitanguí Calixto. Dissertação (Mestrado Profissional em Matemática) –Universidade Federal da Grande Dourados.</p> <p>1. Teoria de Resposta ao Item. 2. Estatística 3. Probabilidade. I. Título.</p>
-------	---

Ficha catalográfica elaborada pela Biblioteca Central da UFGD.

©Todos os direitos reservados. Permitido a publicação parcial desde que citada a fonte.

Dedicado a:

Minha mãe Izaura, que me ensinou a ler e escrever usando papel de pão;

Meu pai João, que comprou o primeiro caderno de brochura;

A minha esposa Talita, que sempre acreditou em mim (muito mais do que eu) e na ideia de que só a Educação é capaz de promover mudanças significativas;

A meus filhos, Dênis, Deiner e Caio, fontes eternas de inspiração.

Agradecimentos

A minha esposa Talita, pela paciência, dedicação, amor e compreensão durante nossa vida conjunta;

A meus filhos, Dênis, Deiner e Caio, por acreditarem no que sempre pregamos, educação é sempre o melhor caminho;

A meus amigos de turma, pela ajuda e companheirismo, em especial, a Márcia pelo sofrimento mútuo nas horas de estudos noturno, Eder e Katiuce, por tornar as viagens menos penosas;

A todos meus colegas do trabalho (UFMS), pela compreensão, principalmente a Márcia Paula e a Leslie, pela amizade e incentivo;

A meu Orientador, Alexandre, pela ajuda, ensinamentos e correta condução dos trabalhos;

A todos os professores do Profmat pelas ajudas, mesmo fora dos horários de aula e, especialmente a professora Irene, por acreditar em nosso potencial;

A minha melhor amiga Alessandra Fujimori, estendido a sua família, inclusive Dona Eva (*in memoriam*), que muito ajudaram em minha formação de vida e acadêmica.

“De fato, nunca se sabe muito bem para que servem as vitórias, suspirou o professor de matemática, Mas as derrotas sabe-se muito bem para que servem.”

José Saramago em O homem duplicado

Resumo

Este estudo objetivou analisar questões aplicadas no Exame Nacional do Ensino Médio - ENEM, referente as áreas de competência seis e sete que contemplam as habilidades de vinte e quatro a trinta, (a avaliação de matemática é dividida em sete competências que são subdivididas em trinta habilidades) comparando o resultado de um teste aplicado com sete itens a alunos de algumas escolas de Mato Grosso do Sul, com os resultados obtidos de examinados que responderam esses mesmos itens no ENEM em anos anteriores, com o intuito de verificar como se desenvolveu a leitura e análise de representações estatísticas. As análises foram realizadas utilizando as teorias clássicas dos testes - TCT e a teoria de resposta ao item - TRI. Foi aplicado o teste a 206 (duzentos e seis) alunos de escolas públicas das cidades de Campo Grande - MS, Bonito - MS e Dourados - MS e comparados com o resultado de 12.347 (doze mil trezentos e quarenta e sete) examinados retirados do relatório microdados do ENEM. Com a teoria clássica do teste foi computada a proporção dos acertos no teste e com a teoria de resposta ao item, modelo logístico unidimensional de três parâmetros, estimativa bayesiana e estimativa a priori calcularam-se os parâmetros de discriminação, dificuldade e acerto casual, correlação bisserial e os gráficos curva característica do item e curva de informação do item. O teste da amostra aplicada apresentou valores aceitáveis enquanto que os do ENEM não. O teste mostrou que os alunos da amostra aplicada nas escolas tiveram melhores resultados do que aqueles obtidos no ENEM, porém revelou que é necessário um ajuste importante no ensino de probabilidade e estatística para que os alunos das escolas públicas tenham um bom letramento e raciocínio estatístico e probabilístico.

Palavras-chave: Teoria de Resposta ao Item. Teoria Clássica dos Testes. Enem. Estatística. Probabilidade.

Abstract

This study aimed to analyse the items from the Exame Nacional do Ensino Médio-ENEM, that belong to the number six and seven areas and involve the skills from twenty-four to thirty (the Mathematics Exam is organized into seven abilities that are subdivided into thirty), comparing the result of a seven-item test applied in some schools in Mato Grosso do Sul to the results obtained in previous ENEM exams, with the objective of verifying how the reading and the analysis of the statistics representations were developed. The analyses were done using the classical test theory – CTT and the item response theory – IRT. The test was applied to 206 (two hundred and six) students of public schools in the cities of Campo Grande – MS, Bonito – MS and Dourados – MS and compared to the result of 12.347 (twelve thousand three hundred and forty-seven) ENEM participants obtained from the Exam microdata. By the classical test theory, the proportions of right answers in the test were computed. By the item response theory, the three parameter logistic model, the Bayesian estimate and the a-priori estimate, the discrimination parameters, the difficulty, the casual right answers, biserial correlation, and the item characteristic curve and item information curve graphics were calculated. The test of the applied sample presented the acceptable values whereas the ones from ENEM didn't. The test showed that the students from the applied sample in schools obtained better results than the students in ENEM, but it revealed that an important change is necessary in the teaching of probability and statistics so that the public school students can have good literacy and statistics and probability thinking.

Keywords: Item Response Theory. Classical Test Theory. Enem. Statistics. Probability.

Lista de ilustrações

Figura 1	– Modelo Curva Característica TRI	28
Figura 2	– Modelo Gráfico Função Informação - TRI	32
Figura 3	– Questão referente a Habilidade 24	50
Figura 4	– Curva Característica do Item H24 - Amostra	51
Figura 5	– Curva Característica do Item H24 - ENEM	52
Figura 6	– Questão referente a Habilidade 25	53
Figura 7	– Curva Característica do Item H25 - Amostra	54
Figura 8	– Curva Característica do Item H25 - ENEM	56
Figura 9	– Questão referente a Habilidade 26	57
Figura 10	– Curva Característica do Item H26 - Amostra	58
Figura 11	– Curva Característica do Item H26 - ENEM	59
Figura 12	– Questão referente a Habilidade 26	60
Figura 13	– Curva Característica do Item H27 - Amostra	61
Figura 14	– Curva Característica do Item H27 - ENEM	63
Figura 15	– Questão referente a Habilidade 28	64
Figura 16	– Curva Característica do Item H28 - Amostra	65
Figura 17	– Curva Característica do Item H28 - ENEM	66
Figura 18	– Questão referente a Habilidade 29	67
Figura 19	– Curva Característica do Item H29 - Amostra	68
Figura 20	– Curva Característica do Item H29 - ENEM	69
Figura 21	– Questão referente a Habilidade 30	70
Figura 22	– Curva Característica do Item H30 - Amostra	72
Figura 23	– Curva Característica do Item H30 - ENEM	73
Figura 24	– CCI de todos os Itens - Amostra	76
Figura 25	– CCI de todos os Itens - ENEM	76
Figura 26	– Curva Característica do Teste	77
Figura 27	– CII de todos os Itens - Amostra	77
Figura 28	– CII de todos os Itens - ENEM	78
Figura 29	– Função Informação do Teste	78
Figura 30	– Parametros com Erro Padrão	84
Figura 31	– Parametros com Erro Padrão	86

Lista de tabelas

Tabela 1	–	Classificação de acordo com o parâmetro a	29
Tabela 2	–	Classificação de acordo com o parâmetro b	30
Tabela 3	–	Tabela de Índices da Habilidade H24 - AMOSTRA	51
Tabela 4	–	Tabela de Índices da Habilidade H24 - ENEM	52
Tabela 5	–	Tabela de Índices da Habilidade H25 - Amostra	54
Tabela 6	–	Tabela de Índices da Habilidade H25 - ENEM	55
Tabela 7	–	Tabela de Índices da Habilidade H26 - Amostra	57
Tabela 8	–	Tabela de Índices da Habilidade H25 - ENEM	59
Tabela 9	–	Tabela de Índices da Habilidade H27 - Amostra	61
Tabela 10	–	Tabela de Índices da Habilidade H27 - ENEM	62
Tabela 11	–	Tabela de Índices da Habilidade H28 - Amostra	64
Tabela 12	–	Tabela de Índices da Habilidade H28 - ENEM	66
Tabela 13	–	Tabela de Índices da Habilidade H29 - Amostra	68
Tabela 14	–	Tabela de Índices da Habilidade H29 - ENEM	69
Tabela 15	–	Tabela de Índices da Habilidade H30 - Amostra	71
Tabela 16	–	Tabela de Índices da Habilidade H30 - ENEM	73

Sumário

	Introdução	11
1	EDUCAÇÃO MATEMÁTICA, EDUCAÇÃO ESTATÍSTICA E ENEM	15
1.1	Educação e Educação Matemática	15
1.2	Educação Estatística e Probabilidade	16
1.3	Exame Nacional do Ensino Médio - ENEM	19
2	MODELOS MATEMÁTICOS	21
2.1	Teoria Clássica dos Testes - TCT	22
2.2	Teoria de Resposta ao Item - TRI	24
2.2.1	Modelos dicotômicos ou dicotomizados	24
2.2.2	O modelo logístico de 3 parâmetros (ML3)	26
2.2.3	Curva Característica do Item - CCI	27
2.2.4	O Parâmetro habilidade (θ)	28
2.2.5	O Parâmetro de Discriminação (a)	28
2.2.6	O Parâmetro de dificuldade (b)	29
2.2.7	O Parâmetro de acerto ao acaso (c)	30
2.2.8	Função de Informação do Item	31
2.2.9	Função de Informação do Teste	32
2.2.10	Unidimensionalidade	33
2.2.11	Independência local	33
2.2.12	Estimação	34
2.2.13	Estimação conjunta: parâmetros dos itens e habilidades	35
2.2.14	Máxima Verossimilhança Marginal	36
2.2.15	Estimação bayesiana	42
2.2.16	Estimação dos parâmetros	43
2.2.17	Estimação das habilidades	46
2.2.18	Aplicação do Algoritmo EM	47
3	ANÁLISE PEDAGÓGICA DOS ITENS	49
3.1	Habilidade 24 - H24	50
3.1.1	Questão H24	50
3.1.2	Tabelas, Gráficos e Comentários	51
3.2	Habilidade 25 - H25	53
3.2.1	Questão H25	53
3.2.2	Tabelas, Gráficos e Comentários	54

3.3	Habilidade 26 - H26	56
3.3.1	Questão H26	57
3.3.2	Tabelas, Gráficos e Comentários	57
3.4	Habilidade 27 - H27	60
3.4.1	Questão H27	60
3.4.2	Tabelas, Gráficos e Comentários	61
3.5	Habilidade 28 - H28	63
3.5.1	Questão H28	64
3.5.2	Tabelas, Gráfico e Comentários	64
3.6	Habilidade 29 - H29	67
3.6.1	Questão H29	67
3.6.2	Tabelas, Gráficos e Comentários	68
3.7	Habilidade 30 - H30	70
3.7.1	Questão H30	70
3.7.2	Tabelas, Gráficos e Comentários	71
4	CONSIDERAÇÕES FINAIS	75
4.1	Conclusão	79
	REFERÊNCIAS	80
	APÊNDICES	82
	APÊNDICE A – AMOSTRA	83
	APÊNDICE B – ENEM	85

Introdução

A Estatística vem sendo utilizada de várias maneiras para demonstrar os acontecimentos do país e do mundo. É necessário um entendimento maior sobre como as análises são realizadas e apresentadas. Dessa forma, o ensino da estatística nas várias séries do ensino básico vem para ajudar os alunos a entenderem melhor tudo que lhes é apresentado na forma de tabela, gráfico e dados probabilísticos. O ensino de matemática, especialmente o ensino de estatística, carece de formas de interpretar o mundo abstrato através do mundo concreto, para que o aluno tenha uma melhor percepção do que está sendo estudado. Desde muitos anos, estudiosos já discutiam maneiras e tinham em mente o quão é necessário essa interligação entre o abstrato e o concreto.

Segundo Comenius, por volta de 1650, o ensino deveria dar-se do concreto ao abstrato, justificando que o conhecimento começa pelos sentidos e que só se aprende fazendo. Locke, em 1680, dizia da necessidade da experiência sensível para alcançar o conhecimento. Cerca de cem anos depois, Rousseau recomendou a experiência direta sobre os objetos, visando à aprendizagem. Pestalozzi e Froebel, por volta de 1800, também reconheceram que o ensino deveria começar pelo concreto; na mesma época, Herbart defendeu que a aprendizagem começa pelo campo sensorial. Pelos idos de 1900, Dewey concordava com o pensamento de Comenius, ressaltando a importância da experiência direta como fator básico para construção do conhecimento, e Poincaré recomendava o uso de imagens vivas para clarear verdades matemáticas. Mais recentemente, Montessori legou-nos inúmeros exemplos de materiais didáticos e atividades de ensino que valorizam a aprendizagem através dos sentidos, especialmente do tátil, enquanto Piaget deixou claro que o conhecimento se dá pela ação repetida sobre o objeto; Vygotsky, na Rússia, e Bruner, nos Estados Unidos, concordaram que as experiências no mundo real constituem o caminho para a criança e o adolescente construírem seu raciocínio (LORENZATO, 2006 apud VENCESLAU, 2015, p. 14) .

O excerto acima, mostra claramente a preocupação de vários pensadores sobre como o ensino matemático deva ser apresentado àqueles que necessitam do aprendizado. É notório que com a prática, a noção daquilo que está escrito fica bem mais claro. Este trabalho foi pensado com o intuito de saber como o ensino de probabilidade e estatística está sendo ministrado e o quanto está sendo aprendido nas escolas públicas de Mato Grosso do Sul.

As motivações vieram, principalmente, com a mudança na Base Nacional Comum Curricular - BNCC, do Ensino Fundamental e do Ensino Médio nos quais, se deu grande abertura para o ensino de probabilidade e estatística, haja visto que a BNCC, contempla esse ensino desde as séries iniciais. Ainda, a necessidade de um levantamento, mesmo que superficial, sobre a absorção, pelos alunos, do conteúdo de probabilidade e estatística no

Ensino Fundamental e Médio e finalmente, a necessidade de observar a forma de cobrança desse conteúdo no Enem - Exame Nacional do Ensino Médio.

A Nova Base Nacional Comum Curricular (BNCC) prega que para o desenvolvimento das competências é necessário que os estudantes possam ser capazes de raciocinar, representar, comunicar-se, argumentar e terem noção de variação e constância, certeza e incerteza, movimento e posição, relações e inter-relações.

A BNCC do Ensino Médio prega que

Certeza e incerteza é um par normalmente associado, na matemática escolar, ao estudo de fenômenos aleatórios, à obtenção de medidas no mundo físico, a estimativas, análises e inferências estatísticas e a argumentações e demonstrações algébricas ou geométricas. Mas ela engloba muitas outras ideias. Inter-relações devem ser destacadas, por exemplo, entre a Estatística e a Probabilidade. As medidas estatísticas também são inter-relações que, por um lado, promovem a ampliação de significado aos tópicos matemáticos e, por outro, propiciam acesso à matemática estrutural, como as relações de equivalência, ordem e semelhança. (BNCC, 2018, p. 519, 521)

As competências específicas de matemática e suas tecnologias para o ensino médio, na competência três definem que devem-se utilizar de estratégias, conceitos e procedimentos matemáticos, entre outros campos, de Probabilidade e Estatística, para que o aluno tenha condições de construir modelos e resolver problemas em diversos contextos, analisando a plausibilidade dos resultados e a adequação das soluções propostas, de modo a construir argumentação consistente.

Além disso, a habilidade “EM13MAT311” pretende que o aluno possa resolver e elaborar problemas que envolvem o cálculo da probabilidade de eventos aleatórios, identificando e descrevendo o espaço amostral e realizando contagem das possibilidades; a habilidade “EM13MAT312”, resolver e elaborar problemas que envolvem o cálculo de probabilidade de eventos em experimentos aleatórios sucessivos; a “EM13MAT316” resolver e elaborar problemas, em diferentes contextos, que envolvem cálculo e interpretação das medidas de tendência central (média, moda, mediana) e das de dispersão (amplitude, variância e desvio padrão).

Portanto, faz-se necessário uma análise prévia para que posteriormente possam haver novos trabalhos que sirvam como meio de comparação sobre a evolução ou a involução do ensino de probabilidade e estatística no ensino fundamental e médio com a implantação da nova BNCC.

Para tanto, em todos os anos em que foi realizado o Enem a partir de 2009, no final o Inep emite arquivo detalhado, que, entre as várias informações, apresenta dados dos inscritos, como sócio-econômico, escolar, resultados das provas, opções marcadas e

cada item e etc. O arquivo é denominado *Microdados do ENEM*¹.

Utilizando desses dados, separamos todas as questões referente ao ensino de estatística, dentro das habilidades específicas, (24 a 30), das provas aplicadas em 2013, 2014 e 2015 e obtivemos os resultados dos inscritos de uma amostra de 12.347 (doze mil, trezentos e quarenta e sete) candidatos de Mato Grosso do Sul, que realizaram a prova.

Segundo dados do Inep e Balanços apresentados pelo Portal MEC, se inscreveram para realizar o Exame Nacional do Ensino Médio - ENEM, em 2013², 7.173.574 sendo que destes, 218.004 de Mato Grosso do Sul, com abstenção média de 29,0%. Em 2014³ foram inscritos 8.721.946 candidatos e de Mato Grosso do Sul 162.865 com abstenção média de 27,0% e 8.356.215 inscritos em 2015⁴ e destes 132.003 referiam-se a inscritos de Mato Grosso do Sul com abstenção média de 30,0%. De modo que, para a obtenção dos resultados e realizar comparações, aplicamos as mesmas questões retiradas do arquivo Microdados, em Escolas Estaduais para alunos do terceiro ano do Ensino Médio, de cidades de Mato do Grosso do Sul (Campo Grande, Dourados e Bonito), para 206 (duzentos e seis) alunos. A maneira mais competente de se obter resultados confiáveis comparando avaliações para alunos diferentes é utilizando a Teoria de Resposta ao Item - TRI.

Vários modelos matemáticos são utilizados como método de avaliação, sendo esta um “calo” no sapato dos estudiosos sobre qual a melhor maneira de avaliar e distinguir entre os respondentes quem conseguiu absorver e quem não está habilitado para responder certas questões de um exame. Dentre as teorias mais utilizadas, a mais comum é a Teoria Clássica dos Testes - TCT, amplamente utilizado em concursos que os itens são de múltipla escolha. A nota é baseada no score bruto e portanto, não permite fazer a distinção se o candidato de maior score é realmente proficiente no conteúdo cobrado. Outro modelo matemático crescente como método avaliativo é a TRI, sendo muito utilizada nas avaliações aplicadas pelo governo brasileiro, como Prova Brasil, Exame Nacional para Certificação de Competências de Jovens e Adultos (Encceja) e ENEM. Essa teoria permite comparação entre indivíduos que fizeram avaliações semelhantes, pois define parâmetros como, grau de dificuldade, de discriminação e acerto ao acaso.

Com a utilização da amostra aplicada, com a amostra do Enem e as duas teorias, pretendemos verificar se o ensino de probabilidade e estatística está sendo desenvolvido a contento para que o aluno tenha habilidade para respondê-las, quando aplicadas no Enem. Analisar quais as questões com maiores índices de discriminação, observando o motivo do alto grau discriminativo e tentar apontar os motivos, quando houver proporções altas de

¹ Disponível em <<http://www.inep.gov.br>>

² Extraído de <http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=13236-indicadoresconsolidadosenem2013a-pdf&category_slug=maio-2013-pdf&Itemid=30192>

³ Extraído de <http://download.inep.gov.br/imprensa/2016/dados_de_abstencao_e_eliminacao_enem_2016.pdf>

⁴ Extraído de <http://download.inep.gov.br/educacao_basica/enem/downloads/2015/enem_2015_balanco_inscricoes.pdf>

acertos e, principalmente, de erros da questão.

Com esse foco, no primeiro capítulo faremos um breve estudo sobre a história da Educação Matemática e da Educação Estatística, revisando a literatura e trabalhos sobre o assunto e um breve histórico sobre o Enem, pois suas questões e teorias também são objetos de nossos estudos. No segundo capítulo, revisaremos a TCT e a TRI. No terceiro capítulo, explicitaremos o método utilizado para o estudo e análise pedagógica de sete itens com a elaboração das tabelas utilizando a TCT e tabela e gráfico da TRI, comentários sobre cada item e finalmente faremos comparação dos resultados e posterior discussão tentando responder as questões abaixo.

Os alunos que responderam as questões da amostra aplicada tem habilidade de respondê-las comparáveis com os alunos que responderam as mesmas questões na amostra ENEM?

Diante do resultado, O Ensino de Probabilidade e Estatística está sendo desenvolvido de forma a permitir que o aluno adquira habilidade suficiente para respondê-las?

Ao responder essas questões, tentaremos verificar o quão o ensino de probabilidade e estatística está sendo desenvolvido nas escolas públicas de Mato Grosso do Sul e, de acordo com as respostas, tentaremos fazer inferências para a melhor aplicação desse conteúdo no ensino básico.

1 Educação Matemática, Educação Estatística e ENEM

1.1 Educação e Educação Matemática

A Educação é um ato político. Ela que promove as transformações no indivíduo, considerando o contexto da atualidade e da nossa escola, imersa em uma sociedade complexa, em cujo “interior se multiplicam os grupos, e cada um deles vivencia e defende seus interesses e seus universos culturais” Lopes (2010, p. 5). Para D’Ambrosio (1996), a cultura nada mais é que um conjunto de mitos, valores, normas de comportamento e estilos que são compartilhados por pessoas que vivem num determinado tempo e espaço. Educar, e principalmente educar matemática, pressupõe alterar nosso conhecimento interno e fazer com que ele interaja com todo o conhecimento externo e bagagem trazida pelo aluno. A comunicação, portanto, facilita com que o docente permeie pela diversidade na qual envolve os estudantes. É preciso prepará-los para a realidade que o rodeia, pois segundo D’Ambrosio (1996, p. 3) “REALIDADE informa INDIVÍDUO que processa e executa uma AÇÃO que modifica a REALIDADE que informa INDIVÍDUO...”. Esse ciclo é considerado o ciclo vital de todo indivíduo. Ainda, a educação só pode ocorrer quando, num meio social, indivíduos podem interagir-se. Assim,

A educação é um processo que ocorre na interação entre as pessoas e a produção de conhecimento ocorre no coletivo. A educação é um processo essencialmente social, e os indivíduos que nela intervêm mantêm uma relação dialética continuada (LOPES, 2010, p. 1).

A Educação necessita ser pensada para que o docente tenha condições de produzir essa interação, promover discussão e fazer com que o conhecimento seja absorvido de modo que o indivíduo realmente produza as ações necessárias para modificar a si e o meio. É necessário esta interação para o conhecimento coletivo pois, segundo D’Ambrosio (1996, p. 25) “nenhum é igual ao outro na capacidade de captar e processar informações de uma mesma realidade”. Dessa forma,

A Educação Matemática e a Educação Estatística têm enorme importância na construção desse indivíduo capaz de modificar a realidade e executar ações, ao desenvolver a capacidade desse indivíduo analisar e tomar decisões assertivas.

Quanto à compreensão de alguns conceitos matemáticos, é importante que os estudantes entendam como eles se relacionam uns com os outros, e nesse sentido a Estatística pode ser muito útil e auxiliar os estudantes a ver a Matemática como um corpo estruturado de conhecimento, em lugar de fatos isolados (VENDRAMINI, 2006, p. 239).

Para ajudar nessa compreensão dos conceitos e na utilização da Estatística, a tecnologia vem de encontro com a necessidade de melhorar a visão dos aprendizes sobre a matemática. A disseminação das mídias sociais, a popularização da internet, dos canais de televisão pagos e com a facilitação do acesso às notícias por meios eletrônicos, é extremamente necessário o aprendizado de outras formas de leituras. Os meios de comunicação, impressos ou *on-line*, são ricos em elementos estatísticos (gráficos e tabelas, por exemplo). Os celulares, cada vez mais completos e acessíveis estão presentes no cotidiano dos alunos e professores. D'Ambrosio (1996) já previa que o professor que não se adequasse em utilizar a informática estaria fadado ao fracasso. Defende uma inovação no ensino da matemática, a qual refere-se a uma estratégia de estímulo ao desenvolvimento e coletivo gerada por diferentes grupos culturais que tem variado de acordo com a geografia e a história dos indivíduos.

O computador pode ser visto como um meio para ampliar três processos fundamentais de comportamento dos estudantes e dos educadores, que são: o processamento da informação, a interação e a comunicação Segundo (LOPES, 2010, p. 11).

Dessa forma D'Ambrosio (1996) discute que a sociedade moderna não será operacional com um instrumento intelectual obsoleto. Nos diz que para um bom aprendizado, necessitamos da matemática de hoje e ainda que a educação matemática e, acrescentamos, a educação estatística são fatores primordiais nesse aprendizado.

A educação para cidadania, que é um dos grandes objetivos da educação de hoje, exige uma “apreciação” do conhecimento moderno, impregnado de ciência e tecnologia. Assim, o papel do professor de matemática é particularmente importante para ajudar o aluno nessa apreciação, assim como para destacar alguns dos importantes princípios éticos a ela associados (D'AMBROSIO, 1996, p. 87).

Sendo assim, é impossível dissociar o ensino da política, pois está entre os papéis da escola formar cidadãos pensantes e capazes de discernir sobre o certo e o errado, o ético e o não ético, o moral e o imoral. A política está impregnada na educação e, nesse sentido, em Lopes (2010, p. 13) “o conhecimento matemático, estatístico e probabilístico é cada vez mais necessário diante das exigências da cidadania, do emprego e da continuidade de estudos”.

1.2 Educação Estatística e Probabilidade

A evolução natural do homem fez surgir a estatística. A curiosidade inerente ao ser humano de explicar o mundo e as coisas, desperta o interesse do homem em organizar e analisar dados.

