



**PUC Minas**

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS**

**Programa de Mestrado em Ensino de Ciências e Matemática**

**Rogério Rodrigues**

**APLICAÇÃO DA TEORIA DA RESPOSTA AO ITEM NA AVALIAÇÃO DAS  
HABILIDADES MATEMÁTICAS EM ALUNOS DO ENSINO MÉDIO DE UMA  
ESCOLA DE BELO HORIZONTE**

Belo Horizonte  
2016

Rogério Rodrigues

**APLICAÇÃO DA TEORIA DA RESPOSTA AO ITEM NA AVALIAÇÃO DAS  
HABILIDADES MATEMÁTICAS EM ALUNOS DO ENSINO MÉDIO DE UMA  
ESCOLA DE BELO HORIZONTE**

Dissertação apresentada ao Programa de Pós-Graduação em Ensino de Ciências e Matemática da Pontifícia Universidade Católica de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Ensino de Ciências e Matemática.

**Orientadora:** Profa. Dra. Tânia Fernandes Bogutchi

**Co-orientadora:** Profa. Dra. Eliane Scheid Gazire

Belo Horizonte  
2016

FICHA CATALOGRÁFICA

Elaborada pela Biblioteca da Pontifícia Universidade Católica de Minas Gerais

R696a Rodrigues, Rogério  
Aplicação da teoria da resposta ao item na avaliação das habilidades matemáticas em alunos do ensino médio de uma escola de Belo Horizonte / Rogério Rodrigues. Belo Horizonte, 2016.  
153 f. : il.

Orientadora: Tânia Fernandes Bogutchi

Coorientadora: Eliane Scheid Gazire

Dissertação (Mestrado) – Pontifícia Universidade Católica de Minas Gerais.  
Programa de Pós-Graduação em Ensino de Ciências e Matemática.

1. Avaliação educacional - Métodos estatísticos. 2. Rendimento escolar. 3. Matemática (Ensino médio) – Teste. I. Bogutchi, Tânia Fernandes. II. Gazire, Eliane Scheid. III. Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Ensino de Ciências e Matemática. IV. Título.

SIB PUC MINAS

CDU: 519.2



PUC Minas

**PROGRAMA DE MESTRADO EM ENSINO DE CIÊNCIAS E MATEMÁTICA**

**FOLHA DE APROVAÇÃO**

**ROGÉRIO RODRIGUES**

Dissertação defendida e aprovada pela seguinte banca examinadora:

Prof.<sup>a</sup> Dr.<sup>a</sup> Tânia Fernandes Bogutchi – Orientadora – (PUC Minas)  
Doutorado em Tratamento da Informação Espacial – (PUC Minas)

Prof.<sup>a</sup> Dr.<sup>a</sup> Eliane Scheid Gazire – Coorientadora – (PUC Minas)  
Doutorado em Educação – (UNICAMP)

Prof. Dr. Leônidas Conceição Barroso – (FUMEC)  
Doutorado em Informática – (PUC RIO)

Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Inês Martins – (PUC Minas)  
Doutorado em Educação – (UNICAMP)

Belo Horizonte, 26 de fevereiro de 2016.

## AGRADECIMENTOS

*Eu vi um menino correndo, eu vi o tempo brincando  
ao redor do caminho daquele menino...  
(Caetano Veloso)*

Na minha feliz infância, no interior dessa Gerais abençoada, meu maior desejo era ser alfabetizado, pois sentia que, com essa ferramenta, no domínio das letras, eu poderia aprender tudo. Minha mãe e minha avó materna, que eram analfabetas, tentavam inserir, nas minhas brincadeiras, as poucas letras que sabiam desenhar, porquanto eu ainda não tinha idade legal para ser matriculado na escola formal. Minha avó não viveu o suficiente para me ver de beca, mas é a ela o meu primeiro agradecimento; D. Ludovina não sabia ler, mas era de uma sabedoria ímpar. Minha mãe quase enfartou de alegria e emoção quando terminei a minha primeira graduação, mas o tempo não nos favoreceu para alinhar, na realidade, nossa aspiração comum, que era a minha pós-graduação. Tive que, por anos, priorizar o trabalho para educar meus filhos, que hoje são adultos e pessoas de bem. Quando consegui o tempo para fazer a pós, minha mãe já tinha partido, deixando um rico legado moral para mim e minha irmã, mas deixando, também, o vazio de sua ausência. D. Maria Inêz não sabia as letras, mas era de uma inteligência invejável; a ela o meu agradecimento.

Por sua vez, meu pai, que há poucos anos também partiu, dedicou todo o seu fatigante trabalho à educação dos filhos. Seu orçamento priorizava as despesas com nossa educação, pelo menos até o segundo grau. Em seu pouco tempo de escola, ele aprendeu muito bem a ler, escrever e fazer as quatro operações matemáticas; tinha uma caligrafia lindíssima. Lembro-me de que meu pai levava do trabalho aqueles livros contábeis inutilizados e nos mobilizava para copiar letras de músicas da época nesses livros; essa era a sua maneira de nos ajudar na escola. Meu sincero agradecimento a meu velho que, do céu, ainda torce por mim.

Hoje, que o tempo entrou no meu caminho, depois de tanto brincar ao redor dele, eu tenho outras tantas pessoas a agradecer, sendo esses os catalisadores da realização dessa aspiração de fazer o Mestrado.

Minha esposa, Rosana, parceira solidária, que sempre me apoiou e até abriu mão de prerrogativas nossas de tempo, para que eu pudesse me dedicar inteiramente à pesquisa. Além disso, foi minha revisora contínua.

Minha gratidão aos meus amigos Marcinha, Mel e Juliano, que só fizeram me estimular cada vez mais.

Há também o orgulho e o entusiasmo de meu sogro, Seu Paulinho, minha sogra, D. Ivone, e minha cunhada Inês, que sempre me colocaram para frente.

Agradeço a oportunidade que essa renomada instituição de ensino me proporcionou e, agradeço, principalmente e em caráter prioritário, aos mestres desse corpo docente de alto ideário e de competência notável; de todos eles, levo um aprendizado de alto nível. Minha gratidão especial às minhas orientadoras neste trabalho, professoras e doutoras Tânia Bogutchi e Eliane Scheid Gazire.

Enfim, minha gratidão maior a Deus, causa primária da existência, que me fez de barro e colocou todos os outros no meu caminho, tudo ao seu tempo, inclusive o tempo para que eu me tornasse, o quanto possível, de ouro.

A todos que em mim acreditaram, me incentivaram e, de algum modo, em mim investiram. A todos que por mim torceram. Ao Senhor de todas as esferas, que, para mim, abriu tanto segredos de Sua criação e, com certeza, muitos outros há de abrir.

## RESUMO

O presente trabalho procura experimentar uma técnica de estatística no âmbito da avaliação escolar. Trata-se da Teoria de Resposta ao Item - TRI, que, apesar da limitação de seus pressupostos, representa um salto quântico em relação à Teoria Clássica de Testes - TCT, no que tange à psicometria. As lacunas que esse salto procura preencher são relativas tanto às limitações, como a dependência do instrumento de medição em relação ao objeto que ele mede. Outros problemas resolvidos pela TRI dizem respeito aos parâmetros de dificuldade e de discriminação dos testes. O primeiro problema pontua que a dificuldade dependerá da amostra de indivíduos, quanto mais capazes, mais fáceis são os testes e o segundo pontua que o indivíduo avaliado não será classificado segundo a sua real capacidade, pode sair um resultado casual. Como na TCT o mesmo grupo pode apresentar desempenhos distintos em avaliações distintas, vem a questão: Qual das avaliações é mais confiável? O objetivo dessa pesquisa é verificar qual dos atributos da TRI lhe conferem a segurança de proficiência na avaliação escolar. Para tanto, foi feito um embasamento bibliográfico no acervo disponível e posteriormente uma aplicação dessas técnicas em uma amostra de alunos do ensino médio de uma escola particular de Belo Horizonte – MG. Esse trabalho gerou um aplicativo que consiste em um banco de itens calibrados para avaliação em matemática. Esse aplicativo insere o item, avalia sua qualidade segundo a TCT e o prepara para a calibragem pela TRI.

**Palavras-chave:** TRI, desempenho escolar, avaliação, proficiência, TCT.

## ABSTRACT

The present work seeks to try a statistical technique in the context of school evaluation. This is the Item Response Theory - IRT, that, despite the limitation of its presuppositions, represents a quantum leap in relation to the classical theory of tests - CTT, regarding the psychometrics. The gaps that this leap tries to fill are relating both to limitations, such as the dependency of the measuring instrument in relation to the object that it measures. The problems solved by IRT relate to parameters of difficulty and discrimination of the tests. The first problem scores that the difficulty will depend on the sample of individuals, the more capable, more easy are the tests and the second scores that the individual evaluated will not be classified according to their real capacity, can leave a chance result. s in the CTT the same group may present distinct performances in separate assessments, comes the question: Which of evaluations is more reliable? The objective of this research is to verify which of the attributes of IRT confer the safety of proficiency in the school evaluation. For both, was done a bibliographic foundations in the collection available and subsequently a application of these techniques in a sample of students from a private school of Belo Horizonte - MG. This work has generated an application that consists of a database items calibrated for evaluation in mathematics. This application inserts the item, assesses its quality according the CTT and prepares for the calibration by IRT.