Antigas civilizações realizavam estudos estatísticos com o objetivo de colher informações sobre agricultura, nascimentos e mortalidade. Dessa forma determinavam parâmetros para taxação e cobrança de impostos, assim como verificavam a quantidade de homens aptos para a guerra(SOUZA; LOPES, 2007, p. 1).

Em Crespo (1993), há um relato breve sobre a história da estatística,

A Estatística surgiu na Antiguidade, quando os vários povos sentiram a necessidade de registrar o número de habitantes, de nascimentos, de óbitos, para cobrar impostos e realizar inquéritos quantitativos por processos que, hoje, chamaríamos de estatísticas. A partir do século XVI, sentiu-se a necessidade de complementar as estatísticas populacionais com análises sistemáticas de fatos sociais, como nascimentos e óbitos, organizando-os em tabelas e calculando suas frequências relativas. No século XVIII, os estudos de tais fatos foram se tornando científicos, sendo batizados por Godofredo Achenwall com o nome de “Estatística”, determinando o seu objetivo e suas relações com as ciências. Surgiram as representações gráficas e o cálculo de probabilidades. A Estatística deixou de ser uma simples catalogação de dados numéricos coletivos para se tornar o estudo de como chegar a conclusões sobre o todo a partir de observações de partes desse todo. Suas aplicações e utilidade encontram-se em quase todos os campos da atividade humana (CRESPO, 1993 apud VENDRAMINI, 2006, p. 240).

As mídias modernas comumente mostram dados da economia, crescimento populacional, pesquisas eleitorais, crescimento ou decréscimo de empregos, saúde mundial, dentre outros, instigando assim a necessidade do aprendizado de novas linguagens para produzir inferências. A Base Nacional Comum Curricular - BNCC, prega que todos os cidadãos precisam desenvolver habilidades para coletar, organizar, representar, interpretar e analisar dados com a finalidade de emitir decisões, julgar e fazer inferências adequadas BNCC (2017).

A Base Nacional Comum Curricular para a Educação Básica - BNCC, também relata que:

Cabe ainda destacar que o desenvolvimento do pensamento numérico não se completa, evidentemente, apenas com objetos de estudos descritos na unidade Números. Esse pensamento é ampliado e aprofundado quando se discutem situações que envolvem conteúdos das demais unidades temáticas: Álgebra, Geometria, Grandezas e medidas e Probabilidade e Estatística (BNCC, 2017, p. 267).

Não é absurdo concluir que o ensino probabilístico e estatístico deva ser contemplado em todos os níveis do ensino, inclusive, na Educação Infantil. Dessa forma, a BNCC (2017, p. 272) - do Ensino Fundamental, anos iniciais, traz em suas diretrizes que, no que concerne as noções de probabilidade, tem como objetivos de “promover a compreensão de que nem todos os fenômenos são determinísticos e desenvolver a noção de aleatoriedade”. Já nos anos finais, deverão ser incluídas a probabilidade frequentista, de modo que façam experimentos

para confrontar os resultados obtidos com a probabilidade teórica. A coleta de dados, a organização de uma pesquisa provocam os alunos e ajudam a compreender o papel da estatística no cotidiano. Em BNCC (2017, p. 273) espera-se que “No Ensino Fundamental – anos finais, a expectativa é que os alunos saibam planejar e construir medidas de tendência central e construção de tabelas e diversos tipos de gráficos”.

A incerteza e o tratamento de dados são estudados na unidade temática probabilidade e estatística. Ela propõe a abordagem de conceitos, fatos e procedimentos presentes em muitas situações-problemas da vida cotidiana, das ciências e da tecnologia. Assim, todos os cidadãos precisam desenvolver habilidades para coletar, organizar, representar, interpretar e analisar dados em uma variedade de contextos, de maneira a fazer julgamentos bem fundamentados e tomar as decisões adequadas. Isso inclui raciocinar e utilizar conceitos, representações e índices estatísticos para descrever, explicar e prever fenômenos. (BNCC, 2017, p. 272)

Já no Ensino Médio, as Diretrizes Curriculares Nacionais de 2012, ainda em vigência, contempla o ensino da probabilidade apenas no primeiro bimestre do terceiro ano e o ensino de estatística no terceiro bimestre do mesmo ano. Isso demonstra que o conhecimento é apresentado de forma descontínua. Em estudo nas coleções de livros didáticos aprovados no Programa Nacional do Livro e do Material Didático - PNLD do triênio 2015-2017, observou que,

as coleções dedicam apenas um pouco mais do que três por cento de suas páginas para a abordagem do conteúdo Probabilidade durante o ciclo do Ensino Médio, ou seja, em uma conta simples, o aluno terá por ano em seu livro didático, apenas 9 páginas dedicadas ao tema Probabilidade (RODRIGUES; MARTINS, 2016, p. 6).

Essa descontinuidade prejudica na formação adequada do educando, pois saber ler, interpretar e analisar dados, ou seja, ter um letramento estatístico é necessário para poder explicar acontecimentos e proceder tomadas de decisões seguras. Nesse sentido é necessário reconhecer a importância da formação estatística como parte significativa de conceitos desenvolvidos no campo da educação matemática e, “compreender a educação estatística como uma necessidade formativa da sociedade atual, ao mesmo tempo, como um conjunto de saberes presentes no currículo da educação básica”, Cavalcante et al. (2016, p. 443).

O fato da estatística e a probabilidade serem assuntos amplamente cobrados no Exame Nacional do Ensino Médio - ENEM, corrobora a necessidade desses conteúdos no currículo da educação básica.

A Estatística é um dos assuntos cobrados no ENEM e outras avaliações, em razão de sua importância para compreensão da sociedade. Mas os alunos, por não terem o hábito de ler, mostram grande dificuldade em responder questões que envolvam textos, gráficos e tabelas (NOGUEIRA et al., 2011, s/p).

A formação acadêmica do cidadão passa pelo Exame Nacional do Ensino Médio - ENEM. A probabilidade e a estatística são conteúdos cobrados, pois estão definidos em duas das áreas de competência, a competência de área 6 e a competência de área 7. Dessa forma, é necessário que a probabilidade e a estatística devam ser apresentadas ao aluno de forma habitual e permanente, com a extrema necessidade de se unir a teoria à prática.

1.3 Exame Nacional do Ensino Médio - ENEM

O Exame Nacional do Ensino Médio - ENEM, é realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - Inep, autarquia federal vinculada ao Ministério da Educação - MEC e tem como missão “subsidiar a formulação de políticas educacionais dos diferentes níveis de governo com intuito de contribuir para o desenvolvimento econômico e social do país”, INEP (2015, s/p).

O ENEM tem como finalidade principal a avaliação do desempenho escolar e acadêmico ao fim do Ensino Médio, podendo seus resultados serem utilizados para: compor a avaliação da qualidade do Ensino Médio no País; subsidiar a implementação de políticas públicas; criar referência nacional para o aperfeiçoamento dos currículos do Ensino Médio; desenvolver estudos e indicadores sobre a educação brasileira; estabelecer critérios de acesso do participante a programas governamentais; selecionar candidatos para o ensino superior; e constituir parâmetros para a autoavaliação do participante, com vistas à continuidade de sua formação e à sua inserção no mercado de trabalho.

Desde sua primeira edição em 1998, até 2008, o ENEM era realizado anualmente, com a aplicação de uma única prova composta por sessenta e três questões interdisciplinares. Durante esse período, algumas instituições de ensino superior passaram a utilizá-lo como instrumento de seleção para o ingresso de seus estudantes.

A partir de 2009, ocorreu a reformulação metodológica do ENEM e de sua utilização como forma de seleção nos processos seletivos de muitas universidades federais. Reformularam-se as Matrizes de Referência do exame, que passaram a ser estruturadas por competências em quatro áreas do conhecimento – Linguagens, Códigos e suas tecnologias, Matemática e suas tecnologias, Ciências Humanas e suas tecnologias e Ciências da Natureza e suas tecnologias, apresentando ainda, eixos cognitivos comuns a todas as áreas.

A partir de então, o ENEM tornou-se uma das principais vias de acesso ao ensino superior público, democratizando as oportunidades e possibilitando a mobilidade acadêmica, além de continuar sendo referência para a autoavaliação dos estudantes. Passou a possibilitar, também, a certificação para conclusão do ensino médio, obedecendo às exigências previstas na Lei de Diretrizes e Bases da Educação Nacional (Lei 9.394/1996) para a Educação de Jovens e Adultos (INEP, 2015, s/p).

Atualmente, instrumento utilizado no exame é composto de uma redação e de

quatro provas com quarenta e cinco questões de múltipla escolha cada, que contempla as quatro áreas de conhecimento já citadas: As Matrizes de Referência de cada área avaliada são o instrumento norteador da construção dos itens (questões) das provas.

No que tange a Matriz Curricular Matemática e suas Tecnologias, as competências que se encontram na Matriz referem-se a conteúdos normalmente veiculados no ensino básico. Elas estão organizadas por temas matemáticos: números, geometria, álgebra, grandezas e medidas, modelagem, tratamento da informação e conhecimentos de estatística e probabilidade. Nesse estudo, nos atentaremos exclusivamente às competências de área 6, que é subdividida em três habilidades, e competência de área 7, que é subdividida em quatro habilidades.

A Competência de área 6 refere-se a: Interpretar informações de natureza científica e social obtidas da leitura de gráficos e tabelas, realizando previsão de tendência, extrapolação, interpolação e interpretação. Suas habilidades são divididas em:

H24 - Utilizar informações expressas em gráficos ou tabelas para fazer inferências.

H25 - Resolver problema com dados apresentados em tabelas ou gráficos.

H26 - Analisar informações expressas em gráficos ou tabelas como recurso para a construção de argumentos.

A Competência de área 7 refere-se a: Compreender o caráter aleatório e não-determinístico dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas, determinação de amostras e cálculos de probabilidade para interpretar informações de variáveis apresentadas em uma distribuição estatística. Suas habilidades são divididas em:

H27 - Calcular medidas de tendência central ou de dispersão de um conjunto de dados expressos em uma tabela de frequências de dados agrupados (não em classes) ou em gráficos.

H28 - Resolver situação-problema que envolva conhecimentos de estatística e probabilidade.

H29 - Utilizar conhecimentos de estatística e probabilidade como recurso para a construção de argumentação.

H30 - Avaliar propostas de intervenção na realidade utilizando conhecimentos de estatística e probabilidade (BRASIL, 2015, p. 6, 7).

Para testar essas habilidades, o ENEM mantém uma Matriz com questões de diferentes níveis que contemplam as competências e as habilidades citadas acima. Para melhor verificar o desempenho dos respondentes, utiliza-se a TRI, uma teoria inovadora, pois utiliza-se de parâmetros de verificação na qual prima avaliar a proficiência do aluno no assunto cobrado ao mesmo tempo em que verifica se o item está realmente aferindo o conhecimento do examinado.

2 Modelos Matemáticos

A avaliação sempre foi um dos pontos críticos do sistema educacional, pois os métodos utilizados largamente não permitem medir o quão verdadeiro é o conhecimento do indivíduo, de modo que, os testes estão sofrendo revisões e novos instrumentos são criados para resolver e ajudar em uma adequada avaliação, principalmente na área educacional. Os resultados dados pelo número ou pelo percentual de acertos em uma escala própria em geral não diz muito sobre o respondente e prejudica o julgamento sobre um determinado conteúdo. Segundo Klein (2013) costuma-se interpretar que as notas acima de setenta por cento são consideradas adequadas e as abaixo de cinquenta por cento como fracasso. Evidente que, quando as notas são abaixo do esperado, associa-se que, provavelmente a prova estava num nível acima dos alunos ou que os itens estavam em um grau de dificuldade demasiado. Esta percepção não leva em consideração o estado de espírito, a habilidade, a competência do aluno ou a discriminação dos itens.

Para lidar com esse problema, modelos matemáticos foram criados, entre eles a Teoria Clássica dos Testes - TCT e a Teoria de Resposta ao Item - TRI. A teoria TCT expressa escores brutos ou padronizados e tem sido utilizada nos processos de avaliação e seleção de indivíduos e, as análises estão sempre ligadas à prova como um todo, ficando inviável a comparação entre itens e indivíduos. Já a teoria TRI propõe modelos que analisa características do indivíduo que não podem ser medida diretamente. O método sugere formas de apresentar a probabilidade de um indivíduo dar uma certa resposta, levando em consideração suas habilidades na área de conhecimento.

A grande vantagem da TRI sobre a TCT é que essa teoria permite a comparação entre populações, desde que as provas submetidas tenham itens comuns e ainda, a comparação entre os resultados dos indivíduos submetidos à provas diferentes em épocas diferentes. Essa vantagem se dá porque esta teoria tem como elemento central o item e não a prova como um todo. Assim, a introdução da TRI trouxe muitas vantagens sobre o método tradicional, pois permite estimar e comparar os resultados dos alunos, mesmo que respondam itens diferentes, colocando-os em uma mesma escala.

O uso da TRI em avaliações educacionais teve início no Brasil com o Sistema de Avaliação da Educação Básica (Saeb), em 1995, e, posteriormente, foi implementado no Exame Nacional para Certificação de Competências de Jovens e Adultos (Encceja), na Prova Brasil e, por último, no Enem. No âmbito internacional, a TRI vem sendo utilizada largamente por diversos países: Estados Unidos, França, Holanda, Coreia do Sul, China e países participantes do Programa Internacional de Avaliação de Estudantes (Pisa). Um dos grandes exemplos de avaliação que utiliza a TRI é o exame de proficiência em língua inglesa (Toefl)(BRASIL, 2015, p.27).

A TRI é um instrumento poderoso que vem sendo introduzida em nosso meio, capaz de melhorar os processos avaliativos quantitativos, permitindo a construção de escalas e de habilidades calibradas.

Desde 2009, quando da mudança do Enem, o Inep vem se utilizando da TRI para avaliar os estudantes que realizam o exame. Para isso foi criado um banco de questões e uma escala específica. A nota não é calculada sobre a quantidade de acertos das questões, mas considera também a coerência das respostas do participante. O objetivo é de medir o conhecimento (proficiência) nas quatro áreas. A avaliação elaborada com quarenta e cinco questões para cada área, considerando os documentos oficiais e Matriz Curricular, portanto são calculadas quatro notas e a escala depende de dois valores, como a seguir.

valor de posição ou de referência, para o qual foi atribuído o valor 500, que representa o desempenho médio dos concluintes do ensino médio da rede pública de 2009 que realizaram o exame naquele ano;
valor de dispersão, para o qual foi atribuído o valor 100, que representa uma medida de variabilidade média das notas desses concluintes em relação ao desempenho médio 500. Esse valor é conhecido como desvio padrão (BRASIL, 2015, p. 8).

A partir disso, pode-se representar graficamente em uma régua, a proficiência de cada indivíduo e a característica de cada item, podendo assim realizar uma interpretação pedagógica dos resultados.

2.1 Teoria Clássica dos Testes - TCT

Segundo Klein (2013, p. 3) a Teoria Clássica dos Testes - TCT, orientou por muito tempo o desenvolvimento dos testes educacionais, psicológicos, avaliativos e etc. A TCT baseia-se no escore bruto de cada indivíduo, ou seja, o resultado é obtido apenas comparando a quantidade de itens respondidos corretamente por cada indivíduo, logo a habilidade é estimada pelo número de acertos, sendo X escore observável obtido por um estudante em um teste.

Ainda em Klein (2013, p. 3-5) a confiabilidade do teste dada por (*test reliability*), denominada R , calculada pela variância do escore observado $V(X)$ dividido pela soma entre a variância do escore verdadeiro $V(T)$ e a variância do erro do escore $V(E)$. O erro padrão de medida (*standard error measurement - SEM*) é dado pelo desvio padrão do escore observável obtido por um estudante $s(X)$ multiplicado pela raiz quadrada da subtração entre um e a confiabilidade R . Utiliza-se como estimativa usual o Alpha de Cronbach, onde todos os itens de múltiplas escolhas reduz-se a fórmula de Kuder-Richardson (KR20).

A confiabilidade depende dos alunos testados, do tamanho e dos itens do teste. Itens com coeficiente ponto bisserial, correlação entre acerto e número de acertos na prova, baixo ou negativo acrescentam pouco à confiabilidade.

No caso de um item em que todos erram ou todos acertam, $V(X)$ não se modifica e a variância do item é zero. Se o item for muito difícil ou fácil, $V(X)$ vai crescer “pouco”, alterando pouco o Alpha de Cronbach ou o KR20. No caso do coeficiente ponto-bisserial baixo ou negativo $V(X)$ cresce “pouco” também. (KLEIN, 2013, p. 5)

O coeficiente de correlação ponto-bisserial permite verificar o poder de discriminação de um item, comparando o escore total dos respondentes do teste com o escore total dos respondentes que acertaram um item em particular. O ponto-bisserial varia no intervalo $[-1, 1]$.

Em Tavares (2014, p. 22) o cálculo dessa correlação é dada por:

$$\rho_{pb} = \frac{S_p - S}{\sigma} \sqrt{\frac{p}{1-p}} \quad (2.1)$$

em que

S_p é a nota média no teste para os indivíduos que acertaram o item;

S é a nota média no teste para todos os indivíduos;

σ é o desvio-padrão das notas obtidas no teste por todos os indivíduos;

p é a proporção de acertos no item.

Valores negativos ou próximos de zero indicam que indivíduos com um bom desempenho total no teste estão dando respostas incorretas. Isso mostra que o item não está cumprindo o seu papel de distinguir alunos com uma grande habilidade dos alunos com uma baixa habilidade.

Pode-se calcular também a correlação para cada uma das opções dos itens, determinando uma medida de correlação entre o escore total no teste e o escore dos respondentes que marcaram cada uma das alternativas. Para isso, basta substituir equação 2.1 a nota média no teste para os indivíduos que acertaram o item pela nota média no teste dos indivíduos que marcaram o item.

Dessa forma, espera-se que a resposta correta tenha coeficiente ponto-bisserial positivo e os demais coeficientes negativos, caso contrário, isso pode ser a indicação de que há problema na construção das respostas ou na construção do item, ou seja, que o referido item não discrimina bem.

A Teoria Clássica dos Testes, apesar de amplamente utilizada nos diversos tipos de avaliação, têm limitações, dentre as quais foram listadas por Klein (2013)

- As estatísticas que descrevem os itens de teste dependem do grupo de estudantes que fazem o teste.

- Os escores de teste que descrevem o desempenho dos alunos dependem dos itens apresentados aos alunos.
- A Teoria Clássica dos Testes só pode ser utilizada em situações nas quais todos os alunos fazem o mesmo teste (ou formas “paralelas” de teste).
- A Teoria Clássica dos Testes não fornece um modelo do desempenho de um aluno em um item.
- A maioria das aplicações da Teoria Clássica dos Testes assume incorretamente que os erros de medida têm a mesma variabilidade para todos os alunos.

Para resolver tais limitações, foi apresentada a Teoria de Resposta ao Item, um modelo matemático que procura representar a probabilidade de um indivíduo dar uma certa resposta a um item, com parâmetros que permitem calcular esta probabilidade.

2.2 Teoria de Resposta ao Item - TRI

A Teoria de Resposta ao Item - TRI, expressa uma relação de tal forma que quanto maior a habilidade, maior a probabilidade de acerto no item. Segundo Andrade et al. (2000), vários modelos propostos na literatura dependem fundamentalmente de três fatores:

- (i) da maneira do item - dicotômicos ou não dicotômicos;
- (ii) do número de populações envolvidas - apenas uma ou mais de uma;
- (iii) e da quantidade de traços latentes que está sendo medida - apenas um ou mais de um.

2.2.1 Modelos dicotômicos ou dicotomizados

Esses modelos podem ser utilizados para a análise de itens de múltipla escolha e para itens de resposta aberta, quando avaliados de forma dicotomizada. Os modelos logísticos para itens dicotômicos são os modelos de resposta ao item mais utilizados, sendo que há três tipos que se diferenciam pelo número de parâmetros. Esses modelos são conhecidos como os modelos logísticos de um (ML1 ou modelo de Rasch), dois (ML2) e três (ML3) parâmetros que consideram, respectivamente a dificuldade do item, a dificuldade e a discriminação e ainda a dificuldade, a discriminação e a probabilidade de acerto ao acaso.

O ajuste da função utilizada na TRI é análogo ao processo utilizado no crescimento populacional que pode ser expresso por duas propostas Costa (2017, p. 23-29):

$$\begin{aligned}\frac{dN}{dt} &= kN \\ \frac{dN}{dt} &= kN \left(1 - \frac{L}{N}\right)\end{aligned}$$

onde N é o número de indivíduos em um tempo t , presentes no ambiente, k é uma constante de proporção e L é a quantidade de indivíduos que completaria a capacidade do ambiente. Para a TRI, a variável que representa o número de indivíduos na dinâmica populacional passa a ser a probabilidade de acerto de um item. Dessa forma, considerando uma pessoa com habilidade θ ; ao resolver um item, sua habilidade de sucesso será P e fracasso $1 - P$. Assim, a taxa de variação da probabilidade é proporcional à probabilidade de sucesso e fracasso. Observando que a Probabilidade de sucesso P é uma função que depende da habilidade θ , ou seja, $P = f(\theta)$, a Equação diferencial ordinária análoga fica

$$\frac{dP}{d\theta} = kP(1 - P) \Rightarrow \frac{dP}{kP(1 - P)} = d\theta \quad (2.2)$$

que, considerando k uma constante de proporção e resolvendo 2.2 utilizando o método de integração por partes (separação de variáveis), chega-se na equação que determina a probabilidade de sucesso em função da habilidade de um sujeito, ou seja

$$P(\theta) = \frac{1}{1 + \exp(-k\theta)} \quad (2.3)$$

Neste caso, o gráfico é uma curva em forma de “S” crescente. Podemos adotar o eixo das abscissas como a escala de proficiência denominada habilidade e o eixo das ordenadas como a probabilidade de acerto $P(\theta)$. Para simplificar o processo, adotamos $k = 1$.

Dessa forma, o modelo logístico com um parâmetro surge realizando a translação horizontal do gráfico da função 2.3. Para isso, acrescentamos o parâmetro denominado (b) ou parâmetro de habilidade ou proficiência. Daí, o modelo 2.3 modifica-se e transforma-se no modelo logístico de um parâmetro - ML1, podendo ser reescrita da seguinte forma:

$$P(\theta) = \frac{1}{1 + \exp(-(\theta - b))} \quad (2.4)$$

Agora, para definir o Modelo Logístico de dois parâmetros - ML2, é necessário um parâmetro capaz de distinguir um indivíduo de outro, com habilidades semelhantes. Esse parâmetro (a) pode ser denominado parâmetro de discriminação. Assim, retomando 2.2, podemos reescrevê-la:

$$\frac{dP}{d\theta} = kP(1 - P)a \quad (2.5)$$

Observe que o parâmetro de discriminação (a) é uma constante no processo de resolução das integrais, logo, o modelo ML2 é escrito

$$P(\theta) = \frac{1}{1 + \exp(-a(\theta - b))} \quad (2.6)$$

Esse parâmetro modifica a curva fazendo com que ela se suavize ou fique íngreme. Por fim, para chegarmos ao modelo logístico de três parâmetros - ML3, é necessário que um parâmetro chamado de acerto ao acaso (c) seja incluído em 2.6. Para introduzir esta propriedade precisa-se de uma transformação linear. Vejamos: A probabilidade mínima na escala do modelo de dois parâmetros é zero e no novo modelo ML3 é igual ao valor do parâmetro de acerto casual. No caso de itens com cinco opções esse acerto casual é 0,2 e, por outro lado, os valores máximos continuam sendo um. Para tanto, deve-se criar uma correspondência entre os modelos ML2 e ML3, no intervalo de 0 a 1, com um único valor de probabilidade no intervalo de c até 1, de forma proporcional. Seja x o valor da probabilidade na função 2.6 e y a nova probabilidade de ML3, a proporção entre as variações de probabilidade será

$$\frac{x - 0}{y - c} = \frac{1 - 0}{1 - c}$$

ou seja,

$$y = c + (1 - c)x$$

Dessa forma, o modelo de três parâmetros - ML3 define-se:

$$P(\theta) = c + (1 - c) \frac{1}{1 + \exp(-a(\theta - b))} \quad (2.7)$$

Esse parâmetro altera a assíntota inferior da curva característica em forma de “S”.

No Brasil e especialmente no Enem, o modelo amplamente utilizado é o modelo logístico unidimensional de três parâmetros - ML3, que definimos a seguir.

2.2.2 O modelo logístico de 3 parâmetros (ML3)

O modelo logístico unidimensional de três parâmetros - ML3, é o mais utilizado em exames como Sistema de Avaliação da Educação Básica – Saeb, Exame Nacional para Certificação de Competências de Jovens e Adultos - Encceja, Prova Brasil e Exame Nacional do Ensino Médio - Enem. Em Andrade et al. (2000) e de 2.7 o ML3 é apresentado pela TRI e definido por:

$$P(U_{ji} = 1 | \theta_j) = c_i + (1 - c_i) \cdot \frac{1}{1 + \exp(-Da_i(\theta_j - b_i))} \quad (2.8)$$

com $i = 1, 2, \dots, I$, e $j = 1, 2, \dots, n$, onde:

U_{ji} = é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente ao item i .

θ_j representa a habilidade (traço latente) do j – *ésimo* indivíduo.

$P(U_{ji} = 1 | \theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Função de Resposta do Item – FRI.

b_i é o parâmetro de dificuldade (ou de posição) do item i , medido na mesma escala da habilidade.

a_i é o parâmetro de discriminação (ou de inclinação) do item i , com valor proporcional à inclinação da Curva Característica do Item — CCI no ponto b_i .

c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i (muitas vezes referido como a probabilidade de acerto casual).

D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

Esse modelo apresenta um gráfico em forma de “S” denominada Curva Característica.