**Keywords:** IRT, school performance, evaluation, proficiency, CTT

## LISTA DE ILUSTRAÇÕES

|  |           |
|--|-----------|
| <b>FIGURA 1 – Exemplo de CCI (Modelo com 1 parâmetro) .....</b>    | <b>26</b> |
| <b>FIGURA 2 – Exemplo de CCI (Modelo de 2 parâmetros).....</b>     | <b>27</b> |
| <b>FIGURA 3 – Exemplo de CCI (Modelo de 3 parâmetros).....</b>     | <b>28</b> |
| <b>FIGURA 4 - Formas de aplicação de testes.....</b>               | <b>30</b> |
| <b>FIGURA 5 – Exemplo de item.....</b>                             | <b>38</b> |
| <b>FIGURA 6 - Exemplo de Funções de Informação .....</b>           | <b>49</b> |
| <b>FIGURA 7 - CCI e curva de informação do item 11 .....</b>       | <b>67</b> |
| <b>FIGURA 8 - CCI e curva de informação do item 13.....</b>        | <b>67</b> |
| <b>FIGURA 9 - CCI e curva de informação do item 2.....</b>         | <b>68</b> |
| <b>FIGURA 10 - CCI e função de informação do item 15 .....</b>     | <b>68</b> |
| <b>FIGURA 11 - Número de acertos x habilidade (dispersão).....</b> | <b>74</b> |
| <b>FIGURA 12 - Número de acertos x habilidade (tendência).....</b> | <b>74</b> |

## LISTA DE QUADROS

|   |           |
|---|-----------|
| <b>QUADRO 1 – Competência para área 1 de Matemática e suas tecnologias.....</b> | <b>40</b> |
| <b>QUADRO 2 - Habilidades dos itens.....</b>                                    | <b>53</b> |
| <b>QUADRO 3 – Quesitos avaliados pelo filtro.....</b>                           | <b>76</b> |

## LISTA DE TABELAS

|   |    |
|---|----|
| TABELA 1 - Amostra para o pré-teste.....                      | 51 |
| TABELA 2 – Amostra para teste .....                           | 52 |
| TABELA 3 - Gabarito e acertantes.....                         | 55 |
| TABELA 4 - Índice de dificuldade por item .....               | 56 |
| TABELA 5 - Distribuição das respostas por alternativa .....   | 56 |
| TABELA 6 - Classificação do pré-teste .....                   | 57 |
| TABELA 7 - Distribuição dos escores .....                     | 58 |
| TABELA 8- Distribuição normal dos escores .....               | 58 |
| TABELA 9 - Cálculo da correlação de normalidade .....         | 59 |
| TABELA 10 - Percentual de cada alternativa por item.....      | 60 |
| TABELA 11 - Correlação bisserial .....                        | 62 |
| TABELA 12 - Fatores do coeficiente de Pearson.....            | 63 |
| TABELA 13 - Coeficientes de Pearson (influência do item)..... | 63 |
| TABELA 14 - Escore global e média de escores .....            | 64 |
| TABELA 15 - Parâmetros dos itens.....                         | 66 |
| TABELA 16 - Escore bruto da amostra no teste .....            | 69 |
| TABELA 17 - Distribuição dos escores brutos .....             | 70 |
| TABELA 18 - Gabarito e respostas dos alunos .....             | 71 |
| TABELA19 - Classificação dos alunos por habilidade .....      | 73 |

## SUMÁRIO

|   |           |
|---|-----------|
| <b>1 INTRODUÇÃO .....</b>   | <b>14</b> |
| <b>1.1 O Trabalho de Pesquisa .....</b>   | <b>16</b> |
| <b>1.2 Metodologia.....</b>   | <b>17</b> |
| <br>  |           |
| <b>2 REVISÃO DE LITERATURA.....</b>   | <b>18</b> |
| <b>2.1 A Teoria Clássica dos Testes (TCT).....</b>                                      | <b>18</b> |
| <i>2.1.1 O Índice de Dificuldade de um item (<math>I_d</math>) .....</i>                | <i>19</i> |
| <i>2.1.2 A confiabilidade de um teste na TCT (<math>\alpha_{XY}</math>) .....</i>       | <i>20</i> |
| 2.1.2.1 Coeficiente de Correlação Linear de Pearson ( $r$ ).....                        | 20        |
| 2.1.2.2 Fórmula de Spearman- Brown ( $\square$ ).....                                   | 21        |
| 2.1.2.3 Fórmula de Kuder- Richardson ( $\square$ ) .....                                | 21        |
| <i>2.1.3 A discriminação de um item .....</i>   | <i>22</i> |
| 2.1.3.1 Correlação Bisserial por ponto ( $\square_{bp}$ ).....                          | 22        |
| 2.1.3.2 Correlação Bisserial.....   | 22        |
| <b>2.2 A Teoria de Resposta ao Item (TRI).....</b>                                      | <b>23</b> |
| <i>2.2.1 Critérios de paralelismo primitivos .....</i>                                  | <i>23</i> |
| <i>2.2.2 Traço Latente e Proficiência - Pressupostos Fundamentais .....</i>             | <i>24</i> |
| <i>2.2.3 Consistência da TRI .....</i>  | <i>24</i> |
| <i>2.2.4 Tipologia dos modelos matemáticos da TRI.....</i>                              | <i>25</i> |
| 2.2.4.1 Modelo unidimensional, dicotômico com um parâmetro .....                        | 25        |
| 2.2.4.2 Modelo unidimensional, dicotômico com dois parâmetros.....                      | 26        |
| 2.2.4.3 Modelo unidimensional, dicotômico com três parâmetros.....                      | 27        |
| <i>2.2.5 Expediente Funcional dos modelos da TRI.....</i>                               | <i>29</i> |
| <i>2.2.6 Estimação dos parâmetros na Teoria de Resposta ao Item – Equalização .....</i> | <i>29</i> |
| 2.2.6.1 Estimação para um grupo submetido a um único teste .....                        | 31        |
| 2.2.6.2 Vantagens e desvantagens dos dois métodos de estimação:.....                    | 32        |
| <br>  |           |
| <b>3 A IMPORTÂNCIA DO ITEM .....</b>  | <b>34</b> |
| <b>3.1 A qualidade do item.....</b>   | <b>35</b> |
| <b>3.2 Banco de itens .....</b>   | <b>36</b> |
| <b>3.3 A Elaboração de itens.....</b>   | <b>37</b> |
| <i>3.3.1 Elementos constitutivos de um item de múltipla escolha.....</i>                | <i>37</i> |

|  |           |
|--|-----------|
| 3.3.2 Exemplo de item.....                                       | 38        |
| 3.3.3 Etapas da elaboração de itens .....                        | 39        |
| 3.3.3.1 Exemplo de item tecnicamente correto .....               | 45        |
| 3.4 Informação do item .....                                     | 47        |
| 3.5 Informação do teste .....                                    | 49        |
| <br>   |           |
| <b>4 ENSAIOS E EXPERIMENTOS .....</b>                            | <b>50</b> |
| 4.1 Descrição e apresentação dos elementos dos ensaios.....      | 50        |
| 4.1.1 Amostras de alunos.....                                    | 50        |
| 4.1.2 Conjunto de itens para pré-teste .....                     | 53        |
| 4.1.3 Primeira análise.....                                      | 57        |
| 4.1.4 Medidas de posição e medidas de dispersão: .....           | 57        |
| 4.1.5 Avaliação dos parâmetros dificuldade e discriminação ..... | 59        |
| 4.2 Aplicação da Teoria de Resposta ao Item - TRI.....           | 65        |
| 4.2.1 Calibragem e Reestruturação do Pré- teste .....            | 65        |
| 4.3 Aplicação final do teste .....                               | 69        |
| 4.4 O produto: um banco de itens com planilhas .....             | 75        |
| <br>   |           |
| <b>5 CONSIDERAÇÕES FINAIS.....</b>                               | <b>77</b> |
| <br>   |           |
| <b>REFERÊNCIAS .....</b>   | <b>79</b> |
| <br>   |           |
| <b>APÊNDICES .....</b>   | <b>81</b> |
| Apêndice 1 - Conjunto de itens utilizado no pré-teste.....       | 81        |
| Apêndice 2 - Curvas características dos itens (pré-teste) .....  | 97        |
| Apêndice 3 - Histograma das habilidades .....                    | 106       |
| Apêndice 4 - Base teórica do BILOG-MG.....                       | 107       |

## 1 INTRODUÇÃO

A missão de atribuir um conceito, inerente a uma competência, a uma pessoa, é, por si somente, complexa, uma vez que o desempenho relativo a qualquer competência avaliada é um produto de inúmeras variáveis, algumas de uma subjetividade sem limites. E mesmo aquelas variáveis, até certo ponto, mais objetivas, ainda são inalienáveis no processo de avaliação como um todo. Portanto, pode-se garantir que qualquer metodologia de avaliação aplicada a pessoas é estimativa. Procura-se, cada vez mais, fechar as lacunas de um processo com o advento de outro, ou seja, cada nova metodologia é o constructo da implementação alavancada pela pesquisa e descoberta de parâmetros e brechas até então desconhecidos.

Na avaliação escolar, isso é de uma responsabilidade sem precedentes, uma vez que as competências avaliadas não se restringem ao domínio escolar como muitos, ingenuamente, acreditam. São competências ligadas à instrumentação para a própria vida do aprendiz. A responsabilidade, neste caso, assume um peso muito maior, se pensarmos em termos de cidadania, no tocante ao compromisso de servir país. Mas é na realização pessoal do aprendiz, no reconhecimento de suas reais potencialidades e na justiça desse reconhecimento que reside a maior responsabilidade. A avaliação que ignora atributos e circunstâncias próprias de cada individualidade, com uma objetividade exclusivamente acadêmica, mostra-se sumária.