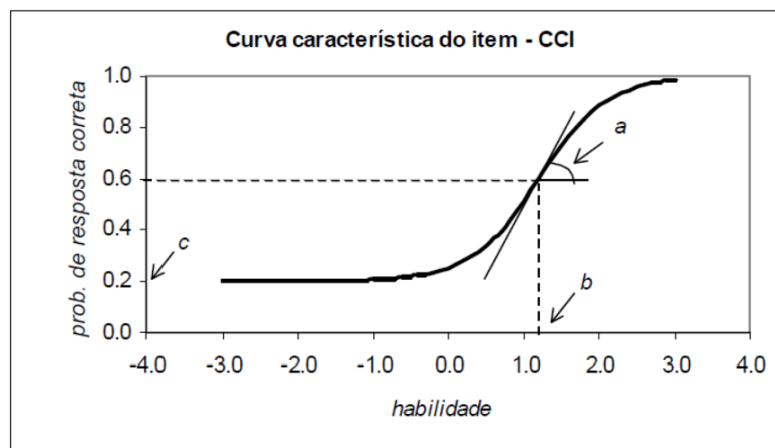
2.2.3 Curva Característica do Item - CCI

A seguir, na figura 1, podemos ver o gráfico da função probabilidade, a qual denomina-se Curva Característica do Item - CCI. Por tratar-se de uma função probabilidade, a imagem está sempre no intervalo $[0, 1]$. A figura apresenta a relação entre a probabilidade de responder corretamente o item (eixo da ordenada) e o grau de habilidade do indivíduo (eixo das abscissas). Pode-se perceber que o gráfico tem forma “S” (curva sigmoide) com inclinação e deslocamento definidos pelos parâmetros a e b , respectivamente.

O eixo das abscissas representa a intensidade do traço latente dos indivíduos (nível de gravidade de problemas com álcool, por exemplo) e o eixo das ordenadas é a probabilidade de o indivíduo responder positivamente ao item (variando de 0 a 1). A curva indica que quanto maior o nível de gravidade do indivíduo, maior sua probabilidade de responder afirmativamente ao item (SARTES; SOUZA-FORMIGONI, 2013, p. 245).

E ainda, o valor de a corresponde ao ângulo de inclinação da curva, com tendência a formar um ângulo de 90^0 e o valor de b a posição na escala de traço latente e finalmente, o local onde a assíntota inferior corta o eixo das ordenadas refere-se à probabilidade de acerto ao acaso Sartes e Souza-Formigoni (2013), (*parâmetro c*), ou seja, o “chute”.

Figura 1 – Modelo Curva Característica TRI



Fonte:Andrade et al. (2000)

2.2.4 O Parâmetro habilidade (θ)

A habilidade θ , também chamada de proficiência define o nível de aptidão de um indivíduo responder corretamente um certo conjunto de itens. É expresso na mesma escala do parâmetro de dificuldade (b), de modo que itens mais difíceis exigem habilidades mais complexas enquanto que itens mais fáceis exigem habilidades mais simples. É importante no cálculo da nota final do respondente, pois entende-se que a forma de aquisição do conhecimento é cumulativo, o que define que respondentes com habilidades maiores devem responder facilmente itens com dificuldades menores.

2.2.5 O Parâmetro de Discriminação (a)

O parâmetro de discriminação a mostra a aptidão do item e distingue indivíduos com proficiências diferentes; O Guia do Participante do Enem define como “o poder de discriminação que cada questão possui para diferenciar os participantes que dominam dos participantes que não dominam a habilidade avaliada naquela questão”BRASIL (2012, p. 13);

A discriminação é a característica capaz de provocar uma diferenciação entre indivíduos com habilidades diversas. De modo que, esse parâmetro é o que aumenta, ou não, a distinção entre as probabilidades de alunos com proficiências diferentes responderem com correção o item proposto, ou seja, deve ser entendido como habilidade do item em distinguir indivíduos quanto ao traço latente.

Apesar dos valores de a poderem variar entre o infinito negativo e o infinito positivo, espera-se que a discriminação seja sempre positiva e não muito próxima de zero, pois mostraria que a probabilidade de acerto do item seria a mesma para indivíduos com proficiências diferentes e isso caracteriza um baixo poder de discriminação.

Na curva característica, o valor de a corresponde a inclinação da curva. As curvas mais inclinadas mostram um maior poder de discriminação e as mais suaves o menor poder discriminativo.

“O parâmetro de discriminação não distingue se o indivíduo apresenta ou não o traço latente daquele nível, mas permite diferenciar pessoas que apresentam níveis semelhantes do traço latente medido por aquele determinado item. Assim quanto maior for o valor do parâmetro de discriminação, mais o item é capaz de detectar diferenças entre indivíduos” (SARTES; SOUZA-FORMIGONI, 2013, p. 245).

São necessários valores de comparação para decidir qual o grau de discriminação do item. Muitos autores apresentam uma tabela que serve como referência, amplamente utilizada. Andrade et al. (2010) apresenta a tabela 1 como referência.

Tabela 1 – Classificação de acordo com o parâmetro a

Valores do parâmetro “ a ”	Discriminação
0	Nenhuma
0,01 a 0,34	Muito Baixa
0,35 a 0,64	Baixa
0,65 a 1,34	Moderada
1,35 a 1,69	Alta
Maior que 1,70	Muito Alta

Fonte: Andrade et al. (2010)

Vejam os que o item com valores acima de 1,70 é um item altamente discriminativo e, portanto, não podem funcionar muito bem para o propósito inicialmente elaborado e o mesmo acontece com os itens que se aproximam de zero.

2.2.6 O Parâmetro de dificuldade (b)

O Parâmetro de dificuldade b mostra a habilidade mínima que um indivíduo precisa para ter uma grande probabilidade de acertar o item; tem como definição no Guia do Participante do Enem, como uma “associação à dificuldade da habilidade avaliada na questão, quanto maior seu valor, mais difícil é a questão. Ele é expresso na mesma escala da proficiência. Em uma prova de qualidade, devemos ter questões de diferentes níveis de dificuldade para avaliar adequadamente os participantes em todos os níveis de conhecimento” BRASIL (2012, p. 13);

O parâmetro b representa o grau de dificuldade do item produzir pelo respondente uma resposta correta (afirmativa). Ou seja, “é o nível de habilidade mínima necessária para um indivíduo ter uma probabilidade alta para acertar o referido item” Tavares (2014, p.12). Quanto maior for o valor de b maior deve ser o nível de dificuldade do item e maior deve ser a habilidade do indivíduo responder positivamente esse item. Essa dificuldade

está totalmente ligada a habilidade θ de resolver um item. A habilidade, alta ou não, pode ser calculada usando $\theta = \frac{1+c}{2}$. O valor de b indica a posição do item na escala do traço latente (proficiência) quando a probabilidade de um indivíduo responder corretamente o item for de 50%. Inicialmente os valores do parâmetro b podem variar entre o infinito negativo até o infinito positivo, mas usualmente utiliza-se a variação de -4 a $+4$.

O parâmetro b pode classificar os itens em muito fácil, fácil, mediano, difícil e muito difícil. Pode-se usar a escala padronizada na qual a média é igual a zero e o desvio padrão é igual a um $[0,1]$, porém pode-se alterar a escala, como no ENEM em que a escala utilizada é a qual a média é igual a 500 e o desvio padrão é igual a 100 $[500,100]$. Nessa última escala, apesar de b poder variar de $-\infty$ a $+\infty$ geralmente o utiliza-se os valores no intervalo $[200, 800]$

Em Tavares (2014) é apresentada a tabela que serve como parâmetro para a dificuldade do item utilizando a escala $[500,100]$, a qual modificamos utilizando a fórmula 2.9 de mudança de escala também apresentada por Tavares (2014, p. 14).

$$x = \frac{y - 500}{100} \quad (2.9)$$

onde x é a proficiência na escala $[0,1]$ e y a proficiência na escala $[500,100]$

Dessa forma chegamos a tabela 2, abaixo com a qual pode-se facilmente graduar a dificuldade dos itens.

Tabela 2 – Classificação de acordo com o parâmetro b

Valor do Parâmetro de Dificuldade (b)	Classificação
Abaixo de $-1,28$	Muito Fácil
De $-1,27$ a $-0,52$	Fácil
De $-0,51$ a $0,51$	Mediano
De $0,52$ a $1,27$	Difícil
Acima de $1,28$	Muito Difícil

Fonte: Tavares (2014) modificada pelo Autor

2.2.7 O Parâmetro de acerto ao acaso (c)

O Parâmetro de acerto ao acaso c mostra a probabilidade de um indivíduo com baixa habilidade responder corretamente um item; Sendo definido no Guia do Participante do Enem, como “em provas de múltipla escolha, um participante que não domina a habilidade avaliada em uma determinada questão da prova pode responder corretamente a esse item por acerto casual. Assim, esse parâmetro representa a probabilidade de um participante acertar a questão não dominando a habilidade exigida”BRASIL (2012, p. 13).

Em questões com cinco opções, a probabilidade de marcar cada uma das alternativas é $\frac{1}{p}$ e isso define que o parâmetro c ideal deve ser igual 0, 20. Quando é superior a esse valor,

significa, claramente, que a alternativa correta atrai respondentes com pouca habilidade e demonstra que a questão ou item tem problemas de elaboração. O parâmetro de acerto ao acaso c “representa a probabilidade de um aluno com baixa habilidade responder corretamente o item” Andrade et al. (2000, p. 11). O parâmetro c é igual a zero quando não é permitido “chutar” e então a probabilidade de acerto é 0,5.

Espera-se que participantes que acertaram as questões difíceis devam também acertar as questões fáceis, pois, entende-se que a aquisição do conhecimento ocorre de forma cumulativa, de modo que habilidades mais complexas requerem o domínio de habilidades mais simples (BRASIL, 2012, p. 17).

Portanto, na elaboração da nota do indivíduo, leva-se em consideração os acertos casuais, pois é razoável pensar que respondentes que acertam as questões com grau de dificuldade alto devam acertar as questões com grau de dificuldade baixo e não o contrário.

2.2.8 Função de Informação do Item

A figura 2 é chamada Curva de Informação do Item (CII), ela permite analisar quanto um item contém de informação para a medida de habilidade. Representa o montante de informação que contém um item em todos os pontos ao longo do traço latente que ele representa. É bastante utilizada juntamente com a Curva Característica do Item. Note que essa informação utiliza-se da informação Fisher que é a maneira de medir a quantidade de informação que uma variável aleatória observável X carrega sobre um parâmetro desconhecido θ do qual a probabilidade de X depende. A função é dada por:

$$I_i(\theta) = \frac{\left[\frac{d}{d(\theta)} P_i(\theta) \right]^2}{P_i(\theta) Q_i(\theta)} \quad (2.10)$$

onde,

$I_i(\theta)$ é a informação fornecida pelo item i na habilidade θ ;

$P_i(\theta) = P(X_{ij} = 1|\theta)$ e $Q_i(\theta) = 1 - P_i(\theta)$.

No caso de modelo ML3, podemos escrever uma equação que mostra a importância que têm os três parâmetros sobre o montante de informação do item. A equação é dada por:

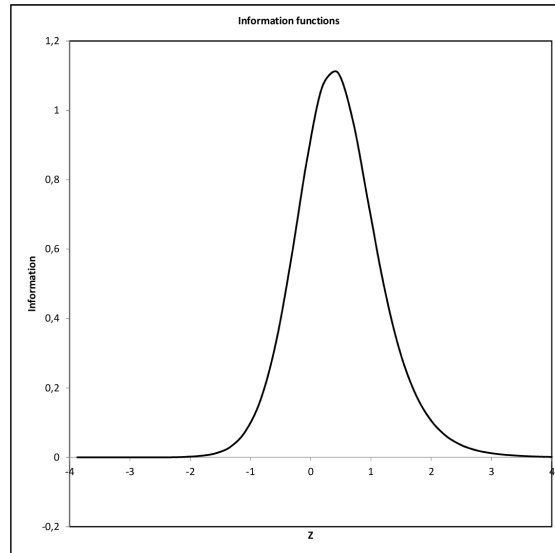
$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[\frac{P_i(\theta) - c_i}{1 - c_i} \right]^2 \quad (2.11)$$

A informação é maior:

- quando b_i se aproxima de θ ;

- quanto maior for a_i
- e quanto mais c_i se aproximar de 0.

Figura 2 – Modelo Gráfico Função Informação - TRI



Fonte: Autor

2.2.9 Função de Informação do Teste

A Informação do Teste é a soma das informações apresentadas pelos itens que compõe o teste. É dada por:

$$I(\theta) = \sum_{i=1}^I I_i(\theta) \quad (2.12)$$

O chamado erro padrão de estimação da TRI também é muito utilizado para representar esta função de informação. O erro padrão é dado por:

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (2.13)$$

Note que esses valores de informação dependem do valor de θ e que a cada item está associado um intervalo de habilidade, no qual cada item tem um poder de discriminação. Este intervalo é definido em torno do parâmetro b e se mostra nos gráficos pela curva de informação. De modo que, a discriminação entre os alunos de maior habilidade para os de menor habilidade é feita pelos itens considerados difíceis e não pelos itens fáceis.

As habilidades e os parâmetros são estimados a partir das respostas dos indivíduos. Estabelecida a escala de medida da habilidade, os valores dos parâmetros dos itens não sofrem variação quando submetidos a diferentes grupos de respondentes, desde que as habilidades de todos os grupos sejam medidas na mesma escala.

2.2.10 Unidimensionalidade

O modelo de três parâmetros pressupõe que os itens são homogêneos e que medem um único traço latente, isto é, supõe-se a unidimensionalidade do teste, ou seja, deve haver somente uma habilidade capaz de realizar todos os itens da prova.

Parece claro que qualquer desempenho humano é sempre multideterminado ou multimotivado, dado que mais de um traço latente entra na execução de qualquer tarefa (ANDRADE et al., 2000, p. 16).

Contudo, a unidimensionalidade admite que haja uma habilidade dominante responsável em realizar todos os itens. Este fator predominante é o que se supõe estar sendo medido no teste. Logo, existe um $\theta = \theta_j$, para algum j , que será dominante, ou seja,

$$\text{Proficiência} = f(\theta)$$

2.2.11 Independência local

Uma suposição fundamental para o processo de estimação dos parâmetros do modelo de três parâmetros é o axioma da Independência Local ou Independência Condicional na qual refere-se que para uma dada habilidade a resposta para diferentes itens da prova serão independentes, ou seja, as respostas dadas a dois itens diferentes são eventos independentes. Considerando que a unidimensionalidade implica independência local, itens devem ser elaborados de forma a atender a suposição de unidimensionalidade. Dessa forma, dada uma série de itens dos quais é elaborado um conjunto de respostas por um sujeito é igual ao produto das probabilidades de sucesso ou fracasso de cada item separadamente. Portanto, $P(u_{ji}|\theta_j)$ é a probabilidade do sujeito j , com habilidade θ_j dar respostas dicotômicas u_{ji} no item i . Logo, a probabilidade de um sujeito obter sucesso em cada item é independente do teste e é um evento independente, da seguinte forma:

$$P(u_{j1}, u_{j2}, u_{j3}, \dots, u_{jI}) = P_1(u_{j1}|\theta_j) \cdot P_2(u_{j2}|\theta_j) \cdot \dots \cdot P_I(u_{jI}|\theta_j),$$

onde N é o total de sujeitos, I é o total de itens, $i \in \{1, 2, \dots, I\}$ e $j \in \{1, 2, \dots, N\}$.

A resposta dada a cada item é dicotômica, ou seja, $u_{ji} = 1$ para resposta correta e $u_{ji} = 0$ para resposta incorreta, portanto a probabilidade de sucesso ou de fracasso é dada por $P_i(u_i = 1|\theta_j)$ e $P_i(u_i = 0|\theta_j)$, respectivamente e θ_j é a habilidade dominante do sujeito j dar a resolução do item i .

A implicação unidimensionalidade independência local pressupõe, então, que tem-se somente uma e não duas suposições a serem verificadas. Quando existe relação entre os itens, ela será apresentada pelo grupo de respondentes que possuem traços latentes próximos e respondem de forma semelhantes.

Segundo Costa (2017) quando um indivíduo responde a uma série de itens, ele produz um padrão de respostas, composto pelos produtos das probabilidades (P_i para acerto e $Q_i = 1 - P_i$ para erro).

2.2.12 Estimação

Uma das etapas mais importantes da TRI é a estimação de parâmetros dos itens e das habilidades dos respondentes Andrade et al. (2000). A probabilidade de uma resposta correta depende da habilidade do indivíduo e dos parâmetros que caracterizam o item, mas geralmente, ambos são desconhecidos. A estimação, no entanto, pode acontecer de três formas: Quando já conhecemos os parâmetros dos itens e então necessitamos estimar apenas a habilidade do indivíduo; Quando conhecemos a habilidade do indivíduo e necessitamos, dessa forma, estimar apenas os parâmetros dos itens e finalmente e mais comum, quando necessitamos estimar os parâmetros dos itens e a habilidade dos indivíduos.

A estimação, geralmente, é feita pelo Método da Máxima Verossimilhança, aplicando algum processo iterativo, como por exemplo o algoritmo de *Newton-Raphson* ou “*Scoring*” de Fisher. Frequentemente também são usados procedimentos Bayesianos. Na estimação, tanto dos parâmetros dos itens quanto da habilidade, há dois mais comuns: estimação conjunta, ou seja, parâmetros dos itens e habilidades ou divididos em duas etapas distintas, primeiro os parâmetros dos itens e depois as habilidades. O procedimento em duas etapas possibilita que a estimação dos itens seja feita pelo Método da Máxima Verossimilhança Marginal, ou seja, considerando uma determinada distribuição para a habilidade dos indivíduos de Π , cuja densidade de probabilidade (*fdp*) é $g(\theta|\eta)$, onde η é o conjunto de parâmetros associados à Π , e integrando a função de verossimilhança com relação a θ . Após a estimação dos parâmetros dos itens, as habilidades são estimadas individualmente por máxima verossimilhança ou pela moda ou média da distribuição condicional de θ_j dado $u_j = (u_{j1}, \dots, u_{jI})$, o vetor de respostas do indivíduo j , $j = 1, \dots, n$, com $\zeta_i = (a_i, b_i, c_i)$, o vetor de parâmetros do item i , $i = 1, \dots, I$, conhecidos. Devido a grande exigência do ponto de vista computacional para a estimação dos parâmetros foi proposto uma modificação¹ da matriz $3I \times 3I$ (no ML3) de derivadas segundas tornando-as bloco diagonal, o que possibilita que os parâmetros dos itens sejam estimados individualmente. Eles propõem também que a estimativas de máxima verossimilhança seja feita com a aplicação do algoritmo EM - Esperança (E) e Maximização (M)².

Neste caso, aplicaremos os procedimentos bayesianos e a estimativa de máxima verossimilhança será aplicado o algoritmo EM. Algumas notações serão necessárias para o desenvolvimento a seguir. Em particular θ_j a habilidade e U_{ji} a variável aleatória representa

¹ Modelo proposto por Bock & Aitkin (1981) referenciado em (ANDRADE et al., 2000)

² Introduzido por Dempster, Laird & Rubin (1977), referenciado em (ANDRADE et al., 2000)

a resposta do indivíduo j ao item i , tal que

$$U_{ji} = \begin{cases} 1, & \text{resposta correta,} \\ 0, & \text{resposta incorreta.} \end{cases}$$

Sejam $U_j = (U_{j1}, U_{j2}, \dots, U_{jI})$ o vetor aleatório de respostas do indivíduo j e $U_{..} = (U_{.1}, \dots, U_{.n})$ o conjunto integral de respostas. De forma similar, representaremos as observações por u_{ji}, u_j e $u_{..}$. Ainda, $\theta = (\theta_1, \dots, \theta_n)$ representará o vetor de habilidades dos n indivíduos e $\zeta = (\zeta_1, \dots, \zeta_I)$ o conjunto de parâmetros dos itens.

As duas principais suposições que usaremos em todo o restante deste texto, são as seguintes:

(S1) as respostas oriundas de indivíduos diferentes são independentes,

(S2) os itens são respondidos de forma independente por cada indivíduo (Independência Local), fixada sua habilidade.

A suposição (S2) necessita de uma discussão um pouco mais detalhada: ela garante que para cada valor de θ , se tomarmos um conjunto de indivíduos com habilidade θ , as covariâncias entre as respostas para cada par de itens serão nulas. Entretanto, se for considerado um conjunto de indivíduos com habilidades variadas, estas covariâncias não necessariamente serão nulas, elas serão positivas.

2.2.13 Estimação conjunta: parâmetros dos itens e habilidades

Este é o caso mais comum, em que nem os parâmetros dos itens e nem as habilidades são conhecidos. A principal diferença da estimação conjunta se dá no tratamento da métrica (escala) em que todos os parâmetros são estimados. Na estimação conjunta não há uma métrica definida e, portanto, deveremos estabelecê-la. Nesse modelo não há uma identificabilidade porque mais de um conjunto de parâmetros produz o mesmo valor no ML3, e conseqüentemente, na verossimilhança. Se $\theta_j^* = \alpha\theta_j + \beta$, $b_i^* = \alpha b_i + \beta$, $a_i^* = a_i/\alpha$ e $c_i^* = c_i$, onde α e β são constantes reais com $\alpha > 0$, então

$$\begin{aligned} P(U_{ji} = 1 | \theta_j^*, \zeta_i^*) &= \frac{c_i^* + (1 - c_i^*) \exp[-Da_i^*(\theta_j^* - b_i^*)]}{1 + \exp[-Da_i^*(\theta_j^* - b_i^*)]} \\ &= \frac{c_i + (1 - c_i) \exp[-D\frac{a_i}{\alpha}(\alpha\theta_j + \beta - (\alpha b_i + \beta))]}{1 + \exp[-D\frac{a_i}{\alpha}(\alpha\theta_j + \beta - (\alpha b_i + \beta))]} \\ &= \frac{c_i + (1 - c_i) \exp[-Da_i(\theta_j - b_i)]}{1 + \exp[-Da_i(\theta_j - b_i)]} \\ &= P(U_{ji} = 1 | \theta_j, \zeta_i). \end{aligned}$$

A não identificabilidade pode ser eliminada de várias formas, como fixando alguns valores para as habilidades, por exemplo. Isso pode ser resolvido especificando uma medida

de posição (média, por exemplo) e outra de dispersão (desvio-padrão, por exemplo) para as habilidades. Dessa forma, estabelece-se uma métrica (unidade de medida), para as habilidades e, por consequência, para os parâmetros dos itens. Neste caso, vamos eliminar o problema da não identificabilidade padronizando as habilidades de forma que estas tenham uma média específica μ e desvio-padrão σ . Assim, as habilidades e os parâmetros dos itens são estimados na métrica (μ, σ) . Adota-se $\mu = 0$ e $\sigma = 1$.

Para aplicação do algoritmo Newton-Raphson são necessárias as derivadas segunda da log-verossimilhança com relação a ζ_i e θ_j , $i = 1, \dots, I$ e $j = 1, \dots, n$. Estas derivadas compõem uma matriz \mathbb{H} quadrada de ordem $(3I + n)$ e essa dimensão pode ser suficientemente grande de forma a causar uma enorme exigência computacional. Por isso, é preciso explorar um pouco mais a estrutura de \mathbb{H} . Notemos que pela independência local, temos

$$\frac{\partial L(\zeta, \theta)}{\partial \zeta_i \partial \zeta'_l} = 0, \quad \text{para } i \neq l. \quad (2.14)$$

Pela independência entre as respostas de indivíduos diferentes, temos que

$$\frac{\partial L(\zeta, \theta)}{\partial \theta_j \partial \theta_l} = 0, \quad \text{para } j \neq l. \quad (2.15)$$

A suposição que simplifica bastante a estrutura de \mathbb{H} é de que não existe correlação entre itens e habilidades. Suposição que condiz com a realidade, pois as habilidades são inerentes dos indivíduos e independem dos itens envolvidos no estudo. A consequência desta suposição é

$$\frac{\partial L(\zeta, \theta)}{\partial \zeta_i \partial \theta_j} = 0, \quad \text{para } i = 1, \dots, I \text{ e } j = 1, \dots, n. \quad (2.16)$$

Assim, a matriz \mathbb{H} torna-se bloco-diagonal, na qual os I primeiros blocos são matrizes 3×3 relativas aos parâmetros dos itens e os n blocos seguintes são escalares relativos às habilidades.

2.2.14 Máxima Verossimilhança Marginal

O método de máxima verossimilhança marginal³, apresenta algumas vantagens em relação ao método da Máxima Verossimilhança Conjunta. A proposta desse método é fazer a estimação em duas etapas: primeiro dos itens e, posteriormente, as habilidades. Como as habilidades não são conhecidas, é preciso utilizar de algum artifício de forma que a verossimilhança não seja mais função das habilidades. Se considerarmos a população Π composta por n indivíduos com habilidades θ_j , $j = 1, \dots, n$, e construirmos a distribuição de frequência acumulada $G(\theta) = (\text{número de } j : \frac{\theta_j \leq \theta}{n})$, então, se n for suficientemente

³ Modelo proposto por *Bock & Lieberman (1970)* referenciado em (ANDRADE et al., 2000)

grande os θ_j estarão bastante próximos de forma que $G(\theta)$ pode ser aproximada por uma distribuição contínua. A densidade $g(\theta)$ relativa à $G(\theta)$, pode realmente ser considerada a função densidade para θ no experimento de retirar um indivíduo ao acaso da população Π e observar seu parâmetro θ .

Um artifício para eliminar as habilidades na verossimilhança consiste em marginalizar a verossimilhança integrando-a com relação à distribuição da habilidade. De forma geral, consideremos que as habilidades, θ_j , $j = 1, \dots, n$, são realizações de uma variável aleatória θ com distribuição contínua e função densidade de probabilidade (*fdp*) $g(\theta|\eta)$, duplamente diferenciável com as componentes de η conhecidas e finitas. Para o caso em que θ tem distribuição Normal, temos $\eta = (\mu, \sigma^2)$, onde μ é a média e σ^2 a variância das habilidades dos indivíduos de Π . Portanto, para que os itens sejam estimados na métrica (0,1), é preciso adotar $\mu = 0$ e $\sigma = 1$.