A expressão da avaliação escolar, como se trata de um processo de mensuração, pode ser um número. Então, trata-se de atribuir ao educando um score que o posicione em relação aos seus pares, que o classifique, segundo a sua competência supostamente avaliada. Subliminarmente à expressão aparentemente ingênua "aluno 10" ou "um zero à esquerda" há um conceito considerado desumano: ninguém é absolutamente vazio de competência e nem absolutamente competente. Portanto, quantificar a qualidade com justeza é uma missão quase impossível.

Entretanto, a avaliação é uma incompetência humana indispensável, até mesmo vital, para a atividade produtiva da espécie. Sem a avaliação, perde-se a referência para se construir o conhecimento e para se construir os homens sob o ponto de vista desse conhecimento e sob o foco de suas potencialidades. Até mesmo as relações humanas são enriquecidas em face das avaliações. A avaliação promove a otimização racional das atividades produtivas e dos processos.

A complexidade de se avaliar quantitativamente uma competência deve-se à sua natureza qualitativa no domínio de suas variáveis. Por exemplo, como avaliar numericamente

a sensibilidade de um indivíduo ou sua consciência ecológica. Em razão de tal perspectiva, os descritores em Educação, que são as habilidades avaliadas, referem-se apenas ao que pode ser avaliado. O ramo do conhecimento encarregado dessa missão sistemática é a Matemática, que, em seu universo de modelagem, trabalha com relações funcionais e relações estatísticas, chamadas de correlações. Enquanto no domínio funcional, as variáveis estão sob certo controle, podendo ser analisadas isoladamente ou não, na Estatística, o objeto da avaliação se apresenta como produto de todas as interferências das variáveis diversas, ou seja, de sua variabilidade conjunta. Esse produto é analisado sob o ponto de vista de seus efeitos, o que torna as técnicas de modelagem estatística convenientes para a avaliação qualitativa escolar.

O Brasil tem larga tradição em avaliações de promoção de um nível escolar para outro. Até a década de 1960, havia o Exame de Admissão ao Ginásio, uma ponte entre os atuais níveis Fundamental I e Fundamental II e, desde 1911, pelo decreto nº 8.659, de 05 de abril de 1911, o exame vestibular, uma ponte entre o atual Ensino Médio e o Ensino Superior. Em ambos os casos, há o reflexo de um problema social ligado ao acesso escolar, uma vez que o número de vagas oferecidas tornou-se muito aquém da demanda no sistema público de ensino. Então, desenvolveu-se uma cultura da avaliação puramente classificatória, baseada nos próprios conteúdos ministrados nas escolas.

Criado em 1998, o Exame Nacional do Ensino Médio – ENEM - veio como uma proposta de avaliar o Ensino Médio, atendendo uma exigência do Banco Mundial, que apontou a necessidade de uma avaliação nacional periódica que indicasse caminhos para a aplicação eficiente dos recursos destinados à Educação, uma vez que o Brasil, no ranque da Educação, estava em situação inferior a nações demograficamente semelhantes como o México e a Índia. Atualmente, em inscrições, o ENEM é o segundo maior exame da Terra, atrás apenas da China (BRASIL, 2015)<sup>1</sup>. No Brasil, ainda há o sistema de Avaliação do Ensino Básico – SAEB e a Prova Brasil, destinados a avaliar os cursos do Ensino Fundamental.

A maioria desses exames, assim como os exames vestibulares das instituições particulares, sempre adotou a Teoria Clássica de Testes – TCT - como metodologia de avaliação e classificação dos participantes. A TCT considera apenas os escores brutos na forma absoluta ou atribui pesos distintos às áreas de conhecimento. Isto foi questionado por educadores como uma metodologia de relativa justiça, uma vez que não discrimina

---

<sup>1</sup> BRASIL. Ministério da Educação e Cultura. **ENEM**: a segunda maior prova de acesso ao Ensino Superior do mundo. 2015. Disponível em: <http://www.brasil.gov.br/educacao/2015/10/enem-a-segunda-maior-prova-de-acesso-ao-ensino-superior-do-mundo>. Acesso em: 13 jan. 2016.

corretamente a posse de uma competência, criando resultados díspares. A TCT considera a fidedignidade dos itens ou questões que são aplicadas na avaliação. Entretanto, a TCT é de grande importância na análise prévia dos itens, complementando o modelo estatístico que a superou: a Teoria de Resposta ao Item - TRI. O ENEM passa a usar a TRI a partir de sua nova versão em 2009, apesar de ela ser conhecida desde o início da década de 1990. Por se tratar de uma modelagem bem complexa do ponto de vista de análise e na robustez dos cálculos, a TRI tornou-se um processo mais fluente com o desenvolvimento de *softwares* como o R e o BILOG-MG.

### 1.1 O Trabalho de Pesquisa

A pretensão do presente trabalho é contemplar, da maneira mais elementar possível, os educadores preocupados com a modelagem TRI, dada a sua complexidade natural. Alguns desses questionam sobre a atribuição de proficiência no que diz respeito à fidedignidade. Talvez, o que incomode esses educadores é a não compreensão dos fatores que conferem à TRI esse *status* de coerência, discriminação e justiça.

Toda metodologia tem procedimentos que procuram garantir a consecução de algum objetivo. No caso da TRI, esse objetivo é a atribuição quantitativa de uma proficiência. A Teoria Clássica de Testes – TCT apresenta limitações que comprometem os critérios de avaliação como a dependência da amostra medida (*test-dependent*), dentre outras. A TRI veio com essas lacunas cobertas, apesar de seus pressupostos, e goza de fidedignidade maior do que a TCT. Portanto, diante do exposto, a questão central que se coloca é: **quais os ingredientes e/ou procedimentos que garantem conceitualmente a proficiência de um aluno, referenciando-se nos itens constituintes das avaliações e admitindo-se que tais itens contemplem as competências e habilidades a que se propõem?**

Tem-se, portanto, como justificativa para essa pesquisa, o fato de que a TRI é um referencial de proficiência consolidado entre os pesquisadores e estatísticos, mas seus atributos e sua consistência teórica estão longe da compreensão dos mais diretamente envolvidos, os educadores da formação básica. Portanto, é essencial que os educadores desses ciclos compreendam, pelo menos em linhas gerais, como funciona essa modelagem estatística. Não se trata de fornecer subsídios para uma preparação direcionada do aluno como fazem os cursos preparatórios para a educação superior, mas fornecer subsídios para que o processo de aprendizagem contemple a compreensão e o desenvolvimento de competências

reais. Há no meio escolar uma crença equivocada de que as competências e habilidades são para o momento de avaliar, mas como avaliar aquilo que não foi trabalhado ao longo do processo ensino-aprendizagem? A compreensão do modelo de avaliação pode ser um argumento definitivo para mudar o foco do processo com um trabalho de aprendizagem real e significativa.

Assim, como objetivo geral dessa pesquisa, procura-se verificar qual dos atributos da Teoria de Resposta ao Item lhe confere a segurança de proficiência na avaliação escolar. Como objetivos específicos, essa pesquisa busca

- estabelecer o conceito de discriminação conceitual para um item de avaliação;
- estabelecer os níveis de discriminação conceitual para os itens de avaliação;
- determinar a relação funcional ou o grau de correlação entre a competência contemplada em um item e o nível de discriminação desse item;
- explicitar as abordagens estatísticas constituintes da TRI.

A TRI usa, como parâmetros, quantificadores referentes ao desempenho (discriminação, dificuldade e acerto por acaso) em comparação com o traço latente (proficiência), calibrados a partir de uma prova piloto (avaliação dos itens), numa prova oficial (avaliação do grupo) e o conjunto de respostas de cada participante (avaliação individual). Portanto, tem-se que como hipótese de que com esses parâmetros, estabelece-se o grau de proficiência de cada participante na habilidade pretendida.

## 1.2 Metodologia

Inicialmente foi feita uma pesquisa teórica no acervo bibliográfico disponível como livros de estatística, documentos oficiais do MEC, dissertações de mestrado e teses de doutorado constantes no acervo da CAPES, planilhas de resultados com utilização de *softwares* e aplicativos. Entre os autores que merecem destaque, nessa pesquisa, estão, entre outros, Pasquali (1996) e Andrade, Tavares e Valle (2000).

Num segundo momento, foram aplicados testes com itens matemáticos em uma amostra de alunos do Ensino Médio de uma escola particular de Belo Horizonte. Concomitantemente a esses dois momentos, foi desenvolvido o produto dessa pesquisa, tratando-se de uma rotina com a utilização de *softwares* que tornaram viável o processamento de cálculos estatísticos e a execução de toda a base teórica relativa à TRI (BILOG-MG e EXCEL) e que será descrito adiante.

## 2 REVISÃO DE LITERATURA

A revisão de literatura contempla uma bibliografia composta de inúmeros artigos, uma vez que o tema, apesar de parecer uma novidade, tem décadas de discussão e reflexão contando com uma vasta bibliografia. No Brasil, duas obras pioneiras abordam a TRI: Pasquali (1996) e Andrade, Tavares e Valle (2000).

A pretensão deste trabalho visa à dinâmica do processo de avaliação das habilidades matemáticas dos alunos, tanto pela TCT quanto pela TRI, pois o perfil do produto almejado é um instrumento que, em pequena escala, viabiliza aplicações da TRI para efeitos de compreensão de habilidades e competências e, conseqüentemente, aperfeiçoamento da produção de bons itens, garantindo-se uma avaliação mais criteriosa.

Os temas estruturantes deste trabalho são a TCT - Teoria Clássica dos Testes e a TRI - Teoria de Resposta ao Item. Nesse universo temático, a rotina de cada teoria em execução será sublinhada, estabelecendo-se comparações e avaliando-se o quão complementares elas são.

A literatura de apoio conta com conteúdos de Estatística e de softwares, uma vez que a execução, tanto da TCT quanto da TRI, exige ferramentas de tecnologia avançadas. Nesses campos, são fundamentais as abordagens relacionadas aos modelos matemáticos, suas hipóteses básicas, seus pressupostos e suas limitações. Esse trabalho visa, preferencialmente, os educadores da formação básica e, tendo em vista o ENEM, o modelo de avaliação dicotômico com três parâmetros é o mais adequado. A seguir, a descrição de cada uma delas.

### 2.1 A Teoria Clássica dos Testes (TCT)

A TCT, base conceitual para os modelos clássicos de psicometria tradicional usados desde o início do século XX, teve a sua sistematização axiomática efetivada nos anos 1950 com os trabalhos de Guilford em 1936 e 1954 e o de Gulliksen em 1950 (PASQUALI, 1996).

A abordagem da TCT - Teoria Clássica dos Testes - fornece os parâmetros de propriedades psicométricas para o modelo, como o chamado **Índice de dificuldade**, que é a proporção de alunos acertantes de um item; o **Índice de discriminação do item**, que diferencia o aluno que domina determinada competência daquele que não domina; e a **correlação bisserial**, que é uma medida que associa o desempenho no item com o desempenho no teste. Isso significa estabelecer uma correlação entre a capacidade de acerto

do item e uma variável latente, não detectável que, por hipótese, é proveniente de uma distribuição de probabilidade normal ou gaussiana. A fidedignidade do teste, na TCT, é avaliada com o uso de fórmulas como, por exemplo, a Fórmula 20 de Kuder-Richardson e a chamada Fórmula de Spearman - Brown, pois se espera os mesmos resultados se um mesmo teste é aplicado várias vezes para uma mesma amostra de estudantes. Conceitualmente, esse modelo tem como axioma fundamental a equação:

$$X_j = T_j + E_j \quad (1)$$

Na qual  $X_j$  é o escore do respondente  $j$ ,  $T_j$  é o escore verdadeiro do respondente  $j$  e  $E_j$  é a componente de erro do respondente  $j$ . O modelo postula que os testes mostram um escore diferente do que seria característico do respondente, uma vez que as circunstâncias ou a ineficiência do teste pode mascarar o desempenho desse respondente. Então, três axiomas têm predominância na abordagem deste modelo: Se uma amostra de respondentes é submetida a dois testes então,

- I - o escore esperado é o verdadeiro;
- II- não há correlação entre o escore verdadeiro e o respectivo erro e
- III- não há correlação entre os respectivos erros dos dois testes.

O modelo da TCT, em sua dinâmica de expediente, processa descritivamente os escores brutos dos respondentes, utilizando as medidas tradicionais da Estatística e verifica a fidedignidade dos testes através de correlações. Os parâmetros avaliados são a dificuldade de cada item e a discriminação do item.

### ***2.1.1 O Índice de Dificuldade de um item ( $I_d$ )***

Esse índice compreende o percentual de alunos que acertaram o item. Quanto maior o seu valor, mais fácil é considerada a questão. Os limites desse índice são 0 (ninguém acerta o item) e 1,0 (todos acertam o item).

Pasquali (1996) e Viana (1982) consideram que os índices de dificuldade médios em torno de 0,5 produzem distribuições de escore com grandes variações. Esses autores utilizam a classificação de Condé para os níveis de dificuldade dos itens, quais sejam:

- ITEM FÁCIL:  $I_d$  maior do que 0,70;
- ITEM MÉDIO:  $I_d$  entre 0,30 e 0,70 (inclusive)
- ITEM DIFÍCIL:  $I_d$  menor ou igual a 0,30

Para a elaboração de um teste com N itens, Pasquali e Primi (2003) sugerem a seguinte distribuição:

- N/10 questões MUITO FÁCEIS (10% do total);
- N/10 questões MUITO DIFÍCEIS (10% do total);
- N/5 questões FÁCEIS (20% do total);
- N/5 questões DIFÍCEIS (20% do total)
- 2N/5 questões MÉDIAS (40% do total)

### 2.1.2 A confiabilidade de um teste na TCT ( $\alpha_{XY}$ )

A confiabilidade  $\alpha$  de um teste é a razão entre a variância do escore verdadeiro ( $s^2_V$ ) e a variância do escore atribuído ao respondente ( $s^2_X$ ), ou seja, é a parte do escore verdadeiro que o escore atribuído pode explicar. Matematicamente, tem-se que:

$$\alpha_{XY} = \frac{s^2_V}{s^2_X} \quad (2)$$

A confiabilidade  $\alpha$  é tanto maior, quanto menor for o erro e seus limites efetivos são:  $\alpha = 0$  e  $\alpha = 1$ .

Para tanto, são adotados os seguintes coeficientes para avaliar a confiabilidade de um teste: coeficiente de correlação linear de Pearson, fórmula de Spearman-Brown e fórmula de Kuder-Richardson, as quais serão descritas a seguir:

#### 2.1.2.1 Coeficiente de Correlação Linear de Pearson (r)

Esse indicador mede a associação entre duas variáveis contínuas x e y, que têm uma relação linear, por meio da relação entre a variação conjunta entre x e y e a variação individual de cada uma delas, se o número de pares de observação não for muito baixo:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum(x_i - \bar{x})^2)(\sum(y_i - \bar{y})^2)}} \quad (3)$$

O coeficiente de correlação linear de Pearson assume valores entre -1 a 1, sendo que os valores positivos apontam uma relação linear no mesmo sentido, ou seja, valores crescentes

de x e de y, e os negativos, uma relação linear inversa, ou seja, valores crescentes em x e decrescentes em y. O valor que determina uma confiabilidade considerável é a partir de |0,9|, entretanto valores maiores que |0,7| também expressam correlações significativas. Antes de se usar a fórmula de Pearson, é conveniente verificar o diagrama de dispersão relacionado aos itens e conferir a existência de uma relação linear entre eles.

### **2.1.2.2 Fórmula de Spearman- Brown ( $\alpha$ )**

A partir da verificação de existência de correlação linear entre as variáveis em análise, pode-se usar a fórmula de Spearman-Brown para verificar a confiabilidade, sendo essa:

$$\alpha_{XY} = \frac{n \cdot r_{XY}}{1 + (n-1)r_{XY}} \quad (4)$$

Nessa fórmula, o número total de itens é dividido, aleatoriamente, em duas metades e a correlação é verificada entre as duas metades ordenadas no mesmo sentido.

### **2.1.2.3 Fórmula de Kuder- Richardson ( $\rho$ )**

Esta fórmula é especialmente indicada para verificar o coeficiente de confiabilidade quando os itens são dicotômicos e as variáveis (os pares ordenados determinados, com respectividade, pelas duas metades discretas descritas no subitem anterior (2.1.2.2)).

$$\rho_{kr} = \frac{n}{n-1} \left( 1 - \frac{\sum_{i=1}^n \theta_i (1 - \theta_i)}{s^2} \right) \quad (5)$$

Em que:

$\theta_i$  é a porcentagem dos examinandos que responderam corretamente ao item i.

n é o número de itens do teste.

$S^2$  é a variância dos escores do teste.

### 2.1.3 A discriminação de um item

É o potencial que o item tem de separar os respondentes que dominaram a competência relativa a esse item dos que ainda não conseguiram esse domínio. O método tradicional de se medir a discriminação do item consiste na diferença entre a dificuldade do item no grupo de alunos formado pelos 27% de escores superiores e o grupo dos 27% inferiores. Na verdade, os escores da amostra são ordenados e a amplitude amostral igualmente dividida em 100 partes (percentis); toma-se, então, a diferença entre os índices de dificuldade calculados em duas subamostras: os 27% maiores e os 27% menores. Entretanto, as correlações bisseriais são mais eficientes na medição da discriminação que será descrita a seguir:

#### 2.1.3.1 Correlação Bisserial por ponto ( $\rho_{bp}$ )

Usado quando uma das variáveis é contínua e a outra é dicotômica.

$$\rho_{bp} = \frac{\bar{X}_c - \bar{X}_e}{S} \cdot \sqrt{\frac{\tilde{\theta}}{1 - \tilde{\theta}}} \quad (6)$$

Em que:

$\bar{X}_c$  é a média dos escores dos examinandos que acertaram o item.

$\bar{X}_e$  é a média global dos escores.

S é o desvio padrão do teste.

$\tilde{\theta}$  é a proporção de examinandos que acertaram o item.

#### 2.1.3.2 Correlação Bisserial

Se uma das variáveis é discreta e a outra é dicotômica, temos.

$$\hat{\rho}_b = \frac{\bar{X}_p - \bar{X}_t}{S} \cdot \frac{\tilde{\theta}}{Y} \quad (7)$$

Em que

$\bar{X}_p$  é a média dos escores dos examinandos que deram resposta certa ao item.

$\bar{X}_t$  é a média global dos escores.

S é o desvio padrão do teste.

$\hat{\theta}$  é a proporção de examinandos que acertaram o item.

Y é a ordenada na curva Normal Padrão, correspondente à área  $\hat{\theta}$ .

## 2.2 A Teoria de Resposta ao Item (TRI)

A TRI - Teoria de Resposta ao Item- apresenta pressupostos e parâmetros e, nesse trabalho, fica evidenciado sua complementaridade à TCT, ou seja, a TRI preenche algumas lacunas da TCT. Porém, torna-se necessário esclarecer a importância da qualidade dos itens (questões), avaliados de forma a estabelecer a proficiência do aluno através da construção de uma escala, levando-se em consideração os três parâmetros que determinam o seu perfil, a saber: **dificuldade do item**, medido na mesma escala do traço latente; **discriminação do item** e probabilidade de **acerto casual**.