A abordagem de Bock & e Lieberman mostra que a probabilidade marginal de U_j é dada por

$$\begin{aligned} P(u_j|\zeta, \eta) &= \int_{\mathbb{R}} P(u_j|\theta, \zeta, \eta)g(\theta|\eta)d\theta \\ &= \int_{\mathbb{R}} P(u_j|\theta, \zeta)g(\theta|\eta)d\theta \end{aligned} \quad (2.17)$$

A última igualdade é usada já que a distribuição de U_j não é função de η e \mathbb{R} representa o conjunto dos números reais. Usando a independência entre as respostas de diferentes indivíduos, podemos escrever a probabilidade associada ao vetor de respostas $U_{..}$ como

$$P(u_{..}|\zeta, \eta) = \prod_{j=1}^n P(u_j|\zeta, \eta) \quad (2.18)$$

Embora a verossimilhança possa ser escrita como 2.18, tem sido frequente utilizar a abordagem de *Padrões de Resposta*. Como tem I itens no total, com duas possíveis respostas para cada item (0 ou 1), há $S = 2^I$ possíveis respostas (padrões de resposta). Nesse sentido, o índice j não mais representa um indivíduo mas sim um padrão de resposta. Seja r_j o número de ocorrências distintas do padrão de resposta j , e ainda $s \leq \min(n, S)$ o número de padrões de respostas com $r_j > 0$. Segue que

$$\sum_{j=1}^s r_j = n. \quad (2.19)$$

Pela independência entre as respostas dos diferentes indivíduos, temos que os dados seguem uma distribuição *Multinomial*, isto é:

$$L(\zeta, \eta) = \frac{n!}{\prod_{j=1}^s r_j!} \prod_{j=1}^s [P(u_j|\zeta, \eta)]^{r_j} \quad (2.20)$$

Então a log-verossimilhança é

$$\log L(\zeta, \eta) = \log \left\{ \frac{n!}{\prod_{j=1}^s r_j!} \right\} + \sum_{j=1}^s r_j \log P(u_j | \zeta, \eta). \quad (2.21)$$

As equações de estimação para os parâmetros dos itens são dadas por

$$\frac{\partial}{\partial \zeta_i} \log L(\zeta, \eta) = 0, \quad i = 1, \dots, I, \quad (2.22)$$

com

$$\begin{aligned} \frac{\partial}{\partial \zeta_i} \log L(\zeta, \eta) &= \frac{\partial}{\partial \zeta_i} \left\{ \sum_{j=1}^s r_j \log P(u_j | \zeta, \eta) \right\} \\ &= \sum_{j=1}^s r_j \frac{1}{P(u_j | \zeta, \eta)} \frac{\partial}{\partial \zeta_i} (P(u_j | \zeta, \eta)) \end{aligned} \quad (2.23)$$

Mas

$$\begin{aligned} \frac{\partial}{\partial \zeta_i} P(u_j | \zeta, \eta) &= \frac{\partial}{\partial \zeta_i} \int_{\mathbb{R}} P(u_j | \theta, \zeta) g(\theta | \eta) d\theta \\ &= \int_{\mathbb{R}} \left(\frac{\partial}{\partial \zeta_i} P(u_j | \theta, \zeta) \right) g(\theta | \eta) d\theta \\ &= \int_{\mathbb{R}} \left(\frac{\partial}{\partial \zeta_i} \prod_{l=1}^I P(u_{jl} | \theta, \zeta_l) \right) g(\theta | \eta) d\theta \end{aligned} \quad (2.24)$$

$$\begin{aligned} \frac{\partial}{\partial \zeta_i} P(u_j | \zeta, \eta) &= \int_{\mathbb{R}} \left(\prod_{l \neq i}^I P(u_{jl} | \theta, \zeta_l) \right) \left(\frac{\partial}{\partial \zeta_i} P(u_{ji} | \theta, \zeta_i) \right) g(\theta | \eta) d\theta \\ &= \int_{\mathbb{R}} \left(\frac{\frac{\partial}{\partial \zeta_i} P(u_{ji} | \theta, \zeta_i)}{P(u_{ji} | \zeta_i)} \right) P(u_j | \theta, \zeta) g(\theta | \eta) d\theta, \end{aligned} \quad (2.25)$$

onde a ordem da derivada e da integral em 2.24 pode ser permutada com base no Teorema de Convergência Dominada de Lebesgue. Reescrevendo $P(u_{ji} | \theta, \zeta_i)$,

$$\begin{aligned} \frac{\partial}{\partial \zeta_i} P(u_{ji} | \theta, \zeta_i) &= \frac{\partial}{\partial \zeta_i} (P_i^{u_{ji}} Q_i^{1-u_{ji}}) \\ &= u_{ji} P_i^{u_{ji}-1} \left(\frac{\partial P_i}{\partial \zeta_i} \right) Q_i^{1-u_{ji}} + (1 + u_{ji}) Q_i^{-u_{ji}} \left(-\frac{\partial}{\partial \zeta_i} P_i \right) P_i^{u_{ji}} \\ &= (u_{ji} P_i^{u_{ji}-1} Q_i^{1-u_{ji}} - (1 - u_{ji}) Q_i^{-u_{ji}} P_i^{u_{ji}}) \frac{\partial P_i}{\partial \zeta_i}. \end{aligned}$$

Note que o termo entre parênteses vale [1] quando $u_{ji} = 1$ e vale [-1] quando $u_{ji} = 0$. Podemos reescrevê-lo como $(-1)^{u_{ji}+1}$.

$$\frac{\partial}{\partial \zeta_i} P(u_{ji} | \theta, \zeta_i) = (-1)^{u_{ji}+1} \left(\frac{\partial P_i}{\partial \zeta_i} \right). \quad (2.26)$$

Notando agora que

$$\frac{(-1)^{u_{ji}+1}P_iQ_i}{P_i^{u_{ji}}Q_i^{1-u_{ji}}} = \begin{cases} Q_i, & \text{se } u_{ji} = 1 \\ -P_i, & \text{se } u_{ji} = 0 \end{cases} \quad (2.27)$$

podemos reescrever este termo $u_{ji} - P_i$. Segue que 2.25 pode ser escrita dessa forma:

$$\frac{\partial}{\partial \zeta_i} P(u_{j.}|\zeta, \eta) = \int_{\mathbb{R}} \left[\frac{(u_{ji} - P_i)}{P_iQ_i} \left(\frac{\partial P_i}{\partial \zeta_i} \right) \right] P(u_{j.}|\theta, \zeta)g(\theta|\eta)d\theta \quad (2.28)$$

Por conveniência, consideremos a seguinte ponderação:

$$W_i = \frac{P_i^*Q_i^*}{P_iQ_i}, \quad (2.29)$$

onde

$$P_i^* = \frac{1}{1 + \exp(-Da_i(\theta - b_i))} \quad \text{e} \quad Q_i^* = 1 - P_i^*$$

Com isso, podemos reescrever a equação 2.28 como

$$\frac{\partial}{\partial \zeta_i} P(u_{j.}|\zeta, \eta) = \int_{\mathbb{R}} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial \zeta_i} \right) \frac{W_i}{P_i^*Q_i^*} \right] P(u_{j.}|\theta, \zeta)g(\theta|\eta)d\theta \quad (2.30)$$

Usando a notação

$$g_i^*(\theta) \equiv g(\theta|u_{j.}, \zeta, \eta) = \frac{P(u_{j.}|\theta, \zeta)g(\theta|\eta)}{P(u_{j.}|\zeta, \eta)} \quad (2.31)$$

Temos que a função verossimilhança (2.23) pode ser escrita como

$$\frac{\partial}{\partial \zeta_i} \log L(\zeta, \eta) = \sum_{j=1}^s r_j \int_{\mathbb{R}} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial \zeta_i} \right) \frac{W_i}{P_i^*Q_i^*} \right] g_j^*(\theta)d\theta \quad (2.32)$$

Definida a verossimilhança, resta a obtenção das equações específicas para cada parâmetro do vetor $\zeta_i = (a_i, b_i, c_i)'$.

Para a estimação do parâmetro de discriminação a_i , de (2.32), temos:

$$\begin{aligned} \frac{\partial}{\partial a_i} \log L(\zeta, \eta) &= \sum_{j=1}^s r_j \int_{\mathbb{R}} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial a_i} \right) \frac{W_i}{P_i^*Q_i^*} \right] g_j^*(\theta)d\theta \\ &= \sum_{j=1}^s r_j \int_{\mathbb{R}} \left[(u_{ji} - P_i) D(1 - c_i)(\theta - b_i) P_i^* Q_i^* \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta)d\theta \\ &= D(1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}} [u_{ji} - P_i](\theta - b_i) W_i] g_j^*(\theta)d\theta. \end{aligned} \quad (2.33)$$

Para a estimação do parâmetro de dificuldade b_i , de 2.32

$$\begin{aligned} \frac{\partial}{\partial b_i} \log L(\zeta, \eta) &= \sum_{j=1}^s r_j \int_{\mathbb{R}} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial b_i} \right) \frac{W_i}{P_i^*Q_i^*} \right] g_j^*(\theta)d\theta \\ &= \sum_{j=1}^s r_j \int_{\mathbb{R}} \left[(u_{ji} - P_i)(-1)Da_i(1 - c_i)P_i^*Q_i^* \frac{W_i}{P_i^*Q_i^*} \right] g_j^*(\theta)d\theta \\ &= -Da_i(1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}} [u_{ji} - P_i] W_i] g_j^*(\theta)d\theta. \end{aligned} \quad (2.34)$$

Para a estimação do parâmetro de acerto ao acaso c_i , de 2.32

$$\begin{aligned} \frac{\partial}{\partial c_i} \log L(\zeta, \eta) &= \sum_{j=1}^s r_j \int_{\mathbb{R}} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial c_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta) d\theta \\ &= \sum_{j=1}^s r_j \int_{\mathbb{R}} \left[(u_{ji} - P_i) Q_i^* \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta) d\theta \\ &= \sum_{j=1}^s r_j \int_{\mathbb{R}} \left[(u_{ji} - P_i) \frac{W_i}{P_i^*} \right] g_j^*(\theta) d\theta. \end{aligned} \quad (2.35)$$

Dessa forma, as equações de estimação para os parâmetros a_i , b_i e c_i são, respectivamente,

$$a_i : D(1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}} [u_{ji} - P_i](\theta - b_i) W_i] g_j^*(\theta) d\theta \quad (2.36)$$

$$b_i : -D a_i (1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}} [u_{ji} - P_i] W_i] g_j^*(\theta) d\theta \quad (2.37)$$

$$c_i : \sum_{j=1}^s r_j \int_{\mathbb{R}} \left[(u_{ji} - P_i) \frac{W_i}{P_i^*} \right] g_j^*(\theta) d\theta \quad (2.38)$$

As quais não possuem solução explícita.

Métodos de quadratura

Em Andrade et al. (2000), diz que podemos notar que as equações 2.36, 2.37 e 2.38 envolvem integrais que não apresentam solução analítica. Dessa forma, algum meio deve ser encontrado para a solução (aproximação) numérica de uma integral. Na TRI têm sido frequente a aplicação do método *Hermite-Gauss*, denominado *método de quadratura gaussiana*. Se $g(\theta|\eta)$ é uma função contínua com integral finita, ela pode ser aproximada, para qualquer grau de precisão, por uma outra função que assume um número finito de pontos. O problema de obter a integral de uma função contínua é substituído pela obtenção da soma das área de um número finito, digamos q , de retângulos. Considerando que os pontos médios de cada retângulo são $\bar{\theta}_k$, $K = 1, \dots, q$, são denominados nós e que cada nó tem um peso $A_k \equiv A(\bar{\theta}_k)$ que leva em conta a altura $g(\bar{\theta}_k|\eta)$ e a largura Δ_k do respectivo intervalo, tal como $A_k = g(\bar{\theta}_k|\eta) \times \Delta_k$. Os valores $\bar{\theta}_k$ e Δ_k são obtidos resolvendo-se um conjunto de equação que envolvem a função $g(\bar{\theta}_k|\eta)$ e o número de nós. Segue que a equação 2.31 pode ser reescrita em forma de quadratura, como

$$g_j^*(\bar{\theta}_k) \simeq \frac{P(u_j|\bar{\theta}_k, \zeta) A_k}{\sum_{k=1}^q P(u_j|\bar{\theta}_k, \zeta) A_k} \cdot \frac{1}{\Delta_k} \quad (2.39)$$

Para que a expressão em forma de quadratura fique o mais parecida possível com a original, podemos redefinir a quantidade $g_j^*(\bar{\theta}_k)$ de 2.39 por

$$g_j^*(\theta_k) = \frac{P(u_j|\bar{\theta}_k, \zeta)A_k}{\sum_{k=1}^q P(u_j|\bar{\theta}_k, \zeta)A_k} \quad (2.40)$$

Desta forma, a função verossimilhança para a_i , b_i e c_i fica

$$a_i : D(1 - c_i) \sum_{j=1}^s \sum_{k=1}^q r_j [(u_{ji} - P_{ki})(\bar{\theta}_k - b_i)W_{ki}] g_j^*(\bar{\theta}_k). \quad (2.41)$$

$$b_i : -Da_i(1 - c_i) \sum_{j=1}^s \sum_{k=1}^q r_j [(u_{ji} - P_{ki})W_{ki}] g_j^*(\bar{\theta}_k) = 0. \quad (2.42)$$

$$c_i : \sum_{j=1}^s \sum_{k=1}^q r_j \left[(u_{ji} - P_{ki}) \frac{W_{ki}}{P_{ki}^*} \right] g_j^*(\bar{\theta}_k) = 0. \quad (2.43)$$

Ressalta-se que a função $g_j^*(\theta_k)$ nas equações 2.41, 2.42 e 2.43 devem ser calculadas por 2.40. Novamente as equações não apresentam soluções explícitas para a estimação de máxima verossimilhança dos parâmetros dos itens. Para a aplicação dos procedimentos de Newton-Raphson ou Scoring de Fisher devemos notar que as derivadas segundas de $\log L(\zeta, \eta)$ com relação a ζ_i e ζ_l , para $i \neq l$, não são nulas, o que leva à necessidade de estimação dos parâmetros dos I itens simultaneamente. Contorna-se este problema adotando uma proposta⁴ que tem por base a suposição de que os itens são independentes, de forma que

$$\frac{\partial^2 \log L(\zeta, \eta)}{\partial \zeta_i \partial \zeta_l} = 0, \quad \text{para } i \neq l \quad (2.44)$$

Essa suposição modifica a matriz $\mathbb{H}_{\Pi}(\zeta)$ tornando-a bloco diagonal, de forma que a independência a possibilita que os itens sejam estimados individualmente. Fazendo as adaptações devidas a 2.44, as obtenções das estimativas de máxima verossimilhança deve ser feita através do algoritmo EM, e por isso algumas alterações nas expressões 2.41 a 2.43 são necessárias.

De 2.36 temos

$$\begin{aligned} \frac{\partial}{\partial a_i} \log L(\zeta, \eta) &= D(1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}} [(u_{ji} - P_i)(\theta - b_i)W_i] g_j^*(\theta) d\theta \\ &= D(1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}} (\theta - b_i) [(u_{ji} g_j^*(\theta) - P_i g_j^*(\theta))] W_i d\theta \end{aligned}$$

⁴ Proposto por *Bock & Aitkin (1981)* referenciado em *(ANDRADE et al., 2000)*

$$\begin{aligned}
\frac{\partial}{\partial a_i} \log L(\zeta, \eta) &= D(1 - c_i) \int_{\mathbb{R}} (\theta - b_i) \left[\sum_{j=1}^s r_j u_{ji} g_j^*(\theta) - P_i \sum_{j=1}^s r_j g_j^*(\theta) \right] W_i d\theta \\
&= D(1 - c_i) \int_{\mathbb{R}} (\theta - b_i) [r_i(\theta) - P_i f_i(\theta)] W_i d\theta
\end{aligned} \tag{2.45}$$

onde

$$r_i(\theta) = \sum_{j=1}^s r_j u_{ji} g_j^*(\theta), \quad f_i(\theta) = \sum_{j=1}^s r_j g_j^*(\theta)$$

De forma análoga, de 2.37 e 2.38 temos

$$\frac{\partial}{\partial b_i} \log L(\zeta, \eta) = -D a_i (1 - c_i) \int_{\mathbb{R}} [r_j(\theta) - P_i f_i(\theta)] W_i d\theta \tag{2.46}$$

$$\frac{\partial}{\partial c_i} \log L(\zeta, \eta) = \int_{\mathbb{R}} [r_j(\theta) - P_i f_i(\theta)] W_i d\theta \tag{2.47}$$

Sendo que $g_j^*(\theta)$ é a distribuição condicional de θ_j dado $u_{j.}$, então $f_i(\theta)$ representa o número esperado de indivíduos, dentre os que responderam o item i em uma população de tamanho n , que têm habilidade θ . Para a quantidade $r_i(\theta)$ contribuem apenas os indivíduos que responderam corretamente ao item i . Logo, esta quantidade representa o número esperado de indivíduos dentre os que responderam corretamente ao item i em uma população de tamanho n , que têm habilidade θ .

2.2.15 Estimação bayesiana

Basicamente, a estimação bayesiana consiste em estabelecer distribuições *a priori* para os parâmetros interessados.

Em Andrade et al. (2000) temos que considerando que a distribuição da habilidade é função de um vetor de parâmetros η , com densidade $g(\theta|\eta)$, e que a distribuição de $\zeta_i, i = 1, \dots, I$, é função de um vetor de parâmetros τ , com densidade $g(\zeta|\tau)$. Podemos, ainda, estabelecer distribuições *a priori* para os parâmetros τ e η , digamos $f(\tau)$ e $g(\eta)$. Podemos adotar uma distribuição degenerada para η em $(0,1)$ ou uma distribuição que tenha vetor de médias $(0,1)$ e variâncias muito pequenas.

Com isso a densidade conjunta desses parâmetros é

$$f(\theta, \zeta, \eta, \tau) = f(\zeta|\tau)g(\theta|\eta)f(\tau)g(\eta) = \left\{ \prod_{i=1}^I f(\zeta_i|\tau) \right\} \left\{ \prod_{j=1}^n g(\theta_j|\eta) \right\} f(\tau)g(\eta).$$

Para fazer inferências sobre todos esses parâmetros, deve-se basear na distribuição *a posteriori*:

$$f(\theta, \zeta, \eta, \tau|u..) \propto L(u..; \theta, \zeta) f(\zeta|\tau) g(\theta|\eta) f(\tau) g(\eta). \tag{2.48}$$

Como, geralmente, interessa-se um número reduzido de parâmetros, trabalha-se com uma posteriori que seja função apenas dos parâmetros de interesse.

2.2.16 Estimação dos parâmetros

Ainda tomando como referência Andrade et al. (2000), para fazer inferências com relação aos parâmetros dos itens, é conveniente “marginalizar” a posteriori integrando com relação a θ e τ , obtendo a distribuição *a posteriori* de ζ e η :

$$\begin{aligned} f^*(\zeta, \eta|u_{..}) &= C \iint P(u_{..}; \theta, \zeta) f(\zeta|\tau) g(\theta|\eta) f(\tau) g(\eta) d\theta d\tau \\ &= C g(\eta) \left\{ \int f(\zeta|\tau) f(\tau) d\tau \right\} \left\{ \int P(u_{..}; \theta, \zeta) g(\theta|\eta) d\theta \right\} \\ &\propto L(\zeta, \eta) f(\zeta) g(\eta), \end{aligned} \quad (2.49)$$

onde C representa uma constante, $L(\zeta, \eta) \equiv P(u_{..}; \zeta, \eta)$ e

$$f(\zeta) = \int f(\zeta|\tau) f(\tau) d\tau.$$

Como estimador de ζ pode-se escolher alguma característica de $f^*(\zeta, \eta|u_{..})$, sendo que as mais adotadas são a média e a moda.

Considerando a *moda da posteriori* como o estimador de ζ , ou seja, o valor de ζ que maximiza a posteriori marginal. Temos que:

$$\log f^*(\zeta, \eta|u_{..}) = C + \log L(\zeta, \eta) + \log f(\zeta) + \log g(\eta)$$

onde o primeiro termo representa uma constante. Pela suposição de independência entre os itens a estimação será feita um item por vez. Observando que a última parcela não é função de ζ_i , temos que as equações de estimação para os parâmetros dos itens $\zeta_i, i = 1, \dots, I$, são dadas por

$$\frac{\partial}{\partial \zeta_i} \log f^*(\zeta, \eta|u_{..}) = \frac{\partial}{\partial \zeta_i} \log L(\zeta, \eta) + \frac{\partial}{\partial \zeta_i} \log f(\zeta) = 0 \quad (2.50)$$

Como o parâmetro a_i deve ser positivo, b_i pode assumir qualquer valor real e c_i deve estar no intervalo $[0,1]$, devemos assumir distribuições que levam em conta essas limitações e isso exige um tratamento diferenciado para cada um dos parâmetros.

Distribuição a priori para a_i

Adota-se, em Andrade et al. (2000), as distribuições “Log-normal” ou “Chi-Quadrado” para a_i . Vamos supor que cada parâmetro a_i tem distribuição Log-normal com parâmetro $\tau = (\mu_a, \sigma_a^2)$. A transformação $\alpha_i = \log a_i$ resulta em cada α_i tendo uma distribuição Normal $(\mu_\alpha, \sigma_\alpha^2)$, onde $\mu_\alpha = e^{\frac{\mu_a + \sigma_a^2}{2}}$ e $\sigma_\alpha^2 = (e^{\sigma_a^2} - 1) e^{2\mu_a + \sigma_a^2}$.

Como a distribuição de a_i é log-normal, sua densidade é

$$f(a_i|\mu_a, \sigma_a^2) = \frac{1}{\sqrt{2\pi a_i \sigma_a}} \exp \left[-\frac{1}{2\sigma_a^2} (\log a_i - \mu_a)^2 \right]$$

Assim, a segunda parcela de 2.50 pode ser reescrita como

$$\frac{\partial}{\partial a_i} \log f(a_i|\mu_a, \sigma_a^2) = -\frac{1}{a_i} \left[1 + \frac{\log a_i - \mu_a}{\sigma_a^2} \right] \quad (2.51)$$

Distribuição a priori b_i

Como os parâmetros de dificuldade estão na mesma escala da habilidade, Andrade et al. (2000) supõe que cada b'_i s tem distribuição *Normal* com vetor de parâmetros $\tau = (\mu_b, \sigma_b^2)$.

Como a distribuição de b_i é normal, sua densidade é

$$f(b_i|\mu_b, \sigma_b^2) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp \left[-\frac{1}{2} \left(\frac{b_i - \mu_b}{\sigma_b} \right)^2 \right]$$

Assim, a segunda parcela de 2.50 pode ser reescrita:

$$\frac{\partial}{\partial b_i} \log f(b_i|\mu_b, \sigma_b^2) = -\frac{(b_i - \mu_b)}{\sigma_b^2} \quad (2.52)$$

Distribuição a priori c_i

Segundo consta em Andrade et al. (2000) uma priori *Beta* foi proposta por “Swaminathan & Gifford (1986)”, pois c_i somente pode pertencer ao intervalo $[0,1]$. Foi proposta a função Beta com parâmetros $s + 1$ e $t + 1$, dada por:

$$f(c_i|s, t) = \frac{\Gamma(s + t + 2)}{\Gamma(s + 1)\Gamma(t + 1)} c_i^s (1 - c_i)^t, \quad (2.53)$$

onde $\Gamma(d)$ é a função Gama, definida por

$$\Gamma(d) = \int_0^\infty x^{d-1} \exp(-x) dx$$

A média desta distribuição é dada por

$$p = \frac{s + 1}{s + t + 2}.$$

É proposto ainda a seguinte parametrização⁵

$$\alpha = mp + 1 \text{ e } \beta = \beta m(1 - p) + 1,$$

⁵ Proposto por “Swaminathan & Gifford (1986)” referenciado em (ANDRADE et al., 2000)

onde $m = s + t + 2$. Desta forma, $p = \frac{(s+1)}{m}$ e, conseqüentemente, $s = mp - 1$ e $t = m - s - 2 = m(1 - p) - 1$. Segue então que

$$s = \alpha - 2 \text{ e } t = \beta - 2$$

Daí, em 2.53, obtemos

$$f(c_i|\alpha, \beta) = \frac{\Gamma(\alpha + \beta - 2)}{\Gamma(\alpha - 1)\Gamma(\beta - 1)} c_i^{\alpha-2} (1 - c_i)^{\beta-2}. \quad (2.54)$$

Neste caso, a média p passa a ser interpretada como a probabilidade de acerto por indivíduos com baixa habilidade. Os parâmetros α e β são definidos para que p tenha o valor desejado. Entretanto, “Swaminathan & Gifford” sugerem que a escolha de m deva se situar no intervalo $15, \dots, 20$, o que leva a uma certa restrição na escolha de α e β (ANDRADE et al., 2000, p. 71).