### 2.2.1 Critérios de paralelismo primitivos

O problema básico da discussão teórica que acometia a TCT era como avaliar as habilidades de testandos submetidos a avaliações diferentes. O problema decorrente dessa lacuna era a relativização dos testes pelo nível de competência da amostra de testandos. De acordo com Andrade, Tavares e Valle (2000), nos anos 1920, Thurstone (1928 - 1959) questionava: "Na sua função de medir, um instrumento não pode ser afetado pelo objeto de medida". Ao encontro dessa demanda, também de acordo com os autores, Guliksen em 1950 postulou que dois testes seriam considerados formas paralelas se, em relação aos escores, convertidos para uma mesma escala, apresentassem a mesma média, o mesmo desvio padrão e a mesma correlação.

Posteriormente, Angoff (1971) iria delimitar mais o critério de Guliksen com as seguintes ponderações:

- os testes devem medir a mesma habilidade ou competência;
- os testes devem apresentar uma equivalência independente dos dados utilizados para estabelecê-la;
- os escores de dois testes considerados equivalentes devem ser intercambiáveis;
- a equivalência dos testes deve apresentar simetria de um em relação a outro.

No final dos anos 1970, ainda de acordo com Vianna (2003), Lord em 1977 conclui que se dois testes têm níveis de dificuldades distintos, então, a relação entre eles não pode ser linear e, conseqüentemente, não apresentam a mesma fidedignidade.

### 2.2.2 Traço Latente e Proficiência - Pressupostos Fundamentais

A TRI conta, fundamentalmente, com dois pressupostos: A **Unidimensionalidade**, segundo a qual o desempenho ( $d$ ) de um indivíduo em uma tarefa é função de habilidades ( $\theta_i$ ), que constituem apenas um traço latente ( $\theta$ ), portanto,  $d = f(\theta_1, \theta_2, \theta_3, \dots, \theta_{n-1}, \theta_n) = f(\theta)$ .

O outro pressuposto, chamado de **Independência Local**, postula que o desempenho de um indivíduo em um item independe de seu desempenho em qualquer outro item. Isso permite, na fase de calibração dos itens, pelo Processo da Verossimilhança Máxima, registrar:

$$P(U=1|\theta) = [P(u_1 =1|\theta). P(u_2 =1|\theta).P(u_3 =1|\theta)... P(u_k =1|\theta)]. [ Q(u_{k+1} =1|\theta). P(u_{k+2} =1|\theta).P(u_{k+3} =1|\theta)... P(u_n =1|\theta)] = \prod_{i=1}^k P(u_{ij} =1|\theta). \prod_{i=k+1}^n Q(u_{ij} =0|\theta), \text{ em que:}$$

- $U = (u_1, u_2, u_3, \dots, u_n)$  é o vetor de respostas do indivíduo,
- $P(u_{ij} =1|\theta)$  e  $Q(u_{ij} =0|\theta)$  são, respectivamente, a probabilidade de acertar o item  $i$ , tendo a habilidade  $\theta$  e a probabilidade de errar o item  $i$ , tendo a habilidade  $\theta$ .

Então,  $P(u_{ij} =1|\theta) = 1 - Q(u_{ij} =0|\theta)$ .

Dessa maneira, pode-se avaliar, com segurança, se um aprendiz dominou a técnica de resolução de uma equação quadrática, assim como é possível avaliar se um atleta sabe nadar em estilo borboleta, por exemplo.

Entretanto, há atributos de um indivíduo sobre os quais a avaliação não pode ser feita objetivamente. A inteligência, por exemplo, pode ser estimada somente por meio de situações sob as quais se submete o indivíduo. A essas habilidades ou competências, dá-se o nome de **Traços Latentes**. A quantificação dessas competências em uma escala adequada gera a **Proficiência**. A TRI considera que, de posse de uma determinada proficiência, o testando está apto para acertar determinado item e todos com um nível de dificuldade inferior a ele.

### 2.2.3 Consistência da TRI

A Teoria da Resposta ao Item consiste em um modelo estatístico cujo foco de avaliação de uma habilidade ou conhecimento de um examinando é desvinculado da prova e se concentra nos itens. Esses, por sua vez, estão constituídos segundo os parâmetros relativos à dificuldade, à discriminação e ao acerto por acaso (ANDRADE; TAVARES; VALLE,

2000). Esses parâmetros são obtidos por um processo chamado de **Calibração**. O expediente começa com um teste piloto ou pré-teste, que vai servir de base para essa calibração. O teste resultante é, então, aplicado aos alunos aos quais se destina.

#### **2.2.4 Tipologia dos modelos matemáticos da TRI**

Os modelos matemáticos se diferenciam de acordo com o número de habilidades que avaliam, o tipo de resposta dos seus itens e o número de parâmetros que utilizam. Assim, tem-se, respectivamente:

1º) Modelos unidimensionais (avaliam apenas uma habilidade) e até multidimensionais (avaliam várias habilidades);

2º) Modelos dicotômicos (Os itens têm respostas do tipo certo ou errado) ou não dicotômicos (itens com respostas abertas ou parcialmente abertas);

3º) Modelos que avaliam somente pelo índice de dificuldade dos itens (1 parâmetro), Modelos com 2 parâmetros (dificuldade e discriminação dos itens) e modelos com 3 parâmetros (dificuldade, discriminação e respostas ao acaso).

##### **2.2.4.1 Modelo unidimensional, dicotômico com um parâmetro**

Matematicamente, o modelo unidimensional é expresso pela sentença:

$$P(X_i = 1 | \theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}} \quad (8)$$

Em que:

$\theta$  é a variável latente.

$P(X_i = 1 | \theta)$  é a probabilidade de acerto do item, dado  $\theta$ .

$b$  é o parâmetro de dificuldade do item.

$D$  é constante relativa à consideração de normalidade da distribuição (vale 1,7).

$e$  é a constante de Euler (aprox. 2,718)

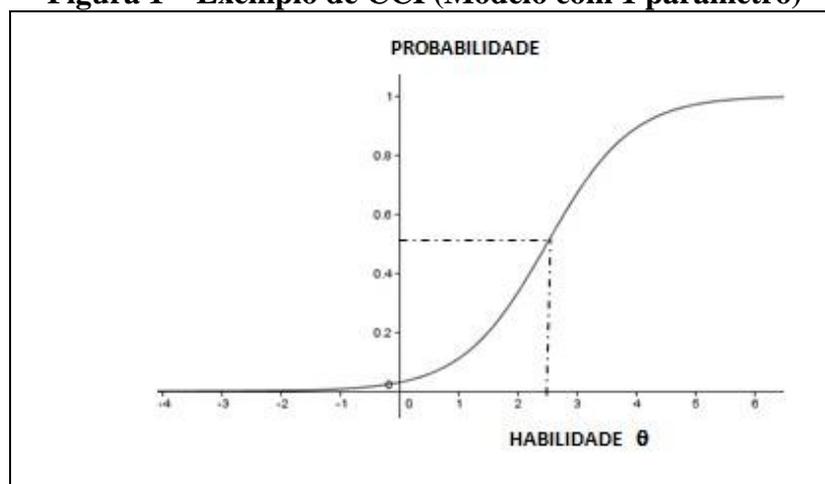
O modelo dicotômico com um único parâmetro que é a dificuldade do item foi criado pelo matemático dinamarquês Georg Rash (1901-1980). Rasch considerou, em particular, a proporcionalidade  $\frac{\theta}{b}$ , entre a habilidade ou traço latente e o índice de dificuldade do item,

pontuando que se dois itens, destinados a dois testandos cujas habilidades estão na razão 2 têm dificuldades na razão 2, então, a probabilidade de ambos acertarem o item é a mesma. Georg Rasch, foi tão importante quanto Lord em 1952, acatando a percepção de D. N. Lawley em 1943 de que muitos constructos da Teoria Clássica dos Testes - TCT poderiam ser expressos em termos de parâmetros da Curva Característica do Item – CCI, como nos modelos dicotômicos da TRI, como mostra a figura 1. (PASQUALI, 2013)

Nesse modelo, o parâmetro de **discriminação**  $a_i$  é igual para todos os itens e o parâmetro da resposta aleatória  $c_i$  é considerado inexistente, ou seja, é igual a zero. No caso da habilidade  $\theta$  ser igual à dificuldade, tem-se uma probabilidade de acerto de 50% , assim:

$$P(X_{i=1}|\theta) = \frac{e^{D(\theta-b_i)}}{1+e^{D(\theta-b_i)}} = \frac{1}{1+1} = \frac{1}{2} = 50\%$$

**Figura 1 – Exemplo de CCI (Modelo com 1 parâmetro)**



Fonte: Elaborado pelo autor.

#### 2.2.4.2 Modelo unidimensional, dicotômico com dois parâmetros

Estabelece, como parâmetros, a **dificuldade**  $b_i$  e a **discriminação**  $a_i$  do item, considerando que não existe a probabilidade de acerto por acaso. Então, tem-se, matematicamente que:

$$P(U_i=1|\theta) = \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \quad (9)$$

Em que:

$\theta$  é a variável latente.

$P(X_i = 1 | \theta)$  é a probabilidade de acerto do item, dado  $\theta$ .

$a$  é o parâmetro de discriminação do item.