Para chegar a expressão para a segunda parcela de 2.50, observamos que

$$\log f(c_i|\alpha, \beta) = C + (\alpha - 2) \log c_i + (\beta - 2) \log(1 - c_i). \quad (2.55)$$

que por conseqüência

$$\frac{\partial}{\partial c_i} \log f(c_i|\alpha, \beta) = \frac{\alpha - 2}{c_i} - \frac{\beta - 2}{1 - c_i} \quad (2.56)$$

Em resumo, com as equações 2.51, 2.52 e 2.56, combinadas com as equações 2.45, 2.46 e 2.47, temos as equações de estimação para as componentes de ζ_i . são elas:

$$a_i : D(1 - c_i) \int_{\mathbb{R}} (\theta - b_i) [r_i(\theta) - P_i f_i(\theta)] W_i d\theta - \frac{1}{a_i} \left[1 + \frac{\log a_i - \mu_a}{\sigma_a^2} \right], \quad (2.57)$$

$$b_i : -D(1 - c_i) \int_{\mathbb{R}} [r_i(\theta) - P_i f_i(\theta)] W_i d\theta - \frac{(b_i - \mu_b)}{\sigma_b^2}, \quad (2.58)$$

$$c_i : \int_{\mathbb{R}} [r_i(\theta) - P_i f_i(\theta)] \frac{W_i}{P_i^*} d\theta + \frac{(\alpha - 2)}{c_i} - \frac{\beta - 2}{1 - c_i}. \quad (2.59)$$

Para efeito de aplicação dos procedimentos iterativos Newton-Raphson ou “Scoring” de Fischer, precisa-se das derivadas segundas das expressões acima. Logo temos,

$$\begin{aligned} \frac{\partial^2 \log f(a_i|\mu_a, \sigma_a^2)}{\partial a_i^2} &= \frac{1}{a_i \sigma_a^2} [\sigma_a^2 + \log a_i - \mu_a - 1], \\ \frac{\partial^2 \log f(b_i|\mu_b, \sigma_b^2)}{\partial b_i^2} &= -\frac{1}{\sigma_b^2}, \\ \frac{\partial^2 \log f(c_i|\alpha, \beta)}{\partial c_i^2} &= -\frac{\alpha - 2}{c_i^2} - \frac{\beta - 2}{(1 - c_i)^2}. \end{aligned} \quad (2.60)$$

Assim, a função probabilidade a posteriori 2.48

$$f(\theta, \zeta, \eta, \tau|u..) \propto L(u..; \theta, \zeta) f(\zeta|\tau) g(\theta|\eta) f(\tau) g(\eta)$$

Podemos reescrever da seguinte forma:

$$\begin{aligned} f(u_{ji}|\theta, a_i, b_i, c_i) &= \left(D(1 - c_i) \int_{\mathbb{R}} (\theta - b_i) [r_i(\theta) - P_i f_i(\theta)] W_i d\theta - \frac{1}{a_i} \left[1 + \frac{\log a_i - \mu_a}{\sigma_a^2} \right] \right) \\ &\times \left(-D(1 - c_i) \int_{\mathbb{R}} [r_i(\theta) - P_i f_i(\theta)] W_i d\theta - \frac{(b_i - \mu_b)}{\sigma_b^2} \right) \\ &\times \left(\int_{\mathbb{R}} [r_i(\theta) - P_i f_i(\theta)] \frac{W_i}{P_i^*} d\theta + \frac{(\alpha - 2)}{c_i} - \frac{\beta - 2}{1 - c_i} \right) \end{aligned} \quad (2.61)$$

Esta expressão define a função probabilidade da Teoria de Resposta ao Item.

2.2.17 Estimação das habilidades

A estimação bayesiana das habilidades é feita numa segunda etapa, considerando os parâmetros fixos. Pela suposição de independência entre as habilidades de diferentes respondentes, podem-se fazer as estimações em separado para cada indivíduo.

Assumindo que a distribuição *a priori* para $\theta_j, j = 1, \dots, n$, é Normal com vetor de parâmetro $\eta = (\mu, \sigma^2)$ conhecidos, a posteriori para a habilidade do indivíduo j pode ser escrita da seguinte forma:

$$g_j^*(\theta_j) \equiv g(\theta_j|u_j, \zeta, \eta) \propto P(u_j|\theta_j, \zeta) g(\theta_j|\eta). \quad (2.62)$$

Pode-se adotar alguma característica de $g_j^*(\theta_j)$ como estimador de θ_j , sendo que as mais adotadas são a média e a moda. Em nossos estudos, adotamos a estimação pela média.

Estimação pela média da posteriori - EAP

A estimação de θ_j pela média da *posteriori*, ou EAP - *Expected a posteori* - consiste em obter a esperança da *posteriori* e pode ser escrita da seguinte forma

$$g(\theta|u_j, \zeta, \eta) = \frac{P(u_j|\theta, \zeta) g(\theta|\eta)}{P(u_j|\zeta, \eta)} \quad (2.63)$$

Daí, temos que a esperança da posteriori é

$$\hat{\theta}_j \equiv E[\theta|u_j, \zeta, \eta] = \frac{\int_{\mathbb{R}} \theta g(\theta|\eta) P(u_j|\theta, \zeta) d\theta}{\int_{\mathbb{R}} g(\theta|\eta) P(u_j|\theta, \zeta) d\theta} \quad (2.64)$$

Esta forma de estimação pode ser calculada diretamente sem a necessidade de iteração, sendo que, por essa vantagem, é recomendada por alguns autores como escolha para a estimação das habilidades.

2.2.18 Aplicação do Algoritmo EM

O algoritmo EM é definido por Andrade et al. (2000) como um processo iterativo para determinação de estimativas de máxima verossimilhança de parâmetros de modelos de probabilidade na presença de variáveis aleatórias não observadas. Cada iteração deste processo é feita em dois passos: Esperança (E) e Maximização (M). No caso da TRI, o objetivo é obter estimativas de ζ na presença das variáveis não observadas θ . Neste caso, $u_{..}$ representa o vetor de dados incompletos e $(u_{..}, \theta)$ o vetor de dados completos. Seja $f(u_{..}, \theta|\zeta)$ a densidade conjunta dos dados completos.

Se $\hat{\zeta}^{(k)}$ é uma estimativa de ζ na iteração t , então os passos EM para obtenção de $\hat{\zeta}^{(k+1)}$ são:

Passo E: Calcular $E \left[\log f(u_{..}, \theta|\zeta) | u_{..}, \hat{\zeta}^{(k)} \right]$

Passo M: Obter $\hat{\zeta}^{(k+1)}$ que maximiza a função do Passo E

No passo M a maximização pode ser feita pelo algoritmo Newton-Raphson ou “Scoring” de Fisher. Há três formas do algoritmo EM, distinguidas pela relação entre a função (densidade) de probabilidade e a forma da família exponencial. A primeira forma se aplica quando a função é um membro regular da família exponencial; a segunda, quando a função não é um membro regular da família exponencial, mas um membro da família exponencial curvada (formada por distribuições em que há restrições no espaço paramétrico) e a terceira, quando a função não tem nenhuma relação com a família exponencial. O modelo logístico de 1 parâmetro (modelo de Rasch) é membro da família exponencial, os modelos de 2 e 3 parâmetros não são. Portanto, a terceira forma do algoritmo EM deve ser aplicada nestes casos.

Suponha as habilidades restritas a um conjunto de q valores, $\bar{\theta}_k$, $k = 1, \dots, q$, com probabilidade π_k , $k = 1, \dots, q$. Seja f_{ki} o número de indivíduos com habilidade $\bar{\theta}_k$ respondendo ao item i , $f_i = (f_{1i}, \dots, f_{qi})'$ com $\sum_{k=1}^q f_{ki} = N$, $f = (f_1, \dots, f_I)$. Seja r_{ki} o número de indivíduos com habilidade $\bar{\theta}_k$ respondendo corretamente ao item i , $r_i = (r_{1i}, \dots, r_{qi})'$ e $r = (r_1, \dots, r_I)$.

Se os n indivíduos responderão ao item i são selecionados ao acaso da população, a probabilidade conjunta que os f_{ki} indivíduos tenham habilidades θ_k , $k = 1, \dots, q$ é dada pela distribuição multinomial:

$$P(f_i|\pi) = \frac{n!}{\prod_{k=1}^q f_{ki}!} \prod_{k=1}^q \pi_k^{f_{ki}}, \quad i = 1, \dots, I.$$

Dados f_{ki} e $\bar{\theta}_k$, a probabilidade de ocorrerem r_{k1} acertos ao item i dentre as f_{k1} tentativas por indivíduos com habilidade $\bar{\theta}_k$ é

$$P(r_{k1}|f_{k1}, \bar{\theta}_k) = \binom{f_{k1}}{r_{k1}} P_{k1}^{r_{k1}} Q_{k1}^{f_{k1}-r_{k1}},$$

onde P_{ki} é a Função de Resposta ao Item - FRI, adotada com θ_j substituída por $\bar{\theta}_k$. A probabilidade conjunta de f e r , dados $\bar{\theta} = (\bar{\theta}_1, \dots, \bar{\theta}_q)$ e π , é

$$P(f, r | \bar{\theta}, \pi) = \left\{ \prod_{i=1}^I P(f_i | \pi) \right\} \left\{ \prod_{i=1}^I \prod_{k=1}^q P(r_{ki} | f_{ki}, \bar{\theta}_k) \right\}$$

$$\log L(\zeta) = C + \sum_{i=1}^I \sum_{k=1}^q \{ r_{ki} \log P_{ki} + (f_{ki} - r_{ki}) \log Q_{ki} \} \quad (2.65)$$

Onde $C = \log P(f | \pi) + \sum_{i=1}^I \sum_{k=1}^q \log \left(\frac{f_{ki}}{r_{ki}} \right)$ é constante com a relação a ζ . Temos que (f, r) são não-observáveis, mas tomando a esperança da log-verossimilhança, condicional em $u_{..}$ e ζ e usando a notação

$$\bar{r}_{ki} = E[r_{ki} | u_{..}, \zeta], \quad \bar{f}_{ki} = E[f_{ki} | u_{..}, \zeta] \quad \text{e} \quad \bar{C} = E[C | u_{..}, \zeta]$$

obtemos,

$$E[\log L(\zeta)] = \bar{C} + \sum_{i=1}^I \sum_{k=1}^q \{ \bar{r}_{ki} \log P_{ki} + (\bar{f}_{ki} - \bar{r}_{ki}) \log Q_{ki} \} \quad (2.66)$$

Maximizar 2.66 é equivalente a maximizar 2.65 e representa o Passo E do algoritmo EM.

Mais especificamente, os passos E e M são:

PASSO E: Usar os pontos de quadratura $\bar{\theta}_k$, os pesos A_k , $k = 1, \dots, q$ e estimativas iniciais dos parâmetros dos itens, ζ_i , $i = 1, \dots, I$, para gerar $g_j^*(\bar{\theta}_k)$ e, posteriormente, \bar{r}_{ki} e \bar{f}_{ki} , $i = 1, \dots, I$ e $k = 1, \dots, q$.

PASSO M: Com r e f obtidos no Passo E, resolver as equações de estimação para ζ_i , $i = 1, \dots, I$, usando o algoritmo Newton-Raphson ou “*Scoring*” de Fisher.

Esses passos compõem cada iteração do algoritmo EM, as quais serão repetidas até que algum critério de parada seja alcançado.

3 Análise Pedagógica dos itens

Com os dados tabulados dos testes aplicados na escolas - amostra aplicada e com os dados retirados do relatório microdados do ENEM, neste capítulo faremos a análise com interpretação pedagógica de cada item. Um de nossos objetivos é avaliar o ensino de probabilidade e estatística no ensino público, especificamente em Mato Grosso do Sul e aferir os conhecimentos dos alunos que estão cursando o terceiro ano do ensino médio comparando com os resultados daqueles que realizaram a prova do ENEM em anos anteriores.

Procedimento

A prova elaborada foi composta de 07 (sete) questões, uma de cada habilidade, de questões aplicadas nos ENEM's dos anos de 2013, H25 e H27, de 2014, H24, H26 e H28 e de 2015, H29 e H30. Aplicada a 206 (Duzentos e seis) alunos de escolas da cidade de Campo Grande, Dourados e Bonito, no mês de Novembro de 2017 corrigida utilizando o gabarito oficial apresentado pelo Inep. Ao mesmo tempo, foram retirados do arquivo Microdados do ENEM 2013, 2014 e 2015, disponibilizados pelo Inep, os dados de 12347 indivíduos que responderam a essas mesmas questões nos anos apresentados.

Para a estimação dos parâmetros, utilizou-se o suplemento do *Microsoft Excel - eirt* - que gerou os relatórios teoria clássica dos testes - TCT e da teoria de resposta ao item - TRI e os gráficos com as curvas características e de informação dos itens e dos testes. O programa se apropriou da estimação Bayesiana para gerar os valores dos parâmetros dos itens e estimou as habilidades utilizando a estimação pela média da posteriori ou *expected a posteriori* - EAP.

Os dados serão apresentados da seguinte forma:

- a) **Questão:** Compreende o texto base, enunciado e alternativas, o assunto abordado, a competência e a habilidade a qual pertence;
- b) **Tabelas:** Uma referente aos dados do ENEM e outra com os dados da amostra aplicada, composta com os seguintes itens:
 - **Gabarito:** refere-se a opção correta do item;
 - **Total:** Representa a quantidade de respondentes ao teste;
 - **Total de Acertos - (T.A.):** Representa a quantidade de respondentes que acertaram o item;
 - **Percentual de Acertos (P.A.):** Indica porcentagem de respondentes que acertaram o item medido pela TCT;

- **Discriminação (DISC):** Representam, respectivamente, o percentual de acertos dos 27% de participantes com as maiores e menores pontuações;
 - **Dificuldade (DIFI):** Dificuldade do item medido pela TCT;
 - **Percentual de Respostas:** Estabelece proporção de respostas dos participantes a cada uma das alternativas do item;
 - **Coefficiente Bisserial:** representam as correlações entre a pontuação dos participantes e a resposta a cada uma das alternativas (gabarito e distratores). Espera-se valor positivo para a opção correta e valores negativos para as demais opções, calculado por 2.1;
- c) **Parâmetros da TRI:** Representa o grau de dificuldade, discriminação e acerto casual de cada item;
- d) **Classificação (Clas.):** Classificação de acordo com o índice de dificuldade e discriminação;
- e) **Dois Gráficos:** Gráficos das curvas características do item - CCI concatenado com o gráfico de informação do item; um das questões aplicadas no ENEM e um da amostra aplicada;
- f) **Comentário:** Uma breve interpretação de cada item e gráfico.

3.1 Habilidade 24 - H24

A habilidade 24 tem como proposta *utilizar informações expressas em gráficos ou tabelas para fazer inferências*. Espera-se que o respondente identifique informações numéricas explicitadas nos gráficos. Com os dados apresentados, pretende-se que tire conclusões e consiga responder as questões sem fórmulas pré-estabelecidas.

3.1.1 Questão H24

Figura 3 – Questão referente a Habilidade 24

A taxa de fecundidade é um indicador que expressa a condição reprodutiva média das mulheres de uma região, e é importante para uma análise da dinâmica demográfica dessa região. A tabela apresenta os dados obtidos pelos Censos de 2000 e 2010, feitos pelo IBGE, com relação à taxa de fecundidade no Brasil.

Ano	Taxa de fecundidade no Brasil
2000	2,38
2010	1,90

Disponível em: www.saladeimprensa.ibge.gov.br. Acesso em: 31 jul. 2013.

Suponha que a variação percentual relativa na taxa de fecundidade no período de 2000 a 2010 se repita no período de 2010 a 2020.

Nesse caso, em 2020 a taxa de fecundidade no Brasil estará mais próxima de

- A 1,14.
- B 1,42.
- C 1,52.
- D 1,70.
- E 1,80.

Solução: A taxa de diminuição é dada por $\frac{1,9}{2,38} \simeq 0,8 \simeq 80\%$
 Logo, houve uma diminuição de 20%. Portanto devemos fazer $1,9 - \frac{20}{100} \cdot (1,9) = 1,9 - 0,38 = 1,52$
 O que nos leva a opção C.

3.1.2 Tabelas, Gráficos e Comentários

Tabela de Índices e Gráfico - Amostra - do item H24

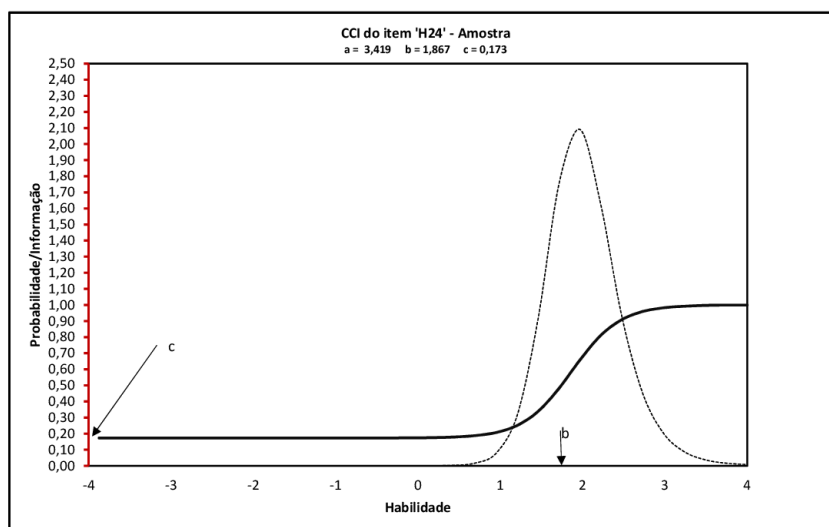
Tabela 3 – Tabela de Índices da Habilidade H24 - AMOSTRA

TCT		TRI					
Gabarito	C	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	206	A	0,095	-0,768	a	3,419	Muito Alto
T.A.	44	B	0,483	0,094			
P.A	21,4%	C	0,204	0,206	b	1,867	Muito Difícil
DISC	0,039	D	0,104	-0,743			
DIFI	0,214	E	0,114	-0,718	c	0,173	

T.A.= Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 4 – Curva Característica do Item H24 - Amostra



Fonte: Autor

Comentário:

Verifica-se pela TCT o índice de acerto da questão, 21,40%. Os coeficientes bisseriais negativos, com exceção das opções B e C e o índice positivo na opção C. O esperado é que o gabarito correto da questão tenha coeficiente positivo. Chama a atenção o coeficiente positivo na opção B com um alto índice de indivíduos que a escolheram (48,3%), mostrando a dificuldade dos respondentes no conceito de Taxa.

A TRI mostra que o parâmetro a tem índice de discriminação Muito Alto e o parâmetro b classifica a questão como Muito Difícil. A Figura 4, (gráfico da curva característica do item) mostra um deslocamento da curva “S” para direita e o gráfico de informação é bem acentuado. O alto índice de escolha na opção B, indica que indivíduos com habilidade de responder a questão foi atraída para esta opção. Talvez pelo respondente intuir que para calcular a taxa basta efetuar uma conta de subtração entre os valores dados.

Verifica-se, ainda, que por ser um item considerado muito difícil, apenas examinados com habilidades altas possuem probabilidades razoáveis de acertá-lo. O gráfico tem destaque apenas nas probabilidades de indivíduos com proficiência acima de 2 e as maiores informações estão entre o intervalo 1 a 3. Mostra também, que até a habilidade 1 não há alteração na curva, sendo bem acentuada a partir desse valor.

Tabela de Índices e Gráfico - ENEM - do item H24

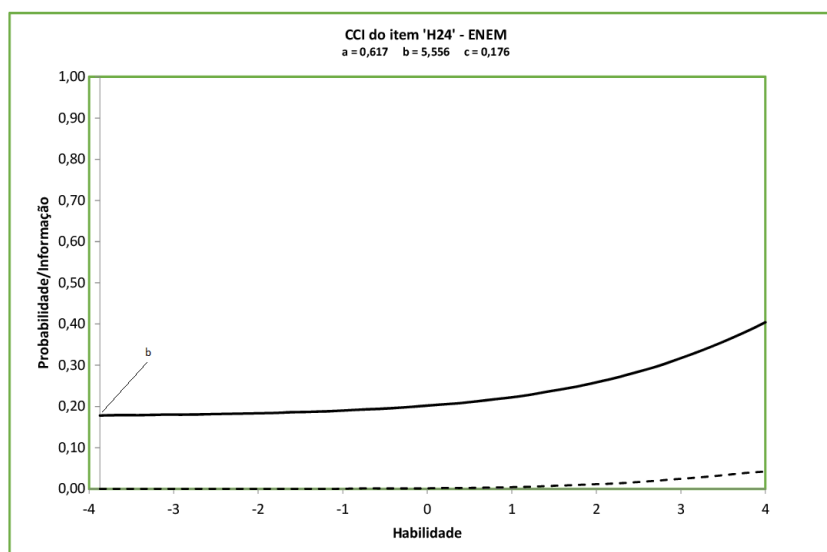
Tabela 4 – Tabela de Índices da Habilidade H24 - ENEM

		TCT			TRI		
Gabarito	C	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	12347	A	0,121	-0,390	a	0,617	Baixa
T.A	2545	B	0,414	1,529			
P.A	20,69%	C	0,206	1,724	b	5,556	Muito Difícil
DISC	0,032	D	0,117	-0,378			
DIFI	0,206	E	0,142	-0,344	c	0,176	

T.A.= Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 5 – Curva Característica do Item H24 - ENEM



Fonte: Autor

Comentário:

Verifica-se pela TCT o índice de acerto da questão, 20, 69%. Os coeficientes bisseriais negativos, com exceção a opção C, que é o esperado, pois o índice positivo nesta opção é o normal, haja vista ser o gabarito correto da questão, assim como os outros coeficientes serem negativos. Chama a atenção o coeficiente positivo na opção B, com um alto índice de indivíduos que marcaram a opção (41, 4%), mostrando a dificuldade dos respondentes no conceito de taxa.

A TRI, indicou que o parâmetro a tem índice de discriminação baixo e o parâmetro b classifica a questão como muito difícil. A Figura 5, (gráfico da curva característica do item), mostra que não se forma o “S” o que corrobora o baixo índice de discriminação. O alto índice de escolha da opção B, indica que indivíduos com habilidade de responder corretamente a questão foram atraídos por ela.

O gráfico mostra ainda, que o item tem informação praticamente nula, ou seja, não há distinção entre os indivíduos com habilidade alta ou habilidade baixa em responder a questão. Logo, a questão não foi competente para aquilo que foi proposto.

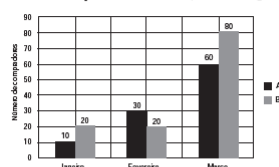
3.2 Habilidade 25 - H25

A habilidade 25 tem como proposta *resolver problema com dados apresentados em tabelas ou gráficos*. Espera-se que o respondente mostre a habilidade de realizar cálculos tendo como base a interpretação das informações presentes nos gráficos e tabelas, sendo que algumas soluções não dependem de fórmulas pré-estabelecidas.

3.2.1 Questão H25

Figura 6 – Questão referente a Habilidade 25

Uma loja acompanhou o número de compradores de dois produtos, A e B, durante os meses de janeiro, fevereiro e março de 2012. Com isso, obteve este gráfico:



A loja sorteará um brinde entre os compradores do produto A e outro brinde entre os compradores do produto B. Qual a probabilidade de que os dois sorteados tenham feito suas compras em fevereiro de 2012?

- A $\frac{1}{20}$
- B $\frac{3}{242}$
- C $\frac{5}{22}$
- D $\frac{6}{25}$
- E $\frac{7}{15}$

Solução: Probabilidade do comprador A ser de fevereiro: $P(A_f) = \frac{30}{100} = \frac{3}{10}$
 Probabilidade do comprador B ser de fevereiro: $P(B_f) = \frac{20}{120} = \frac{1}{6}$
 Probabilidade dos compradores A e B serem de fevereiro: $P(A_f) \cdot P(B_f) = \frac{3}{10} \cdot \frac{1}{6} = \frac{1}{20}$
 O que nos leva a opção A.

3.2.2 Tabelas, Gráficos e Comentários

Tabela de Índices e Gráfico - Amostra - do item H25

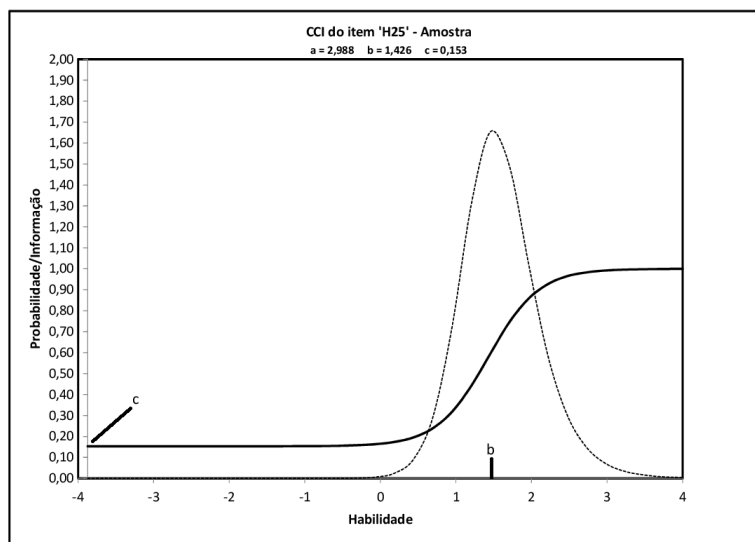
Tabela 5 – Tabela de Índices da Habilidade H25 - Amostra

		TCT			TRI		
Gabarito	A	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	206	A	0,244	0,294	a	2,988	Muito Alto
T.A	50	B	0,112	-0,789	b	1,426	Muito Difícil
P.A	21,4%	C	0,371	-0,203			
DISC	0,094	D	0,239	-0,479	c	0,153	
DIFI	0,243	E	0,034	-0,987			

T.A.= Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 7 – Curva Característica do Item H25 - Amostra



Fonte: Autor

Comentário

Verifica-se pela TCT que o grau de dificuldade do item 0,243 corrobora o índice de acerto da questão, 24,40%. Os coeficientes bisseriais negativos, com exceção da opção A, é o esperado, pois é o gabarito correto da questão, mas chama a atenção a grande quantidade de indivíduos que escolheram as opções C e D, em torno de 60%.

A TRI mostra que o parâmetro a tem índice de discriminação Muito Alto e o parâmetro b classifica a questão como Muito Difícil. A Figura 7, (gráfico da curva característica do item) mostra um deslocamento da curva “S” para direita e o gráfico de informação é bem acentuado. Observa-se que a curva do item é bem íngreme mostrando que o item é bem discriminativo e, ainda, mostra que o deslocamento no eixo da habilidade altera consideravelmente a probabilidade de acertar o item.

Podemos observar que no primeiro trecho da curva, ou seja de -4 até 1, ela se mantém paralela ao eixo das habilidades, ficando em torno de 15% que é praticamente o valor do parâmetro c , considerado chute. Após esse valor, houve uma alteração considerável na curva. Verifica-se, ainda, que por ser um item considerado muito difícil, apenas examinados com habilidades altas, acima de 1 possuem probabilidades razoáveis de acertá-lo, fato evidenciado também no gráfico informação do item, onde as maiores informações estão em torno das habilidades 1 e 2. Portanto, o item é ideal para que se possa aferir o conhecimento do indivíduo no conteúdo proposto.

Tabela de Índices e Gráfico - ENEM - do item H25

Tabela 6 – Tabela de Índices da Habilidade H25 - ENEM

		TCT			TRI		
Gabarito	A	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	12347	A	0,283	3,155	a	0,297	Muito Baixa
T.A	3547	B	0,186	-0,120			
P.A	28,39%	C	0,240	0,565	b	9,895	Muito Difícil
DISC	-0,118	D	0,204	0,025			
DIFI	0,278	E	0,084	-0,414	c	0,238	

T.A.= Total Acerto

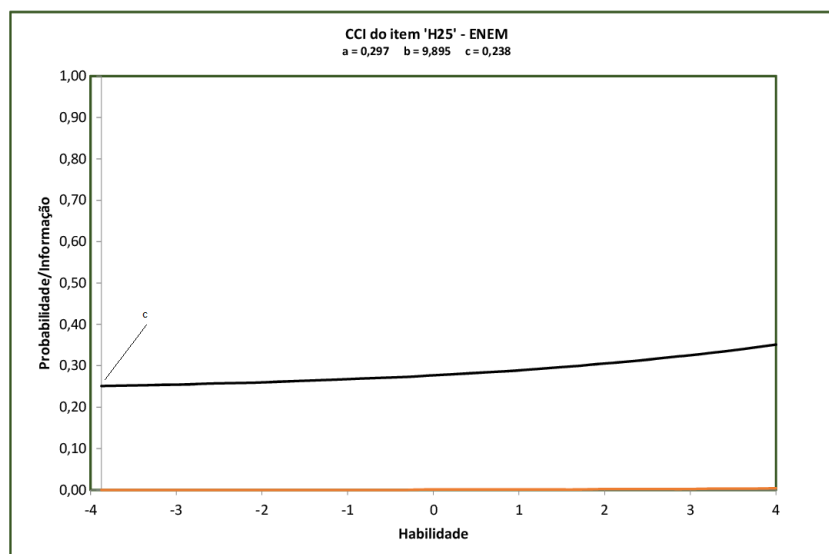
P.A. = Percentual Acerto

DISC = Discriminação

DIFI = Dificuldade

Fonte: Autor

Figura 8 – Curva Característica do Item H25 - ENEM



Fonte: Autor

Comentário

Verifica-se pela TCT que o grau de dificuldade do item 0,278 corrobora o índice de acerto da questão, 28,39%. Os coeficientes bisseriais são negativos nas opções B e E, e positivos nas opções A, C e D. O esperado é ser positivo apenas para o gabarito A, correto da questão. Chama a atenção a grande quantidade de indivíduos que escolheram as opções C e D, em torno de 44%.

A TRI mostra que o parâmetro a tem índice de discriminação Muito Baixo e o parâmetro b classifica a questão como Muito Difícil. A Figura 8, (gráfico da curva característica do item) mostra que não há curva em forma de “S” e que o gráfico de informação não existe, ou seja, o item não dá nenhuma informação pertinente. Observa-se que a curva do item é praticamente toda paralela ao eixo da habilidade e que alterações consideráveis nesse eixo não alteram a probabilidade de acertar o item, igualando os respondentes com habilidade alta com os de habilidade baixa.

Podemos observar que não há alteração da curva nas diversas habilidades estando todos em torno de 23% que é praticamente o valor do parâmetro c , considerado “chute”. De acordo com o apresentado no gráfico, o item não é ideal para que se possa aferir o conhecimento do indivíduo no conteúdo proposto.

3.3 Habilidade 26 - H26

A habilidade 26 tem como proposta *analisar informações expressas em gráficos ou tabelas como recurso para a construção de argumentos*. Espera-se que ao analisar as

alternativas juntamente com gráficos ou tabelas, o examinado consiga realizar cálculos para chegar ao resultado esperado.

3.3.1 Questão H26

Figura 9 – Questão referente a Habilidade 26

O Ministério da Saúde e as unidades federadas promovem frequentemente campanhas nacionais e locais de incentivo à doação voluntária de sangue, em regiões com menor número de doadores por habitante, com o intuito de manter a regularidade de estoques nos serviços hemoterápicos. Em 2010, foram recolhidos dados sobre o número de doadores e o número de habitantes de cada região conforme o quadro seguinte.

Região	Doadores	Número de habitantes	Doadores/habitantes
Nordeste	820 959	53 081 950	1,5%
Norte	232 079	15 864 454	1,5%
Sudeste	1 521 766	80 364 410	1,9%
Centro-Oeste	362 334	14 058 094	2,6%
Sul	690 391	27 386 891	2,5%
Total	3 627 529	190 755 799	1,9%

Os resultados obtidos permitiram que estados, municípios e o governo federal estabelecessem as regiões prioritárias do país para a intensificação das campanhas de doação de sangue.

A campanha deveria ser intensificada nas regiões em que o percentual de doadores por habitantes fosse menor ou igual ao do país.

Disponível em: <http://bvsms.saude.gov.br>. Acesso em: 2 ago. 2013 (adaptado).

As regiões brasileiras onde foram intensificadas as campanhas na época são

- A Norte, Centro-Oeste e Sul.
- B Norte, Nordeste e Sudeste.
- C Nordeste, Norte e Sul.
- D Nordeste, Sudeste e Sul.
- E Centro-Oeste, Sul e Sudeste.

Fonte: Extraído de: Avaliação ENEM/2014

Solução:

Pede-se valor menores ou iguais do que valores nacionais. Analisando a Tabela temos:

Doadores do Nacionais: 1,9%.

Doadores do Nordeste: 1,5%.

Doadores do Norte: 1,5%.

Doadores do Sudeste: 1.9%.

Logo, opção B.

3.3.2 Tabelas, Gráficos e Comentários

Tabela dos Índices e Gráfico - Amostra - do item H26

Tabela 7 – Tabela de Índices da Habilidade H26 - Amostra

		TCT			TRI		
Gabarito	B	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	206	A	0,183	-0,421	a	0,524	Baixa
T.A	109	B	0,540	0,674			
P.A	54,0%	C	0,040	-0,817	b	0,513	Mediana
DISC	0,126	D	0,079	-0,754			
DIFI	0,529	E	0,158	-0,486	c	0,165	

T.A.= Total Acerto

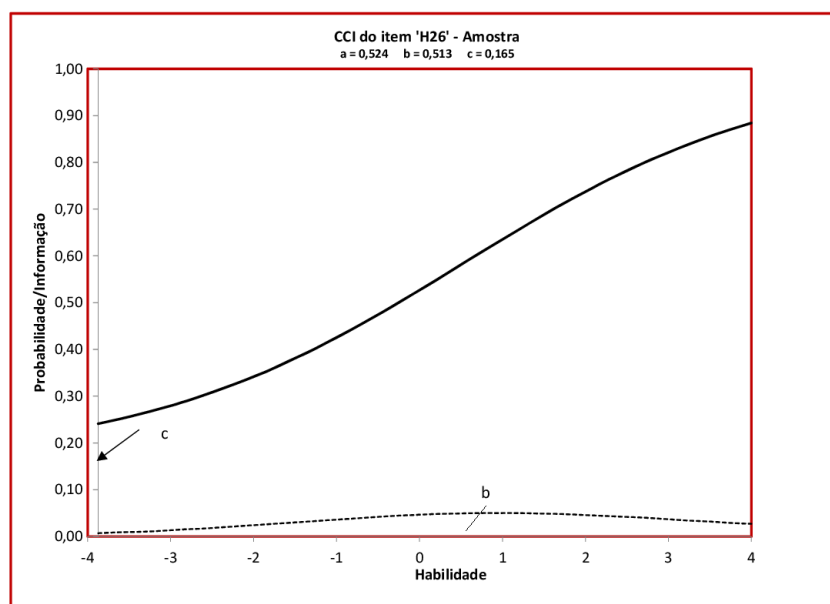
P.A. = Percentual Acerto

DISC = Discriminação

DIFI = Dificuldade

Fonte: Autor

Figura 10 – Curva Característica do Item H26 - Amostra



Fonte: Autor

Comentário

Verifica-se pela TCT o índice alto de acerto da questão, 54,00%. Os coeficientes bisseriais todos negativos sendo positivo apenas na opção A, o que é esperado, pois o gabarito A é o correto da questão. As opções A e E também atraíram respondentes, mas com porcentagem dentro do esperado no parâmetro c da TRI.

A TRI mostra que o parâmetro a tem índice de discriminação Baixa e o parâmetro b classifica a questão como Mediana. A Figura 10, (gráfico da curva característica do item) mostra uma curva muito suave e o gráfico de informação praticamente não fornece dados pertinentes. A curva suave indica que deslocamentos no eixo da habilidade provoca alterações discretas na probabilidade de acerto, sendo alteradas com um pouco mais de intensidade a partir da habilidade 0 (habilidade média).

Podemos concluir, então, que apenas indivíduos com habilidades acima da média tenham probabilidade de acertar o item (alteração no final da curva). O gráfico de informação é bem baixo, mostrando poucos dados, pois há apenas uma leve curvatura em torno da habilidade média, porém sem relevância. O item não discrimina bem.

Tabela dos Índices e Gráfico - ENEM - do item H26

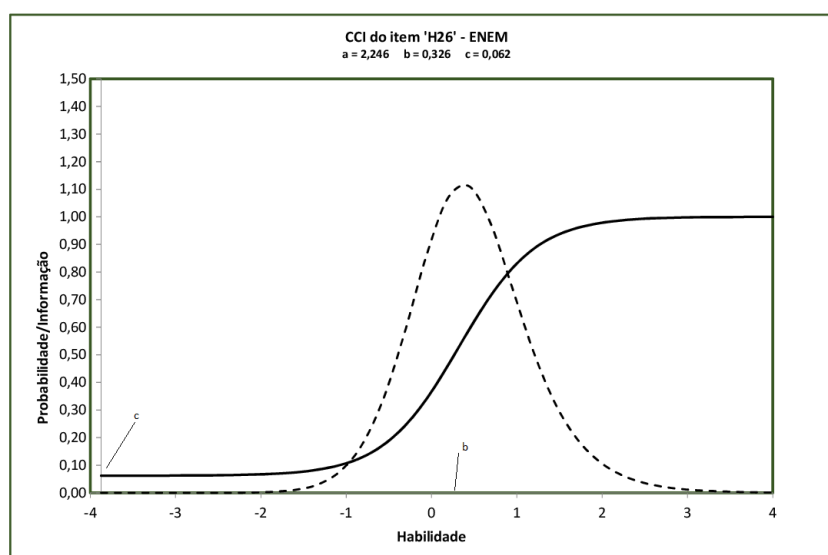
Tabela 8 – Tabela de Índices da Habilidade H25 - ENEM

TCT		TRI					
Gabarito	B	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	12246	A	0,147	-0,589	a	2,246	Muito Alta
T.A	5331	B	0,434	6,815			
P.A	43,4%	C	0,096	-0,812	b	0,326	Mediana
DISC	0,169	D	0,133	-0,655			
DIFI	0,432	E	0,190	-0,292	c	0,062	

T.A.= Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 11 – Curva Característica do Item H26 - ENEM



Fonte: Autor

Comentário:

Pela TCT verifica-se o índice alto de acerto da questão, 43,4%. Os coeficientes bisseriais todos negativos e positivo apenas na opção A, o que é esperado, pois o gabarito A é o correto da questão. As opções A e D também atraíram examinados, mas com porcentagens baixas, dentro do esperado com o parâmetro c da TRI.

A TRI mostra que o parâmetro a tem índice de discriminação muito alto e o parâmetro b classifica a questão como mediana. A Figura 11, (gráfico da curva característica do item) mostra uma curva bem acentuada e o gráfico de informação também é bem característico, com informações variando no intervalo -1 a 2. Vê-se, que os dois gráficos estão bem centralizados, sem deslocamentos à direita ou à esquerda, demonstrando que a maioria das probabilidades de acertos estão dentro da habilidade média, habilidade zero.

Podemos concluir, então, que apenas indivíduos com habilidades mediana em diante tenham probabilidade alta de acertar o item. As curvas apresentadas intui que pequeno deslocamento no eixo da habilidade causa alterações importantes na probabilidade de acerto. Com índice de discriminação alto, o item é ideal para que se possa aferir o aprendizado dos indivíduos, sendo competente em cumprir aquilo para que foi proposto.

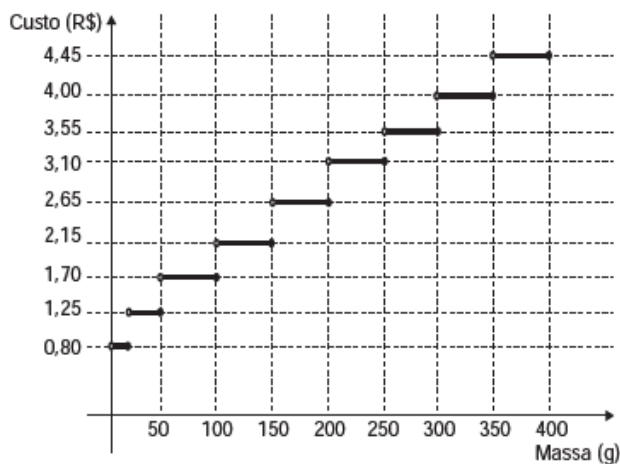
3.4 Habilidade 27 - H27

A proposta da habilidade 27 é *Calcular medidas de tendência central ou de dispersão de um conjunto de dados expressos em uma tabela de frequências de dados agrupados (não em classes) ou em gráficos*. Requer do indivíduo a capacidade de calcular as medidas de posição e dispersão, classificadas como média, mediana, moda, desvio padrão e variância, ou resolver problemas analisando gráficos, sem fórmulas previamente estabelecidas.

3.4.1 Questão H27

Figura 12 – Questão referente a Habilidade 26

Deseja-se postar cartas não comerciais, sendo duas de 100 g, três de 200 g e uma de 350 g. O gráfico mostra o custo para enviar uma carta não comercial pelos Correios:



Disponível em: www.correios.com.br. Acesso em: 2 ago. 2012 (adaptado).

O valor total gasto, em reais, para postar essas cartas é de

- A 8,35.
- B 12,50.
- C 14,40.
- D 15,35.
- E 18,05.

Fonte: Extraído de: ENEM/2013

Solução:
 Analisando a Gráfico temos:
 Valor da carta de 100g é 1,70, logo $2 \cdot 1,70 = 3,40$
 Valor da carta de 200g é 2,65, logo $3 \cdot 2,65 = 7,95$
 Valor da carta de 350g é 4,00, logo $1 \cdot 4,00 = 4,00$.
 Portanto o valor pago é $2,40 + 7,95 + 4,00 = 15,35$.
 O que nos leva a opção D.

3.4.2 Tabelas, Gráficos e Comentários

Tabela de Índices e Gráfico - Amostra - do Item H27

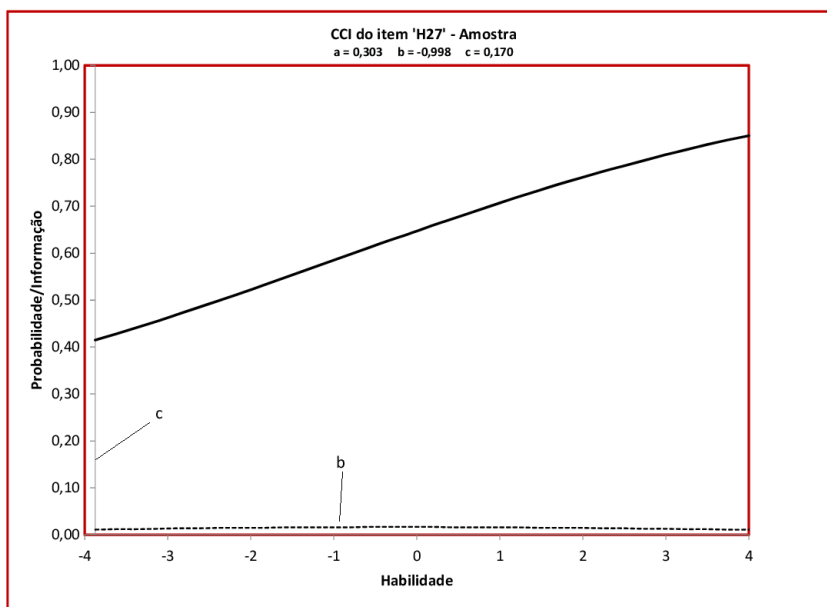
Tabela 9 – Tabela de Índices da Habilidade H27 - Amostra

		TCT			TRI		
Gabarito	D	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	206	A	0,132	0,495	a	0,303	Muito Baixa
T.A	133	B	0,093	-0,561			
P.A	65,20%	C	0,088	-0,537	b	-0,998	Fácil
DISC	0,006	D	0,652	0,769			
DIFI	0,646	E	0,034	-0,660	c	0,170	

T.A.= Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 13 – Curva Característica do Item H27 - Amostra



Fonte: Autor

Comentário

Verifica-se pela TCT que o índice de acerto da questão, 65,20%, é bastante alto. Verifica-se ainda, coeficientes bisseriais negativos, com exceção das opções A e D. O índice

positivo na opção D é o esperado, pois é o gabarito correto da questão, mas chama a atenção o coeficiente positivo na opção A com um índice de indivíduos que escolheram a opção dentro dos parâmetros c da TRI.

A TRI mostra que o parâmetro a tem índice de discriminação muito baixa e o parâmetro b classifica a questão como fácil. A Figura 13, (gráfico da curva característica do item) não mostra uma curva em formato de “S” proeminente, sendo extremamente suave e discreto, mantendo-se na parte superior e o gráfico de informação praticamente não existe, sendo quase uma reta.

Verifica-se, ainda, que o baixo índice de discriminação indica que deslocamentos no eixo da habilidade não causam alterações importantes na probabilidade de acerto. A curva mantém-se na parte superior, ou seja, as habilidades baixas tem probabilidade de acertos acima de 40%. Isto indica que o item não é discriminante e portando não cumpre o papel de aferir o conhecimento do examinado no conteúdo cobrado e nem diferencia indivíduos pelas habilidades.

Tabela de Índices e Gráfico - ENEM - do Item H27

Tabela 10 – Tabela de Índices da Habilidade H27 - ENEM

		TCT			TRI		
Gabarito	D	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	12347	A	0,146	-0,513	a	1,045	Moderada
T.A	5361	B	0,146	-0,513			
P.A	42,93%	C	0137	-0,551	b	0,721	Difícil
DISC	0,000	D	0,429	6,414			
DIFI	0,419	E	0,093	-0,754	c	0,109	

T.A.= Total Acerto

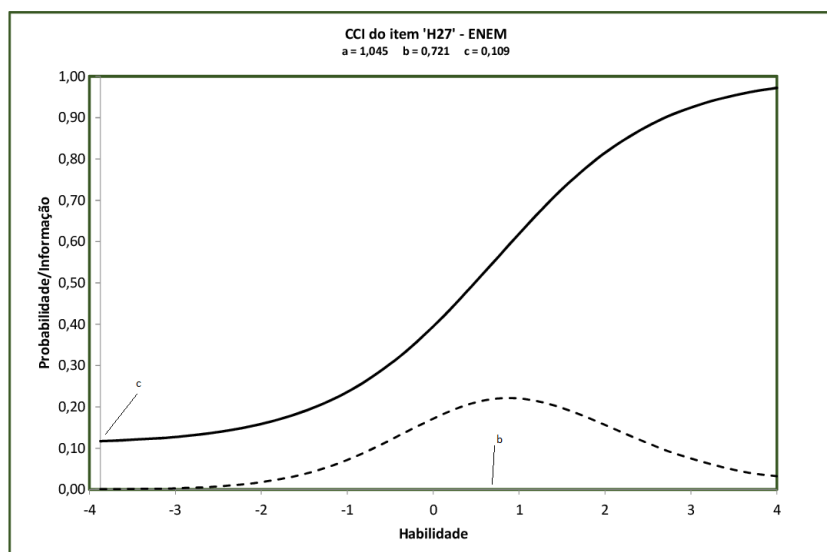
P.A. = Percentual Acerto

DISC = Discriminação

DIFI = Dificuldade

Fonte: Autor

Figura 14 – Curva Característica do Item H27 - ENEM



Fonte: Autor

Comentário

Verifica-se pela TCT que índice alto de acerto da questão, 42,93%. Todos os coeficientes bisseriais negativos, com exceção da opção D. O índice positivo na opção D é o esperado, pois é o gabarito correto da questão.

A TRI mostra que o parâmetro a tem índice de discriminação moderada e o parâmetro b classifica a questão como difícil. A Figura 14 (gráfico da curva característica do item) mostra uma curva em forma de “S” bem acentuada e ligeiramente deslocada para direita, assim como o gráfico informação do item.

Verifica-se, ainda, que o índice de discriminação moderado é claramente apresentado no gráfico, pois mostra que as maiores probabilidades de acertos estão acima da média zero (habilidade média), a partir da qual a curva fica mais íngreme. Isso mostra que pequenos deslocamentos no eixo da habilidade causam alterações de destaque na probabilidade de acerto, facilmente verificada nas habilidades acima de 2. O gráfico de informação mostra que maior probabilidade de acerto está entre o -1 a 3. Isto indica que o item é bem discriminante e portanto, cumpre o papel de aferir o conhecimento do examinado no conteúdo cobrado.

3.5 Habilidade 28 - H28

A proposta da habilidade 28 consiste em *Resolver situação-problema que envolva conhecimentos de estatística probabilidade*. O respondente necessita demonstrar se consegue uma situação-problema por meio do cálculo das medidas de dispersão, as vezes necessitando

realizar os cálculos de cada uma delas.

3.5.1 Questão H28

Figura 15 – Questão referente a Habilidade 28

O psicólogo de uma empresa aplica um teste para analisar a aptidão de um candidato a determinado cargo. O teste consiste em uma série de perguntas cujas respostas devem ser verdadeiro ou falso e termina quando o psicólogo fizer a décima pergunta ou quando o candidato der a segunda resposta errada. Com base em testes anteriores, o psicólogo sabe que a probabilidade de o candidato errar uma resposta é 0,20.

A probabilidade de o teste terminar na quinta pergunta é

- A 0,02048.
- B 0,08192.
- C 0,24000.
- D 0,40960.
- E 0,49152.

Fonte: Extraído de: Avaliação ENEM/2014

Solução:

Considerando que:

1. A probabilidade de errar é $0,20 = \frac{20}{100}$
2. A probabilidade de acertar é $0,80 = \frac{80}{100}$
3. Pergunta-se a probabilidade do teste terminar na quinta pergunta.

Teremos:

Fixando o segundo erro na quinta pergunta, o primeiro erro pode estar na primeira ou na segunda ou na terceira ou na quarta per gunda. Daí, temos que para as quatro opções anteriores, resultará:

$$4 \cdot \frac{2}{10} \cdot \frac{8}{10} \cdot \frac{8}{10} \cdot \frac{8}{10} \cdot \frac{2}{10} \Rightarrow \frac{8192}{100000} = 0,08192$$

O que nos leva a opção B.

3.5.2 Tabelas, Gráfico e Comentários

Tabela de Índices e Gráfico - Amostra - do item H28

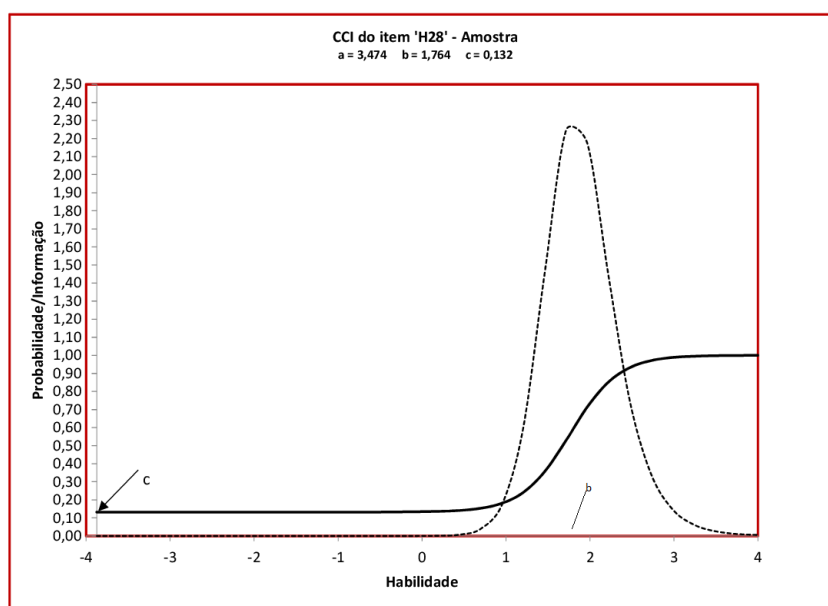
Tabela 11 – Tabela de Índices da Habilidade H28 - Amostra

		TCT			TRI		
Gabarito	B	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	206	A	0,249	-0,329	a	3,474	Muito Alta
T.A	37	B	0,184	0,012			
P.A	18,40%	C	0,259	-0,292	b	1,764	Muito Difícil
DISC	0,038	D	0,214	-0,482			
DIFI	0,180	E	0,095	-0,720	c	0,132	

T.A.= Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 16 – Curva Característica do Item H28 - Amostra



Fonte: Autor

Comentário

Verifica-se pela TCT o baixo índice de acerto da questão, 18,40% e podemos ver que os coeficientes bisseriais são todos negativos, com exceção a opção B, sendo o esperado, pois é a opção correta. Chama a atenção que a porcentagem de escolha das opções A, C e D serem maiores que da opção, B. Isso mostra que indivíduos com habilidade para resolver a questão foram atraídos por essas opções.