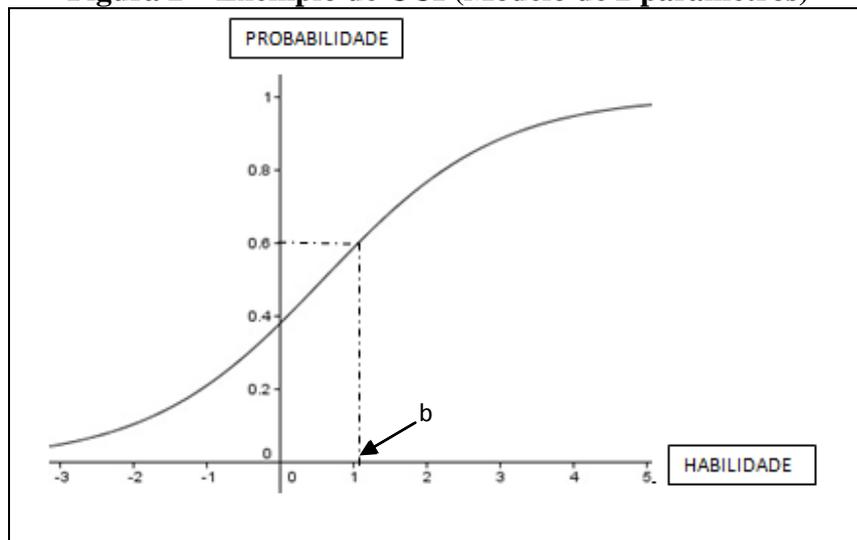
$b$  é o parâmetro de dificuldade do item.

$D$  é constante.

$e$  é o número de Euler (aprox. 2,718)

O parâmetro de discriminação  $a_i$  corresponde à inclinação da reta tangente no ponto de inflexão da CCI. (ANDRADE, TAVARES, VALLE, 2000). Dessa forma, quanto mais acentuada essa inclinação, maior é a discriminação do item e mais distinta é a separação dos respondentes mais hábeis, como pode ser notado na figura 2, a seguir, em que  $b$  é aproximadamente 1:

**Figura 2 – Exemplo de CCI (Modelo de 2 parâmetros)**



FONTE: Elaborado pelo autor.

#### **2.2.4.3 Modelo unidimensional, dicotômico com três parâmetros**

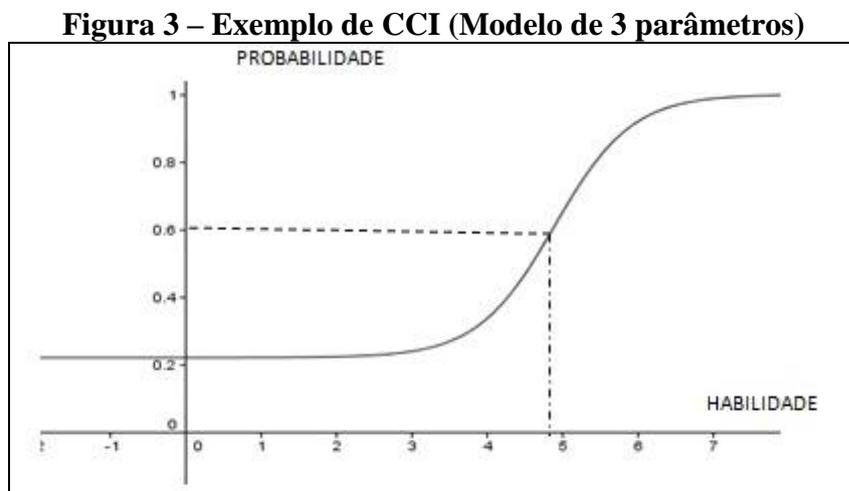
A representação gráfica deste modelo é a Curva Característica do Item - CCI, mostrada na figura 3. Nessa curva, os três parâmetros são identificados:

- O índice de dificuldade  $b$ , na mesma escala da habilidade  $\theta$ , corresponde a uma probabilidade de acerto equivalente à média da probabilidade de resposta aleatória  $c$  e a probabilidade máxima (1), ou seja, uma habilidade de medida  $b$  tem uma probabilidade de

acerto equivalente a  $\frac{1+c}{2}$ . Por outro lado, esse ponto assim determinado é o ponto de inflexão da curva.

- O parâmetro  $a$  equivale à inclinação da curva no ponto de inflexão. Assim, quanto maior o valor de  $a$ , mais íngreme é a curva, maior é o poder de discriminação do item e mais separados por  $b$  estarão os mais aptos (à direita de  $b$ ) e os menos aptos (à esquerda de  $b$ ).

A figura 3 apresenta a probabilidade de acerto, 60%, de um item com dificuldade  $b$  de aproximadamente 4,8 considerando um acerto casual de aproximadamente de 20%, e com discriminação equivalente à curva no ponto de inflexão, tangente aproximadamente 1.



Fonte: Elaborado pelo autor.

Matematicamente, o modelo de 3 parâmetros é expresso pela sentença:

$$P(U_i=1|\theta) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta - b_i)}} \quad (10)$$

Em que:

$\theta$  é a variável latente.

$P(X_i = 1 | \theta)$  é a probabilidade de acerto do item, dado  $\theta$ .

$a$  é o parâmetro de discriminação do item.

$b$  é o parâmetro de dificuldade do item.

$c$  é o parâmetro da resposta ao acaso.

$D$  é constante.

$e$  é a constante de Euler (aprox. 2,718).

### ***2.2.5 Expediente Funcional dos modelos da TRI***

A aplicação do modelo começa com a elaboração de itens, uma missão que apresenta um grande nível de complexidade, uma vez que cada item e o conjunto de itens devem contemplar alguma competência ou habilidade que se pretende avaliar. Portanto, é uma atribuição criteriosa, que norteia todo o processo.

O produto dessa elaboração é um teste para avaliar uma ou mais habilidades. Entretanto, precisa ser confrontado com a realidade. Para isso, aplica-se um pré- teste a uma amostra de alunos. Essa experiência gera um diagnóstico sobre o equilíbrio dessa avaliação por meio das respostas desse grupo amostral. O próximo passo é a estimação dos parâmetros  $a$ ,  $b$  e  $c$  e da habilidade  $\theta$ , processo conhecido como **Calibragem**. Vários são os métodos de estimação, por meio dos quais, essencialmente, é construída uma escala de valores coerente que ajustam, por iteração, a curva resultante de tal maneira que ela se aproxime o máximo possível da curva gerada pelas respostas dadas no pré-teste. Posteriormente, avalia-se o ajuste efetuado para garantir a validade das escalas e parâmetros construídos. Após essa calibragem, o teste é aplicado em uma turma destino.

### ***2.2.6 Estimação dos parâmetros na Teoria de Resposta ao Item – Equalização***

O passo mais importante dos modelos da TRI, através do qual são geradas as referências que caracterizam os itens, é o estabelecimento de valores estimados, em escalas apropriadas, que determinarão os marcos de referência para a proficiência do respondente, posicionando-o, em cada item. Nesse processo, também chamado de calibragem, o procedimento básico é substituir os itens com dificuldade muito grande e os itens com dificuldade muito pequena por outros que confirmam mais equilíbrio ao teste como um todo.

Há duas calibrações a serem feitas: a do item e a da habilidade do respondente. Se for conhecida uma delas, o problema é determinar a outra, caso contrário, é necessário determinar as duas. Então, há três hipóteses a considerar, conforme preconizam Andrade, Tavares e Valle (2000):

- o item está estimado e se quer estimar a habilidade;
- a habilidade é conhecida e se quer estimar o item;
- o item e a habilidade devem ser estimados.

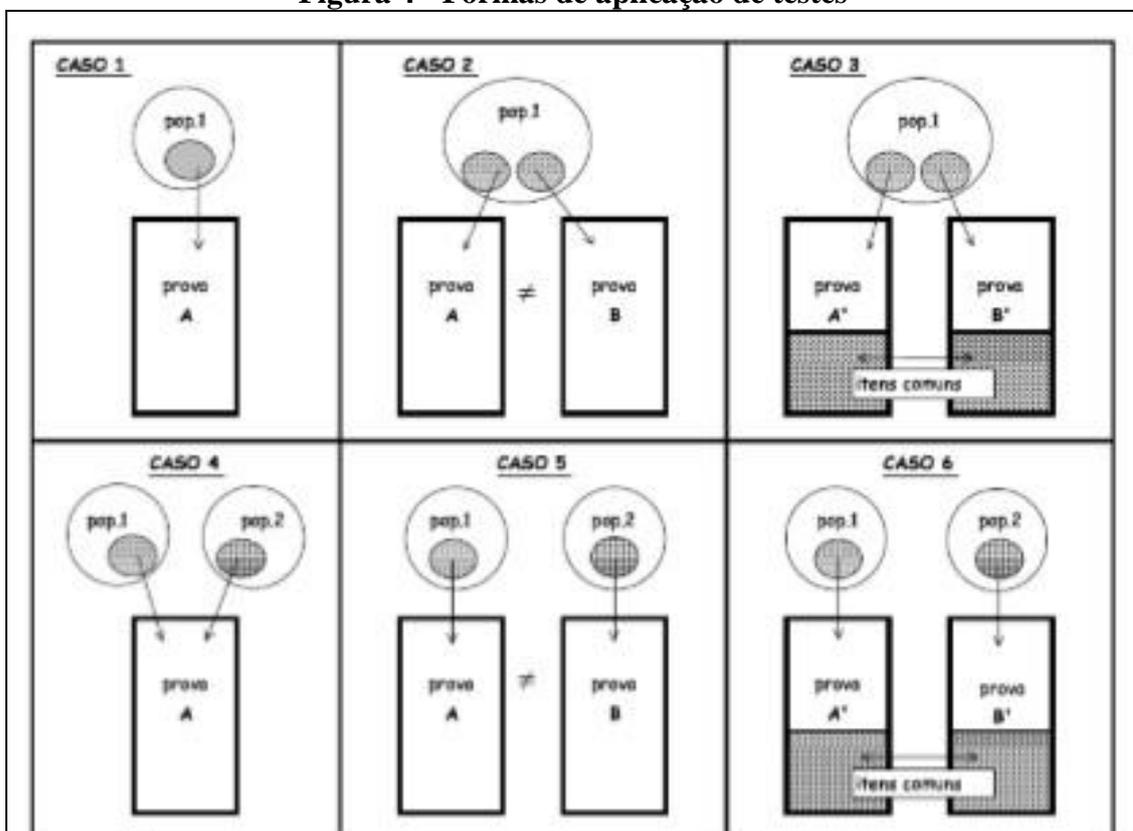
Entretanto, como os únicos dados disponíveis, na maioria dos casos, são as respostas dos testandos do pré-teste, a terceira hipótese é a maior demanda prática.