A TRI, mostra que o parâmetro a tem índice de discriminação muito alta e o parâmetro b demonstra que a questão é graduada como muito difícil, o que claramente transparece na Figura 16 (gráfico da curva característica do item). Vemos que de -4 a 1 a curva não se altera, sendo paralela ao eixo da habilidade, ou seja, o índice de probabilidade de acerto é praticamente o mesmo para indivíduos com habilidade inferior a um, com variações acentuadas para habilidades acima desse valor.

Logo, por ser um item considerado muito difícil, examinados com habilidades altas possuem probabilidades razoáveis de acertá-lo. A probabilidade de indivíduo com proficiência dois acertar o item está acima de 50%. Claramente se vê que o gráfico é deslocado para direita devido a dificuldade e a alta discriminação do item. O gráfico de informação, é bem deslocado para a direita demonstrando que a maioria dos respondentes estão na habilidade entre 1 a 3. Logo, é um bom item para aferir conhecimento.

Tabela de Índices e Gráfico - ENEM - do item H28

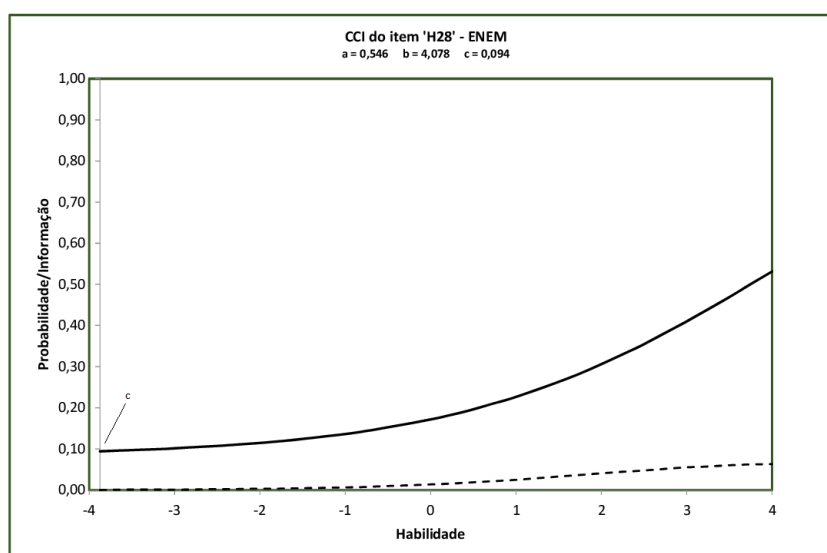
Tabela 12 – Tabela de Índices da Habilidade H28 - ENEM

		TCT			TRI		
Gabarito	B	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	12347	A	0,259	0,919	a	0,546	Baixa
T.A	2229	B	0,181	1,017			
P.A	18,16%	C	0,230	0,226	b	4,078	Muito Difícil
DISC	0,051	D	0,209	0,347			
DIFI	0,181	E	0,118	-0,105	c	0,082	

T.A. = Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 17 – Curva Característica do Item H28 - ENEM



Fonte: Autor

Comentário:

Verifica-se pela TCT o baixo índice de acerto da questão, em torno de 18,16% e podemos ver que os coeficientes bisseriais são todos positivos, com exceção a opção E, não sendo o esperado, pois apenas a opção correta (gabarito B) deveria ter coeficiente bisserial positivo. Chama a atenção que a porcentagem de escolha das opções A, C e D serem maiores que da opção B. Isso mostra que indivíduos com habilidade para resolver a questão foram atraídos por essas opções.

A TRI, mostra que o parâmetro a tem índice de discriminação baixa e o parâmetro b demonstra que a questão é graduada como muito difícil, o que claramente transparece na Figura 17, (gráfico da curva característica do item). Vemos que de -4 a 1 , aproximadamente, a curva não se altera, sendo praticamente paralela ao eixo da habilidade, ou seja, o índice de probabilidade de acerto é praticamente a mesma para indivíduos com habilidade

inferior um, apenas com variações discretas em habilidades acima desse valor. O gráfico de informação praticamente não existe.

Logo, por ser um item considerado muito difícil, apenas examinados com habilidades altas possuem probabilidades razoáveis de acertá-lo, porém com porcentagens baixas, inferior a 60%. Claramente se vê que o gráfico tem alterações na direita, onde há maior a habilidade. Portanto o item é pouco discriminante e não cumpre o papel de aferir o conhecimento do respondente sobre o conteúdo cobrado.

3.6 Habilidade 29 - H29

A habilidade 29 tem como proposta *utilizar conhecimentos de estatística e probabilidade como recurso para a construção de argumentação*. O indivíduo deve chegar ao resultado esperado analisando cada argumento apresentado nas alternativas, por meio de cálculos que comprovem qual a opção correta.

3.6.1 Questão H29

Figura 18 – Questão referente a Habilidade 29

Uma competição esportiva envolveu 20 equipes com 10 atletas cada. Uma denúncia à organização dizia que um dos atletas havia utilizado substância proibida. Os organizadores, então, decidiram fazer um exame *antidoping*. Foram propostos três modos diferentes para escolher os atletas que irão realizá-lo:

Modo I: sortear três atletas dentre todos os participantes;

Modo II: sortear primeiro uma das equipes e, desta, sortear três atletas;

Modo III: sortear primeiro três equipes e, então, sortear um atleta de cada uma dessas três equipes.

Considere que todos os atletas têm igual probabilidade de serem sorteados e que $P(I)$, $P(II)$ e $P(III)$ sejam as probabilidades de o atleta que utilizou a substância proibida seja um dos escolhidos para o exame no caso do sorteio ser feito pelo modo I, II ou III.

Comparando-se essas probabilidades, obtém-se

- A $P(I) < P(III) < P(II)$
- B $P(II) < P(I) < P(III)$
- C $P(I) < P(II) = P(III)$
- D $P(I) = P(II) < P(III)$
- E $P(I) = P(II) = P(III)$

Fonte: Extraído de: Avaliação ENEM/2015

Solução

Temos que:

$$P(I). \text{ Espaço Amostral é } 20 \cdot 10 = 200. \text{ Logo } P(I) = 3 \cdot \frac{1}{200} = \frac{3}{200}$$

$$P(II). \text{ Sortear umas das equipes: } \frac{1}{20}; \text{ Sortear nessa equipe 3 atletas: } 3 \cdot \frac{1}{10}. \text{ Logo } P(II) = \frac{3}{200}$$

$$P(III). \text{ Sortear 3 equipes: } 3 \cdot \frac{1}{20}; \text{ Sortear um atleta dessas equipes: } \frac{1}{10}. \text{ Logo } 3 \cdot \frac{1}{20} \cdot \frac{1}{10} = \frac{3}{200}$$

Ou seja, opção E.

3.6.2 Tabelas, Gráficos e Comentários

Tabela de Índices e gráfico - Amostra - do Item H29

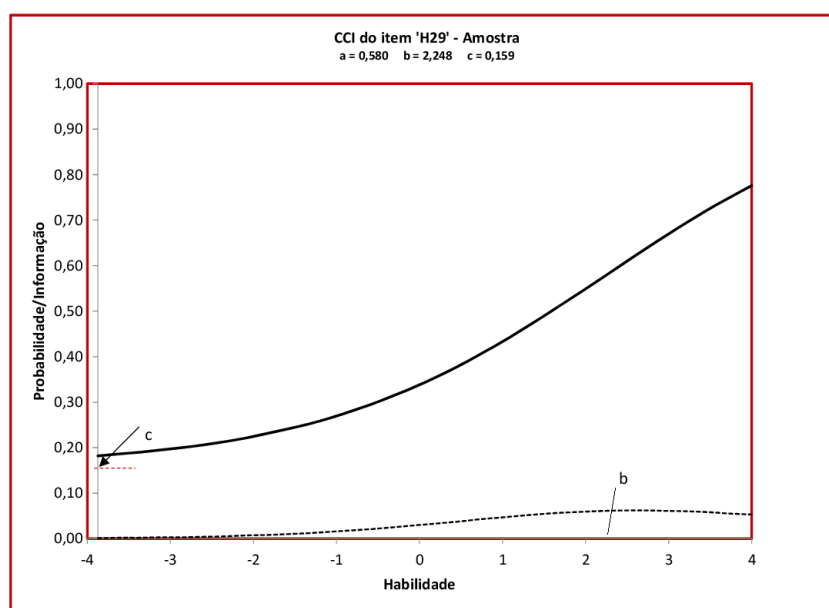
Tabela 13 – Tabela de Índices da Habilidade H29 - Amostra

		TCT			TRI		
Gabarito	E	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	206	A	0,158	-0,907	a	0,580	Baixa
T.A	72	B	0,222	-0,670			
P.A	35,50%	C	0,163	-0,808	b	2,248	Muito Difícil
DISC	0,136	D	0,103	-0,959			
DIFI	0,350	E	0,355	-0,530	c	0,159	

T.A.= Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 19 – Curva Característica do Item H29 - Amostra



Fonte: Autor

Comentário:

Verifica-se pela TCT um bom índice de acerto da questão, 35,50% e podemos ver que os coeficientes bisseriais são todos negativos, sendo que o ideal é que a opção correta (gabarito E) tenha coeficiente bisserial positivo. Isso indica que talvez as alternativas apresentadas devam ser reelaboradas.

Pela TRI, verifica-se que o parâmetro *a* tem índice de discriminação baixa e o parâmetro *b* classifica a questão como muito difícil, o que claramente transparece na Figura 19, (gráfico da curva característica do item). Vemos que a curva “S” é suave e se mantém na parte superior do gráfico, indicando que deslocamento positivo no eixo das

proficiências praticamente conduz a mudanças na probabilidade de acertos, principalmente nas habilidades maiores. O gráfico de informação não fornece dados relevantes.

Logo, por ser um item considerado muito difícil, apenas examinados com habilidades altas possuem probabilidades razoáveis de acertá-lo. Verifica-se que o gráfico tem leves alterações apenas na probabilidade de indivíduo com proficiência acima de dois. Mesmo não sendo o ideal, o item tem relativa discriminação necessitando ajuste.

Tabela de Índices e gráfico - ENEM - do Item H29

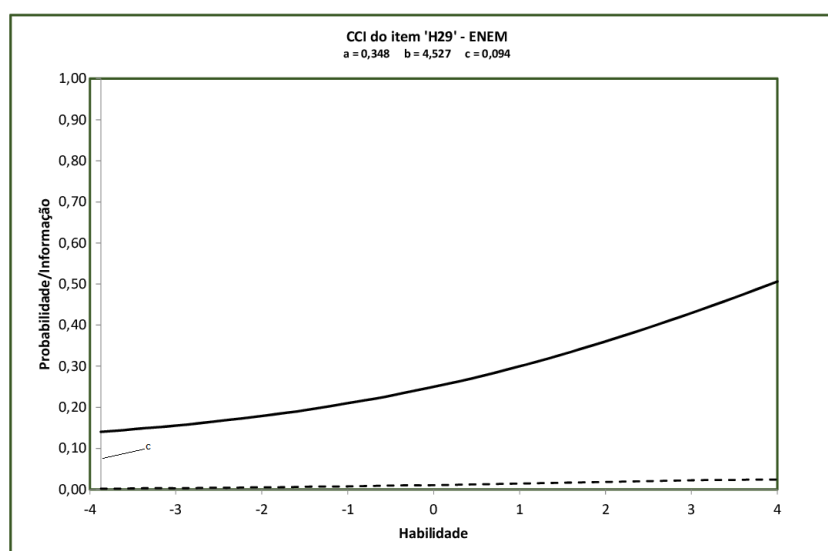
Tabela 14 – Tabela de Índices da Habilidade H29 - ENEM

		TCT			TRI		
Gabarito	E	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	12347	A	0,163	0,416	a	0,348	Muito Baixa
T.A	2739	B	0,192	0,128			
P.A	22,35%	C	0,217	0,254	b	4,527	Muito Difícil
DISC	0,004	D	0,203	0,585			
DIFI	0,219	E	0,223	1,354	c	0,094	

T.A. = Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 20 – Curva Característica do Item H29 - ENEM



Fonte: Autor

Comentário

Verifica-se pela TCT o índice de acerto da questão, 22,30% e podemos ver que há problemas com os coeficiente bisseriais, pois para a opção correta, gabarito E, o coeficiente é positivo, como esperado, mas as outras opções também são positivas, exibindo,

inclusive, uma proximidade entre os escores, ou seja, as alternativas atraíram indivíduos com habilidade para responder, ou utilizou-se o recurso do “chute”.

Pela TRI, verifica-se que o parâmetro a tem índice de discriminação baixa e o parâmetro b demonstra que a questão é graduada como muito difícil. A curva característica, figura 20, não apresenta o “S”, mostrando que a probabilidade de acertar o item não sofre um aumento proporcional.

Por apresentar baixo índice de discriminação, vemos que o gráfico de informação é praticamente uma reta, e portanto, probabilidade de acertar o item praticamente não apresenta variação entre os indivíduos, nem mesmo naqueles com habilidade alta, pois a variação é pequena, não ultrapassando de 50%. Portanto, a questão não tem condições de cumprir o papel de uma boa avaliação.

3.7 Habilidade 30 - H30

A Habilidade 30 tem como proposta *Avaliar propostas de intervenção na realidade utilizando conhecimentos de estatística e probabilidade*. O aluno deve conhecer os conceitos de probabilidade, tais como espaço amostral, evento, amostra, experimento aleatório, utilizando-se dos meios estatísticos (medidas de dispersão) para realizar os cálculos e resolver cada uma da proposta que foi apresentada.

3.7.1 Questão H30

Figura 21 – Questão referente a Habilidade 30

O HPV é uma doença sexualmente transmissível. Uma vacina com eficácia de 98% foi criada com o objetivo de prevenir a infecção por HPV e, dessa forma, reduzir o número de pessoas que venham a desenvolver câncer de colo de útero. Uma campanha de vacinação foi lançada em 2014 pelo SUS, para um público-alvo de meninas de 11 a 13 anos de idade. Considera-se que, em uma população não vacinada, o HPV acomete 50% desse público ao longo de suas vidas. Em certo município, a equipe coordenadora da campanha decidiu vacinar meninas entre 11 e 13 anos de idade em quantidade suficiente para que a probabilidade de uma menina nessa faixa etária, escolhida ao acaso, vir a desenvolver essa doença seja, no máximo, de 5,9%. Houve cinco propostas de cobertura, de modo a atingir essa meta:

Proposta I: vacinação de 90% do público-alvo.

Proposta II: vacinação de 55,8% do público-alvo.

Proposta III: vacinação de 88,2% do público-alvo.

Proposta IV: vacinação de 49% do público-alvo.

Proposta V: vacinação de 95,9% do público-alvo.

Para diminuir os custos, a proposta escolhida deveria ser também aquela que vacinasse a menor quantidade possível de pessoas.

Disponível em: www.virus HPV.com.br. Acesso em: 30 ago. 2014 (adaptado).

A proposta implementada foi a de número

- A. I.
- B. II.
- C. III.
- D. IV.
- E. V.

Solução

Como a vacina tem eficácia de 98%, então:

Pessoas vacinadas e imunes: $\frac{98}{100}$

Pessoas vacinadas e não imunes: $\frac{2}{100}$

Mas, o HPV acomete apenas 50% dos não imunes. Logo, de $\frac{2}{100}$, temos:

Pessoas vacinadas e saudáveis: $\frac{1}{100}$

Pessoas vacinadas e não saudáveis: $\frac{1}{100}$

Portanto:

Pessoas vacinadas e saudáveis: $\frac{98}{100} + \frac{1}{100} = \frac{99}{100}$

Pessoas vacinadas e doentes: $\frac{1}{100}$

Pessoas não vacinadas e saudáveis: $\frac{50}{100}$

Pessoas não vacinadas e doentes: $\frac{50}{100}$

Probabilidade máxima de desenvolver a doença: $\frac{5,9}{100}$

Total de pessoas vacinadas é x e total de pessoas não vacinadas é y

Logo devemos somar as probabilidades de doentes e igualar com a probabilidade máxima a ser atingida.

$\frac{1}{100}x + \frac{50}{100}y = \frac{5,9}{100}(x + y)$, que resultará $x = 9y \Rightarrow \frac{x}{y} \Rightarrow \frac{9}{1}$

Ou seja, a relação é de 9 por 1, o que implica que necessita-se vacinar 90% das meninas que nos leva a opção A.

3.7.2 Tabelas, Gráficos e Comentários

Tabela de Índices e Gráfico - Amostra - do Item H30

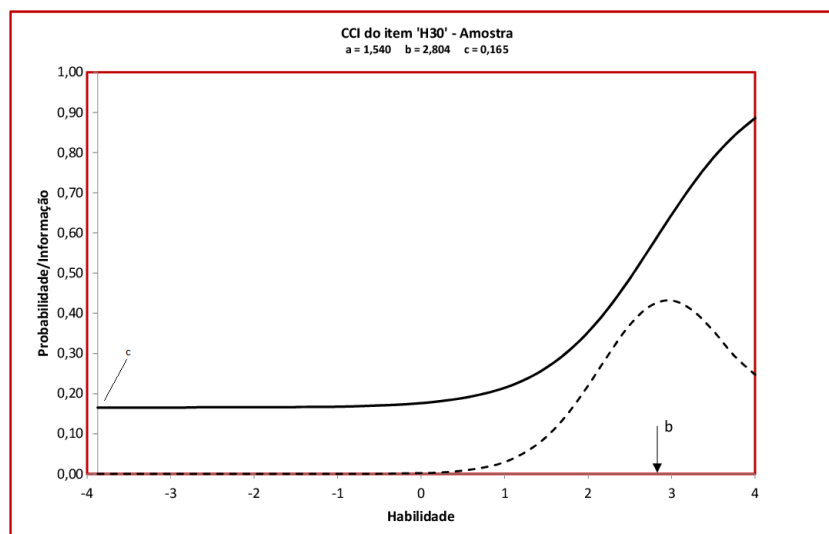
Tabela 15 – Tabela de Índices da Habilidade H30 - Amostra

		TCT			TRI		
Gabarito	A	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	206	A	0,093	-0,195	a	1,540	Alta
T.A	19	B	0,195	-0,301			
P.A	9,30%	C	0,205	-0,273	b	2,804	Muito Difícil
DISC	-0,035	D	0,434	-0,010			
DIFI	0,194	E	0,073	-0,439	c	0,165	

T.A. = Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 22 – Curva Característica do Item H30 - Amostra



Fonte: Autor

Comentário

Verifica-se pela TCT a baixa proporção de acerto da questão, em torno de 9,30% e podemos ver que os coeficientes bisseriais são todos negativos, sendo que o ideal é que a opção correta (E) tenha coeficiente bisserial positivo. Isso indica que talvez as alternativas apresentadas devam ser reelaboradas, pois a alternativa D atraiu indivíduos com habilidade suficiente para resolver a questão (coeficiente próximo de zero), talvez pela questão inferir a frase “*para diminuir custos*”.

Pela TRI, verifica-se que o parâmetro a tem índice de discriminação alto observado pela curva ingreme do gráfico. O parâmetro b demonstra que a questão é graduada como muito difícil, o que claramente transparece na Figura 22, (gráfico da curva característica do item). Vemos que de -4 a 1 a curva não se altera, ou seja, o índice de probabilidade de acerto é praticamente o mesmo para indivíduos com habilidade inferior um, apenas com variações para habilidades acima desse valor, sendo a CCI e o gráfico informação do item totalmente deslocado para direita. O gráfico traz informação de que os respondentes estão em intervalo de habilidades muito altas, superiores a um.

Logo, por ser um item considerado muito difícil, apenas examinados com habilidades altas possuem probabilidades razoáveis de acertá-lo. A probabilidade de indivíduo com proficiência dois acertar o item gira em torno de 30%, o que indica que mesmo com uma proficiência alta a probabilidade de acerto é baixa, haja visto que dos 206 alunos que responderam esse item, apenas 19 responderam corretamente, corroborando que o índice é bem discriminativo.

Tabela de Índices e Gráfico - ENEM - do Item H30

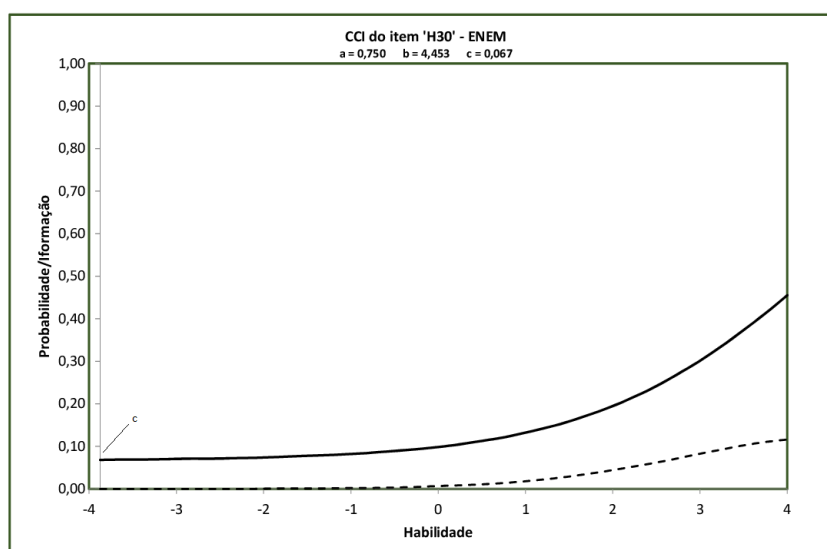
Tabela 16 – Tabela de Índices da Habilidade H30 - ENEM

		TCT			TRI		
Gabarito	A	Opções	Frequência	Bisserial	Parâmetro	Índice	Classificação
TOTAL	12347	A	0,106	-0,012	a	0,750	Moderada
T.A	1336	B	0,192	0,153			
P.A	10,66%	C	0,213	0,766	b	4,453	Muito Difícil
DISC	-0,086	D	0,345	0,546			
DIFI	0,106	E	0,141	0,308	c	0,067	

T.A.= Total Acerto P.A. = Percentual Acerto DISC = Discriminação DIFI = Dificuldade

Fonte: Autor

Figura 23 – Curva Característica do Item H30 - ENEM



Fonte: Autor

Comentário:

Verifica-se pela TCT o baixo índice de acerto da questão, em torno de 10,66% e podemos ver que há problemas com os coeficiente bisseriais, pois para a opção correta, gabarito A, o coeficiente é negativo, sendo positivo para todas as outras opções, quando o ideal deveria ser o contrário. Observa-se que as outras alternativas atraíram respondentes com habilidades para responder a questão, especialmente a opção D.

Pela TRI, verifica-se que o parâmetro a tem índice de discriminação moderada e o parâmetro b demonstra que a questão é graduada como muito difícil. A curva característica, figura 23, mostra que independente da habilidade do avaliado, a probabilidade de acertar o item não sofre um aumento proporcional importante. O gráfico de informação baixo e totalmente deslocado para a direita não traz informações relevantes.

Logo, mesmo sendo um item de discriminação moderada, verificamos que entre -4 a aproximadamente 1, a probabilidade de acertar o item praticamente permanece inalterada, ficando próxima do parâmetro a (acerto ao acaso), havendo pouca alteração apenas nas proficiências altas. A probabilidade tem um aumento importante com os respondentes com habilidades acima de dois. Por ser muito difícil, dos 12.347 respondentes deste item, apenas 1.336 responderam corretamente.

4 Considerações Finais

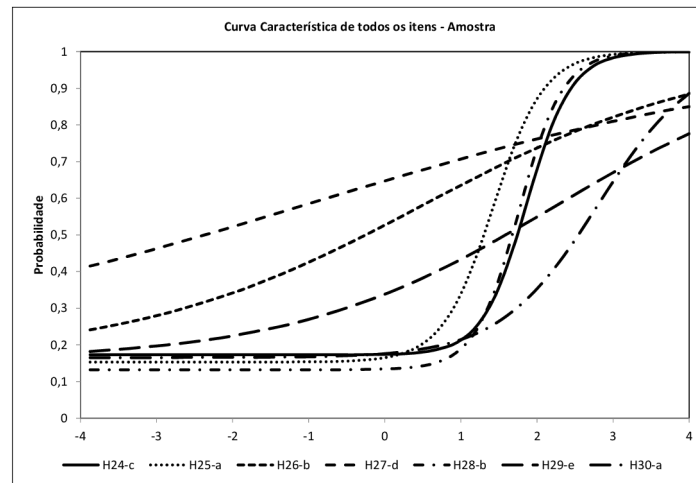
Após a formatação dos dados recolhidos das provas aplicadas nas Escolas das cidades de Campo Grande, Bonito e Dourados em Mato Grosso do Sul e no ENEM (Arquivo Microdados), foi realizada a dicotomização dos itens (aplicando “um” para as respostas certas e “zero” para as respostas incorretas). Utilizando o programa *Microsoft Excel - eirt* foram elaboradas as tabelas, anexos A, B e C nas quais constam dados da Teoria Clássica dos Testes - TCT e da Teoria de Resposta ao Item - TRI.

Pela teoria TCT, uma tabela contendo dados, como o número de sujeitos, número de itens, média da pontuação e o alfa de Cronbach dos itens dos testes. Outra tabela com as proporções de acertos, a dificuldade média, o desvio padrão e a correlação polisserial (discriminação) de cada item. O alfa de Cronbach baixo para as duas amostras, 0,154 para a amostra aplicada nas escolas e 0,024 para o ENEM, pode ser devido a quantidade muito baixa de itens, pois o número de questões afeta o valor de alfa. Testes com muitos itens aumentam o valor de alfa, sem que isso signifique aumento de consistência interna; um valor baixo de alfa pode significar apenas número pequeno de itens, o que certamente é o caso em questão.

Podemos notar que, no caso da amostra aplicada nas escolas, pela TCT, o item considerado mais fácil pelos examinados foi o H27 e o mais difícil o H28 enquanto que a amostra retirada do ENEM o considerado mais fácil foi H27 seguido com proporções próximas por H28, mostrando assim paridade entre os respondentes. O item menos discriminativo, na amostra aplicada, foi o H30, enquanto que pela amostra retirada do ENEM foi o item H25 e H30, mostrando, novamente, a similaridade entre os itens.

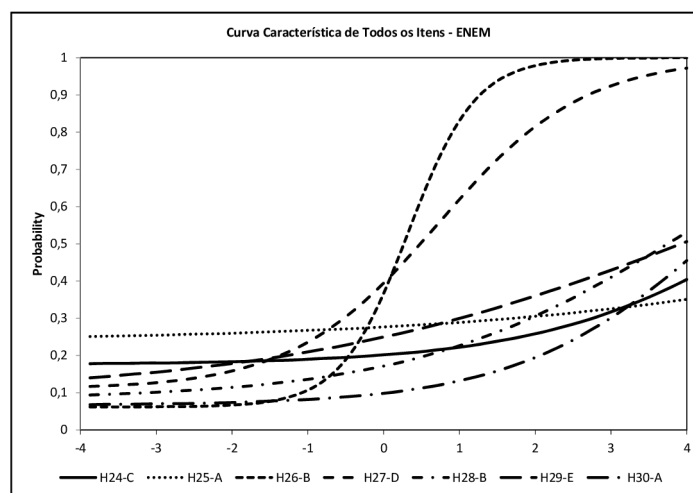
Analisando pela teoria de resposta ao item - TRI - vemos que a curva característica dos itens - CCI - mostra grande diferença entre os dois testes. Nas CCIs da amostra verifica-se que cinco, dos sete itens, foram significantes para o teste, H24, H25, H26, H28 e H30, enquanto que no teste aplicado no ENEM apenas dois dos sete itens foram significativos, H26 e H27, como mostram as figuras 24 e 25. Isso repercutiu na curva característica do teste - CCT, figura 26.

Figura 24 – CCI de todos os Itens - Amostra



Fonte: Autor

Figura 25 – CCI de todos os Itens - ENEM

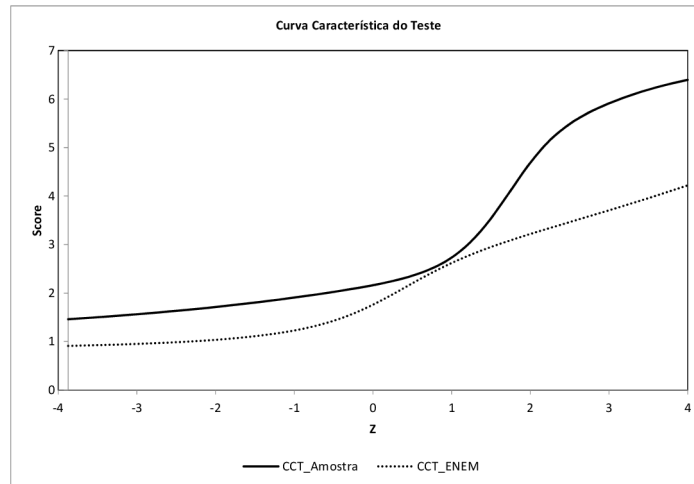


Fonte: Autor

A curva CCT, segundo Pasquali (2003), representa a probabilidade de obtenção de um determinado *score* total em função de θ e é dada pela soma das CCIs de todos os itens. Na figura 26 podemos ver a CCT dos testes objetos desse estudo. Verificamos que a amostra aplicada apresentou valores melhores para os respondentes com proficiência acima de 1, enquanto que para os examinados do ENEM a curva é mais suave e começa com leves alterações em habilidades a partir de -1. Percebe-se, nos dois testes, que há uma proximidade nos escores dos examinados com proficiência -4 até 0 e que não há diferença entre os escores dos examinados dos dois testes para aqueles que tem habilidade 0,952, havendo apenas um distanciamento para as habilidades acima de 1. Isso mostra, que os respondentes da amostra obtiveram melhores e maiores resultados, aqueles com habilidade

acima de 1, enquanto que os respondentes do ENEM não houveram grandes alterações a medida que aumentava a habilidade do examinado.

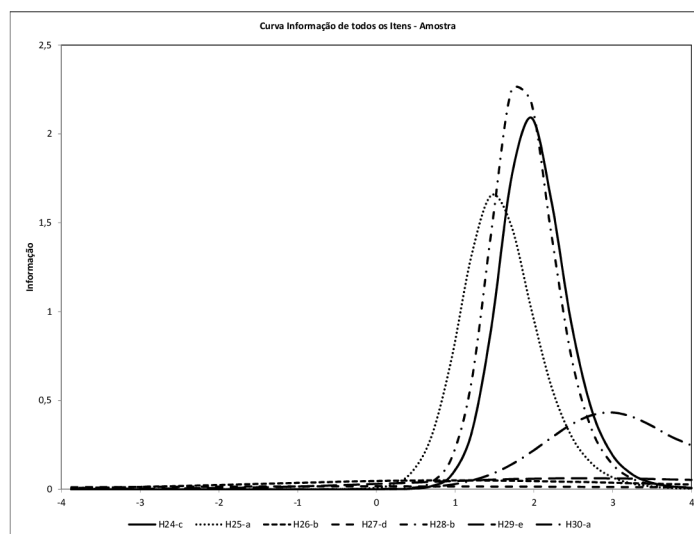
Figura 26 – Curva Característica do Teste



Fonte: Autor

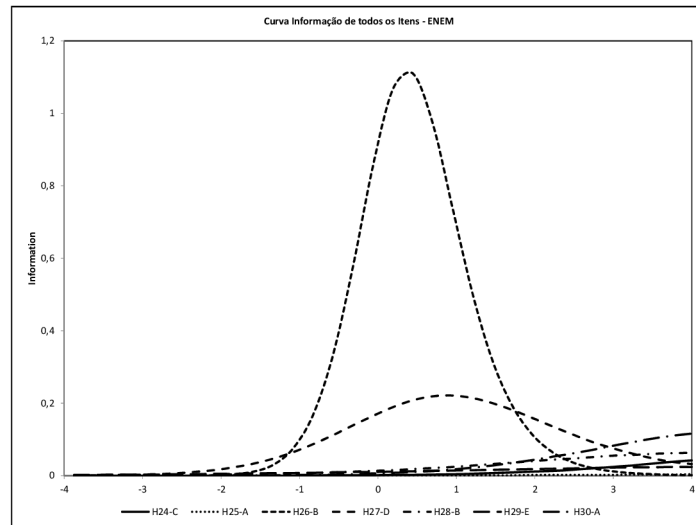
A mesma repercussão se deu no gráfico de informação do teste. Esse gráfico é obtido pela soma da informação de cada item e a precisão da estimativa é dada pelo inverso da informação. A curva de informação do teste especifica para que faixa de θ ele é especialmente válido. As figuras 27 e 28 mostram os gráficos de informação dos itens, e da mesma forma, quatro dos sete itens mostram informações relevantes na amostra aplicada, H24, H25, H28 e H30, enquanto no ENEM apenas dois dos sete itens mostraram informações significativas, H26 e H27.

Figura 27 – CII de todos os Itens - Amostra



Fonte: Autor

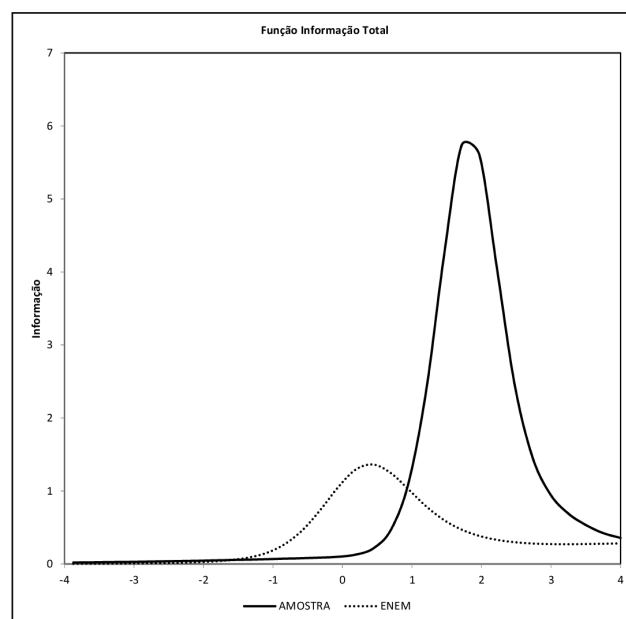
Figura 28 – CII de todos os Itens - ENEM



Fonte: Autor

Pode-se ver diferença enorme entre os dois gráficos de informações dos testes concatenados na figura 29, onde, enquanto o gráfico da amostra fornece muita informação para os valores latentes no intervalo $[1, 3]$, com inflexão no ponto $(1,714, 5,737)$, o gráfico do teste aplicado no ENEM fornece maiores informações para valores latentes no intervalo $[-1, 1]$, sendo a inflexão no ponto $(0,444, 1,361)$, porém nesse caso, o erro padrão supera a informação. Verifica-se claramente, então, que o gráfico da amostra é mais informativo do que o gráfico do ENEM.

Figura 29 – Função Informação do Teste



Fonte: Autor

4.1 Conclusão

Diante do exposto, podemos intuir que os alunos que responderam a amostra aplicada estavam melhores preparados que os examinados que responderam esses mesmos itens no ENEM de anos anteriores. Isso responde a primeira indagação realizada no início desse trabalho, ou seja, “SIM”, os alunos que responderam os testes na amostra aplicada estavam melhores preparados para responder as questões de probabilidade e estatística que os alunos que responderam as mesmas questões no exame nacional do ensino médio. Essa afirmação nos leva a resposta da segunda indagação, ou seja, “SIM”, o ensino de probabilidade e estatística está sendo desenvolvido de forma que os alunos sejam capazes de responder as questões que envolvam esse conteúdo nas provas do exame nacional do ensino médio.

Apesar do resultado positivo na realização desse teste, não podemos deixar de observar que esses mesmos alunos, ao responderem as questões, tiveram grandes dificuldades em itens que envolviam conteúdos básicos, tais como, não conseguir distinguir o conjunto universo em probabilidade, calcular taxa de porcentagem e até mesmo a leitura errada de uma tabela. Como já descrito nesse trabalho, é necessário que, para um melhor desenvolvimento dessa disciplina nas escolas públicas, a prática seja incluída no cotidiano escolar, para que o ensino da matemática e, especificamente de probabilidade e estatística, deixe de ser exclusivamente conteudista, teórica. A prática, aliada a teoria, permite que o aluno aprenda a desenvolver, analisar, interpretar e inferir sobre o conteúdo estudado. Portanto, com a implantação da nova Base Nacional Comum Curricular - BNCC, espera-se que o ensino de probabilidade e estatística tenha um grande salto de qualidade, haja visto, que será aplicado nas grades curriculares dos ensinos, desde as séries iniciais até o final do ensino médio.

Diante da nova BNCC, tanto do ensino fundamental quanto do ensino médio, faz-se necessário que, no futuro, trabalhos semelhantes a este sejam elaborados para que assim, realize-se novas comparações frente ao ensino de probabilidade e estatística. O microdados do ENEM, publicado anualmente, fornece inúmeras informações as quais podem ser utilizadas de forma mais detalhadas. Trabalhos que contemplem amostras com maiores quantidades de alunos nas escolas e também um maior refinamento na coleta de dados, como tabular por idade, renda ou exclusivamente com alunos oriundos de escolas particulares, pode permitir um olhar mais afinado sobre o ensino da matemática e especificamente de probabilidade e estatística e ainda, com maior importância, se houve evolução ou involução com a aplicação da nova BNCC.

Referências

- ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. *Teoria da Resposta ao Item: conceitos e aplicações*. [s.n.], 2000. Disponível em: <<http://egov.ufsc.br/portal/sites/default/files/livrotri.pdf>>. Acesso em: jan/2018. Citado 14 vezes nas páginas 24, 26, 28, 31, 33, 34, 36, 40, 41, 42, 43, 44, 45 e 47.
- ANDRADE, J. M. d.; LAROS, J. A.; GOUVEIA, V. V. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. *Avaliação Psicológica*, v. 9, n. 3, p. 421–435, 2010. Citado na página 29.
- BNCC. *Base Nacional Comum Curricular - Ensino Fundamental*. 2017. Ministério da Educação. Disponível em: <http://basenacionalcomum.mec.gov.br/images/pdf/4.2_BNCC-Final_MA.pdf>. Acesso em: 03 out. 2017. Citado 2 vezes nas páginas 17 e 18.
- BNCC. *Base Nacional Comum Curricular - Ensino Médio: Educação é a base*. 2018. Ministério da Educação. Disponível em: <http://basenacionalcomum.mec.gov.br/wp-content/uploads/2018/04/BNCC_EnsinoMedio_embaixa_site.pdf>. Acesso em: 09 jun. 2018. Citado na página 12.
- BRASIL. *Entenda sua nota do ENEM - Guia do Participante*. [s.n.], 2012. Ministério da Educação - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Disponível em: <http://download.inep.gov.br/educacao_basica/enem/manuais/2012/guia-do-participante-Notas.pdf>. Acesso em: 06 fev. 2018. Citado 4 vezes nas páginas 28, 29, 30 e 31.
- BRASIL. *Relatório pedagógico: Enem 2011-2012*. Brasília - DF: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2015. Acesso em: 02 out 2017. Citado 3 vezes nas páginas 20, 21 e 22.
- CAVALCANTE, J. L.; ANDRADE, V. L. V. X. de; REGNIER, J.-C. O conceito de probabilidade na formação docente: uma reflexão apoiada pela análise estatística implicativa. *VIDYA*, v. 36, n. 2, p. 441–455, 2016. Citado na página 18.
- COSTA, S. T. S. Teoria de resposta ao item aplicada no enem. Universidade Federal de Goiás, 2017. Citado 2 vezes nas páginas 25 e 34.
- CRESPO, A. A. *Estatística fácil*. [S.l.]: Saraiva, 1993. Citado na página 17.
- D'AMBROSIO, U. *Educação matemática: Da teoria à prática*. Papirus. Campinas - SP: [s.n.], 1996. Citado 2 vezes nas páginas 15 e 16.
- INEP. *Micro dados do Enem 2015*. 2015. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Disponível em: <<http://portal.inep.gov.br/web/guest/microdados>>. Acesso em: 30 set. 2017. Citado na página 19.
- KLEIN, R. Alguns aspectos da teoria de resposta ao item relativos à estimação das proficiências. *Ensaio: Avaliação e Políticas Públicas em Educação*, SciELO Brasil, v. 21, n. 78, p. 35–56, 2013. Disponível em: <http://www.scielo.br/pdf/ensaio/2013nahead/aop_0213.pdf>. Citado 3 vezes nas páginas 21, 22 e 23.

- LOPES, C. E. Educação estatística na escola básica e suas interfaces com a educação matemática, a cultura e a diversidade. 2010. Disponível em: <www.lematec.net.br/CDS/ENEM10/artigos/MR/MR5_Lopes.pdf>. Acesso em: 20 nov. 2017. Citado 2 vezes nas páginas 15 e 16.
- LORENZATO, S. *O Laboratório de Ensino de Matemática na Formação de Professores: Coleção formação de professores*. Autores associados. Campinas - SP: [s.n.], 2006. Citado na página 11.
- NOGUEIRA, P. A.; VICTER, E. das F.; NOVIKOFF, C. Roteiro didático para o ensino da estatística: A cidadania na/pela matemática. 2011. Disponível em: <http://www.pucrs.br/famat/viali/tic_literatura/relatorios/produto-paulo-apolinario.pdf>. Acesso em: 03 out. 2017. Citado na página 18.
- PASQUALI, L. *Psicometria: teoria dos testes na psicologia e na educação*. [S.l.]: Editora Vozes Limitada, 2003. Citado na página 76.
- RODRIGUES, M. R.; MARTINS, E. G. A abordagem do tema probabilidade nos livros aprovados pelo pnld para o triênio 2015–2017 e suas implicações no processo de ensino e aprendizagem. 2016. Disponível em: <http://www.sbemrasil.org.br/enem2016/anais/pdf/5327_3409_ID.pdf>. Acesso em: 05 dez. 2017. Citado na página 18.
- SARTES, L. M. A.; SOUZA-FORMIGONI, M. L. O. Avanços na psicometria: da teoria clássica dos testes à teoria de resposta ao item. *Psicologia: Reflexão e Crítica*, Universidade Federal do Rio Grande do Sul, v. 26, n. 2, 2013. Citado 2 vezes nas páginas 27 e 29.
- SOUZA, A. C.; LOPES, C. E. A construção de ideias estatística na educação infantil. In: *16º Congresso de Leituras do Brasil, Campinas/SP*. [S.l.: s.n.], 2007. Citado na página 17.
- TAVARES, C. M. *A Teoria De Resposta ao Item na Avaliação em Larga Escala: Um estudo sobre o exame nacional de acesso ao Mestrado Profissional em Matemática em Rede Nacional - PROFMAT*. Dissertação (Mestrado), 2014. Disponível em: <https://sca.profmt-sbm.org.br/sca_v2/get_tcc3.php?id=298>. Acesso em: fev/2018. Citado 3 vezes nas páginas 23, 29 e 30.
- VENCESLAU, A. W. N. *Curvas parametrizadas, ciclóides, experimentos e aplicações*. Universidade Federal de Sergipe, 2015. Citado na página 11.
- VENDRAMINI, C. M. M. *Contribuições da Educação Estatística para a Educação Matemática.: Solução de problemas e a matemática escolar*. Alínea. Campinas - SP: Márcia Regina Ferreira de Brito, 2006. Citado 2 vezes nas páginas 15 e 17.

Apêndices

APÊNDICE A – Amostra

Classical test theory statistics

Number of subject	Number of item	Number of missing response	Score's mean	Score's standard deviation	Cronbach's Alpha
206	7	0	2,354	1,252	0,154

Fonte: Autor

ITEM	MEAN (DIFFICULTY)	STANDARD DEVIATION	CORRELATION (DISCRIMINATION)
H24-c	0,214	0,410	0,039
H25-a	0,243	0,429	0,094
H26-b	0,529	0,499	0,126
H27-d	0,646	0,478	0,006
H28-b	0,180	0,384	0,038
H29-e	0,350	0,477	0,136
H30-a	0,194	0,396	-0,035

Fonte: Autor

Estimation summary

Data source Range \$A\$2:\$G\$207 in sheet Planilha1 of file Amostra.xlsx

Method: Bayes modal estimator (BME)

Model: Three parameters logistic model (3PLM)

Latent variables estimator: EAP

Version 1.3.1

No item were ignored by the estimation process.

Test of fit			
Item	Chi-square	Degrees of freedom	P-Value
H24-c	0,309	10	1,000
H25-a	0,251	10	1,000
H26-b	0,511	10	1,000
H27-d	1,836	10	0,997
H28-b	0,225	10	1,000
H29-e	0,750	10	1,000
H30-a	0,216	10	1,000
Global	4,097	70	1,000

Fonte: Autor

Test of local independance				
Item	Item	Chi-square	Degrees of freedom	P-Value
H24-c	H25-a	0,430	1	0,512
H24-c	H26-b	0,400	1	0,527
H24-c	H27-d	1,363	1	0,243
H24-c	H28-b	0,030	1	0,863
H24-c	H29-e	1,036	1	0,309
H24-c	H30-a	0,247	1	0,619
H25-a	H26-b	1,205	1	0,272
H25-a	H27-d	0,446	1	0,504
H25-a	H28-b	0,112	1	0,738
H25-a	H29-e	0,148	1	0,701
H25-a	H30-a	0,059	1	0,809
H26-b	H27-d	3,466	1	0,063
H26-b	H28-b	0,298	1	0,585
H26-b	H29-e	2,347	1	0,126
H26-b	H30-a	0,119	1	0,730
H27-d	H28-b	6,810	1	0,009
H27-d	H29-e	2,931	1	0,087
H27-d	H30-a	3,701	1	0,054
H28-b	H29-e	0,255	1	0,613
H28-b	H30-a	1,393	1	0,238
H29-e	H30-a	1,017	1	0,313

Fonte: Autor

Figura 30 – Parametros com Erro Padrão

Parameter estimates - Amostra							
Item	Slope (a)	s.e.	Threshold (b)	s.e.	Item	Asymptote (c)	s.e.
H24-c	3,419	0,938	1,867	0,192	H24-c	0,173	0,000
H25-a	2,988	0,740	1,426	0,145	H25-a	0,153	0,000
H26-b	0,524	0,190	0,513	0,407	H26-b	0,165	0,039
H27-d	0,303	0,170	-0,998	0,914	H27-d	0,170	0,061
H28-b	3,474	0,896	1,764	0,167	H28-b	0,132	0,000
H29-e	0,580	0,248	2,248	0,861	H29-e	0,159	0,018
H30-a	1,540	0,768	2,804	0,724	H30-a	0,165	0,002

APÊNDICE B – ENEM

Classical test theory statistics

Number of subject	Number of item	Number of missing response	Score's mean	Score's standard deviation	Cronbach's Alpha
12347	7	0	1,840	1,137	0,024

Fonte: Autor

ITEM	MEAN (DIFFICULTY)	STANDARD DEVIATION	CORRELATION (DISCRIMINATION)
H24-C	0,206	0,405	0,032
H25-A	0,278	0,448	-0,118
H26-B	0,432	0,495	0,169
H27-D	0,419	0,493	0,000
H28-B	0,181	0,385	0,051
H29-E	0,219	0,414	0,004
H30-A	0,106	0,307	-0,086

Fonte: Autor

Estimation summary

Data source Range \$A\$2:\$G\$12348 in sheet Planilha1 of file ENEM_AM_pronta.xlsx

Method: Bayes modal estimator (BME)

Model: Three parameters logistic model (3PLM)

Latent variables estimator: EAP

Version 1.3.1

No item were ignored by the estimation process.

Test of fit			
Item	Chi-square	Degrees of freedom	P-Value
H24-C	0,015	10	1,000
H25-A	93,374	10	0,000
H26-B	0,309	10	1,000
H27-D	0,294	10	1,000
H28-B	0,698	10	1,000
H29-E	85,126	10	0,000
H30-A	59,914	10	0,000
Global	239,730	70	0,000

Fonte: Autor

Test of local independence				
Item	Item	Chi-square	Degrees of freedom	P-Value
H24-C	H25-A	1,406	1	0,236
H24-C	H26-B	0,592	1	0,442
H24-C	H27-D	0,494	1	0,482
H24-C	H28-B	2,440	1	0,118
H24-C	H29-E	82,804	1	0,000
H24-C	H30-A	0,726	1	0,394
H25-A	H26-B	22,139	1	0,000
H25-A	H27-D	1055,372	1	0,000
H25-A	H28-B	0,418	1	0,518
H25-A	H29-E	87,193	1	0,000
H25-A	H30-A	0,880	1	0,348
H26-B	H27-D	3,198	1	0,074
H26-B	H28-B	10,868	1	0,001
H26-B	H29-E	89,183	1	0,000
H26-B	H30-A	70,506	1	0,000
H27-D	H28-B	36,473	1	0,000
H27-D	H29-E	84,698	1	0,000
H27-D	H30-A	66,958	1	0,000
H28-B	H29-E	84,681	1	0,000
H28-B	H30-A	8,118	1	0,004
H29-E	H30-A	551,513	1	0,000

Fonte: Autor

Parameter estimates

Item	ENEM				Amostra			
	Slope(a)	Threshold(b)	Asymptote(c)	Difficulty	Slope(a)	Threshold(b)	Asymptote(c)	Difficulty
H24-C	0,617	5,556	0,176	Muito Difícil	3,419	1,867	0,173	Muito Difícil
H25-A	0,297	9,895	0,238	Muito Difícil	2,988	1,426	0,153	Muito Difícil
H26-B	2,246	0,326	0,062	Mediana	0,524	0,513	0,165	Mediana
H27-D	1,045	0,721	0,109	Difícil	0,303	-0,998	0,170	Fácil
H28-B	0,546	4,078	0,082	Muito Difícil	3,474	1,764	0,132	Muito Difícil
H29-E	0,348	4,527	0,094	Muito Difícil	0,580	2,248	0,159	Muito Difícil
H30-A	0,750	4,453	0,067	Muito Difícil	1,540	2,804	0,165	Muito Difícil

Fonte: Autor

Figura 31 – Parametros com Erro Padrão

Parameter estimates - ENEM						
Item	Slope (a)	s.e.	Threshold (b)	s.e.	Asymptote (c)	s.e.
H24-C	0,617	0,120	5,556	0,913	0,176	0,001
H25-A	0,297	0,112	9,895	3,592	0,238	0,001
H26-B	2,246	0,049	0,326	0,012	0,062	0,000
H27-D	1,045	0,031	0,721	0,026	0,109	0,002
H28-B	0,546	0,043	4,078	0,281	0,082	0,001
H29-E	0,348	0,042	4,527	0,600	0,094	0,002
H30-A	0,750	0,079	4,453	0,386	0,067	0,000