Por outro lado, a estimação está condicionada a dois fatores, uma vez que é implementada no pré-teste: número de amostras (grupos de respondentes) distintas e número de testes (grupos de itens) destinados a esses respondentes. Então, a questão, desta forma colocada, é chamada de **equalização** e trata-se de contemplar a fidedignidade da avaliação, adaptando-se os processos de estimação à realidade que envolve grupos de respondentes e grupos de itens. Assim, equalizar é tornar comparável, o que, no caso da TRI, significa colocar parâmetros de itens vindos de provas distintas ou habilidades de respondentes diferentes. São exemplos de situações a equalizar, de acordo com Andrade, Tavares e Valle (2000, p.79):

- Um único grupo fazendo uma única prova.
- Um único grupo, dividido em dois subgrupos, fazendo duas provas totalmente distintas.
- Um único grupo, dividido em dois subgrupos, fazendo duas provas apenas parcialmente distintas, ou seja, com alguns itens comuns.
- Dois grupos fazendo uma única prova.
- Dois grupos fazendo duas provas, totalmente distintas (nenhum item comum).
- Dois grupos fazendo duas provas, apenas parcialmente distintas, ou seja, com alguns itens comuns.

A figura 4 representa graficamente as situações descritas:

**Figura 4 - Formas de aplicação de testes**



Fonte: Andrade, Tavares e Valle, 2000, p.79.

O objetivo deste trabalho refere-se exclusivamente ao primeiro caso, uma vez que o foco principal é o ENEM.

O problema fundamental da equalização é garantir que os itens e as habilidades sejam calibrados na mesma métrica. Isso sugere dois tipos de calibração, a de itens e a de habilidades. Por exemplo, se há dois testes com itens totalmente distintos, devem-se calibrar os itens como se fossem de um só teste. Por outro lado, se os itens não são totalmente distintos, deve-se focar a calibragem nos itens comuns. Seja como for, a calibragem deve observar a equalização.

### 2.2.6.1 Estimação para um grupo submetido a um único teste

Este é o caso geral de estimação. Pelo **Método da Máxima Verossimilhança Marginal (MVM)**, tem-se a estimação dos parâmetros  $a_i$ ,  $b_i$  e  $c_i$ :

$$a_i : D(1 - c_i) \sum_{j=1}^n [(u_{ji} - P_{ji})(\theta_j - b_i)w_{ji}] = 0 \quad (11)$$

$$b_i : -D(1 - c_i) \sum_{j=1}^n [(u_{ji} - P_{ji})w_{ji}] = 0 \quad (12)$$

$$c_i : \sum_{j=1}^n [(u_{ji} - P_{ji}) \frac{w_{ji}}{P_{ji}^*}] = 0 \quad (13)$$

Em que:

$$w_{ji} = \frac{P_{ji}^* Q_{ji}^*}{P_{ji} Q_{ji}} \quad (14)$$

$$P_{ji}^* = [1 + e^{-D\theta_i(\theta_j - b_i)}]^{-1} \quad (15)$$

$$Q_{ji}^* = 1 - P_{ji}^* \quad (16)$$

Pelo Método Bayesiano, ter-se-ia o seguinte expediente e condições para os parâmetros:

→  $P(x|\theta)$  conhecida e,  $\theta$  desconhecida e  $P(\theta)$  conhecida;

→ Para um conjunto de amostras  $D = x_{ki}$ ,  $k = 1, 2, 3, \dots, N$ , deve-se:

1º) Integrar todo o espaço de parâmetros (Monte Carlo - Algoritmo de Gibbs)

$$P(x|D) = \int P(x|\theta)P(\theta|D) d\theta \quad (17)$$

Em que:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta} \quad (18)$$

E usando a independência das amostras:

$$P(D|\theta) = \prod_{k=1}^n P(x_k|\theta) \quad (19)$$

2º) O método pode ser desenvolvido *on-line*, utilizando-se da forma recursiva do aprendizado Bayesiano:

$$P(D^k|\theta) = P(x_k|\theta).P(D^{k-1}|\theta) \quad (20)$$

$$P(\theta|D^k) = \frac{P(D_k|\theta).P(\theta)}{\int P(D^k|\theta)P(\theta)d\theta} \quad (21)$$

Substituindo-se (20) em (21), tem-se:

$$P(\theta|D^k) = \frac{P(x_k|\theta).P(\theta|D^{k-1})}{\int P(x_k|\theta)P(\theta|D^{k-1})d\theta} \quad (22)$$

### **2.2.6.2 Vantagens e desvantagens dos dois métodos de estimação**

O MVM tem como pontos positivos: (i) A consistência dos parâmetros estimados, dada à natureza assintótica de suas propriedades. (ii) Uma vez estimados os parâmetros, a

estimação da habilidade fica objetiva. Seus pontos negativos são: (i) Não é definido para os casos extremos de acerto total ou de erro total. (ii) Sua matemática é trabalhosa e pesada, exigindo complexas ferramentas computacionais. (iii) É apropriada apenas para o caso de ser grande o número de respondentes.

O método Bayesiano tem como pontos positivos: (i) É compatível com qualquer padrão de resposta. (ii) Uma vez estimados os parâmetros, a estimação da habilidade fica objetiva. Seus pontos negativos são: (i) Tem necessidade de distribuição e probabilidade *a priori* para os parâmetros dos itens. (ii) Seus cálculos matemáticos são mais trabalhosos e pesados do que a do MVM.

### 3 A IMPORTÂNCIA DO ITEM

Como visto, a Teoria Clássica dos Testes centra sua metodologia na prova, ou seja, no instrumento de avaliação como um todo, enquanto a Teoria da Resposta ao Item foca o item como instrumento referencial de aquisição de competências, uma vez que seus parâmetros como **difficuldade, discriminação e probabilidade do acaso** são critérios intrínsecos de segura proficiência. Quando se faz a calibragem dos itens, posteriormente ao pré-teste, cria-se um referencial de avaliação experimentado e aperfeiçoado, cuja essência é a contemplação real das habilidades e competências almejadas.

O chamado **Traço Latente** - habilidade individual que não aparece e que habilita um indivíduo a responder corretamente um item - pode ser indiretamente avaliado através dos parâmetros citados, desde que estes sejam características dos itens calibrados (ANDRADE; TAVARES; VALLE, 2000). Do ponto de vista estatístico, as variáveis latentes são aquelas que não podem ser medidas diretamente, pois não constam da base de dados coletados. Essa definição contempla a inteligência e outras características mais abordadas, como nível socioeconômico, qualidade de vida, efeitos aleatórios e dados faltantes (COIMBRA, 2006).

Em geral, um grande desafio para a Estatística é a medição relacionada a variáveis qualitativas; muito maior se forem latentes. Historicamente, o método científico do século XVII foi, formalmente, a primeira iniciativa a trazer, em essência, a proposta de avaliar e medir qualidades; o Projeto *Characteristica*, de Leibnitz, com sua proposta de uma linguagem universal para a filosofia e a ciência trouxe, em seu bojo, uma homogeneização que facilitaria a construção de escalas referenciais de medição (COIMBRA, 2006).

Na verdade, o traço latente é algo característico do indivíduo que aumenta a sua probabilidade de responder corretamente um item, apesar de tantas interferências de outros fatores. A interação (positiva ou negativa) de todas as variáveis (fatores) gera uma probabilidade maior ou menor de resposta correta. A qualidade de um item é essencialmente a sua capacidade de definir, em um grupo, os níveis de posse de habilidades e competências pré-determinadas. Esse processo é como avaliar um doente pelos seus sintomas ou pelas suas respostas que ele dá em face dos tratamentos, uma vez que não é possível ver o nome da doença escrito em sua testa. Para garantir uma crescente fidedignidade ao processo, é preciso criar uma maneira de cercar bem a competência que se quer avaliar e, para responder a essa demanda, o item deve ter qualidade, sob pena de ser necessário multiplicar o número de itens destinados a avaliar uma mesma competência. Uma imagem adequada é a seguinte:

Deseja-se verificar se um indivíduo tem facilidade de interação com os outros. Então, pergunta-se a ele: Você gosta de trabalhar em grupo? Certamente a resposta não é o suficiente para classificá-lo em IN (interativo) ou NI (não interativo). Seu "sim" pode esconder restrições como, por exemplo, gostar de trabalhar apenas com pessoas conhecidas, o que não o faz interativo. Por outro lado, seu "não" pode ser motivado por outros fatores que o classificam mais do que NI. O impasse é resolvido de uma das duas formas: filtrar cada vez mais a resposta, convergindo-se para o que se quer ou elaborar uma única pergunta capaz de expor eficientemente o indivíduo. A primeira forma significa aumentar, em um mesmo teste, o número de itens e a segunda implica aumentar a qualidade do item.

### **3.1 A qualidade do item**

Em termos gerais, um item tem qualidade quando é eficaz e suficiente para avaliar em que nível de domínio real um indivíduo é classificado em relação a uma habilidade ou competência. Nesse caso, seus parâmetros de dificuldade, discriminação e incidência do acaso atestam sua fidedignidade. Torna-se, portanto, necessário verificar a estabilidade desses parâmetros nos itens, independente da população a ele submetida; por isso, os resultados são comparáveis numa mesma escala. Pela Teoria Clássica dos Testes, cujo foco é o conjunto de itens, os resultados das avaliações só são comparáveis se duas populações são submetidas ao mesmo conjunto de itens ou se são submetidas a provas paralelas, quase sempre muito complexas na elaboração (KLEIN, 2009). Então, tendo-se um banco de dados calibrados numa mesma escala de proficiência, garante-se que dois grupos avaliados com subconjuntos distintos de itens desse banco são comparáveis. Entretanto, espera-se sempre um erro diferente na estimativa de proficiência devido à imprecisão eventual da medição relativa a cada item, ou seja, as qualidades individuais dos itens “não absolutamente correspondentes” (HAMBLETON, JONES, ROGERS, 1993).

Do ponto de vista quantitativo, a essência do item é expressa por seus parâmetros  $a$  (discriminação),  $b$  (dificuldade) e  $c$  (acaso), em face de uma habilidade ou proficiência  $\theta$ , sendo  $b$  e  $\theta$  medidos numa mesma escala. Por outro lado, esses parâmetros serão mais adequados ao avaliar o domínio de determinada habilidade, quanto mais o item proposto estiver, de fato, contemplando essa habilidade. Nesse caso, o foco pedagógico é o principal e, segundo esse foco, procedimentos técnicos indispensáveis devem ser observados e implementados.

### 3.2 Banco de itens

Pode ser considerado um banco de itens qualquer base de dados que tenha uma parte descritiva (enunciados, respostas corretas, respostas incorretas, etc.) e uma parte psicométrica (indicadores e parâmetros da TCT e da TRI, etc.).

Qualquer instituição que tenha uma demanda envolvendo Teoria da Resposta ao Item necessitará, pela exigência periódica de ineditismo, alimentar continuamente um banco de itens visando compor um pré-teste para experimentação e posterior calibragem ou itens calibrados para um teste definitivo. Nesse último caso, os itens disponíveis estarão todos calibrados na mesma escala, de forma a preservar os perfis desejáveis de candidatos para a instituição. Um banco de itens é considerado bem calibrado se as estimativas dos parâmetros dos itens forem adequadas e seus respectivos erros padrões forem baixos.

Em geral, os bancos de itens têm grandes acervos, uma vez que o número de itens de um teste é diretamente proporcional à eficiência na atribuição de uma proficiência. Quanto mais vezes uma habilidade é contemplada nos itens de um teste, maior é a confiabilidade de avaliação da mesma. Este é o caso das avaliações de currículos de algumas disciplinas e séries, nas quais os itens são tão numerosos que demandam mais de duas horas de resolução (KLEIN, 2009).

Um bom banco de itens deve ter, em seu acervo, itens diversificados ao máximo, de modo a satisfazer as demandas de várias naturezas, tais como níveis de dificuldade, níveis de discriminação e enunciados apropriados aos hábitos, costumes e características regionais. Depois da aplicação do pré-teste, as eventuais substituições de itens, na calibração, podem demandar itens com características muito específicas que só um banco bem equipado pode atender.

Os bancos de itens são utilizados bem antes do desenvolvimento da TRI. Entretanto, o advento da informática permitiu um melhor desempenho na construção, na administração e no armazenamento dos itens, além da possibilidade da elaboração dos tipos inovadores de itens, que utilizam os recursos multimídia do computador. O surgimento da TRI também contribuiu para o desenvolvimento dos bancos de itens, adicionando informações psicométricas (parâmetros dos itens) que descrevem mais precisamente as características de cada item (PASQUALI, 2013).

Instituições como o Instituto Nacional de Pesquisas Educacionais Anísio Teixeira - INEP, responsável pela realização do ENEM - mantém um banco permanente de itens

elaborados por educadores de todo o país. Segundo o INEP, a existência de um Banco Nacional de Itens (BNI) é necessária para que se tenha uma quantidade expressiva de itens com comprovada qualidade técnico-pedagógica e psicométrica para compor os testes de uma avaliação em larga escala.

### 3.3 A Elaboração de itens

Para que contemple, de fato, uma determinada habilidade e, possa refletir um justo juízo, um item exige uma apurada técnica de elaboração com sólidas bases pedagógicas que lhe dão fidedignidade. Há diversos manuais de elaboração de itens, principalmente de instituições ligadas ao ensino público e, sobretudo, ao MEC. A essência de todos eles é a mesma: objetividade e precisão de foco em uma habilidade ou competência.

#### 3.3.1 Elementos constitutivos de um item de múltipla escolha

Um item de múltipla escolha é estruturalmente constituído, basicamente, de texto-base, enunciado e alternativas, a seguir definidas:

**Texto-base:** É o texto que se presta a compor ou motivar a situação-problema, sendo que, na maioria dos casos, contextualiza a situação-problema. Esse elemento pode se traduzir em um texto verbal ou não verbal, uma figura, um gráfico, uma tabela, um diagrama, uma charge ou qualquer elemento que comunique uma ideia.

**Enunciado:** É a instrução da tarefa a ser executada pelo testando. Deve ser expressa de maneira clara e objetiva, podendo vir em forma de pergunta a ser respondida pelo testando ou frase a ser completada pelo mesmo, através de uma das alternativas oferecidas.

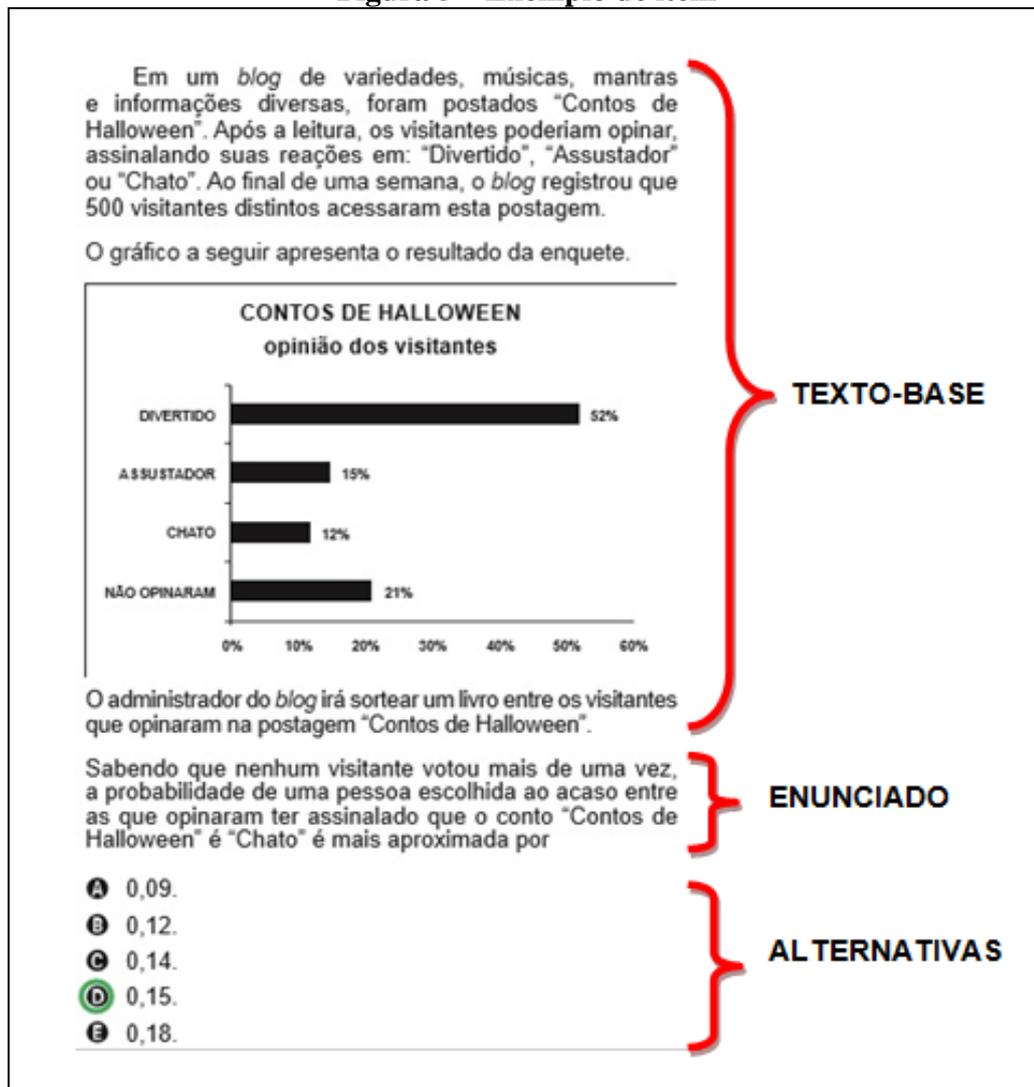
**Alternativas:** São as opções de resposta à situação-problema e seu número é, em geral, quatro (A, B, C e D) ou cinco (A, B, C, D e E). As alternativas são classificadas como *resposta correta* ou *distratores*. No primeiro caso, é a única alternativa que consolida, no testando, o domínio da habilidade contemplada pelo item; trata-se da alternativa que indica um grau de discriminação que atende plenamente o enunciado como resposta da situação-problema. No segundo caso, são as alternativas com respostas plausíveis, ou seja, atendem, com restrições conceituais, o enunciado. A essas, o testando deve ficar atento, pois elas têm como função verificar o pleno domínio de uma competência através do grau de discriminação

do item. Geralmente, os distratores são os erros e distrações mais comuns na abordagem de um conceito ou técnica relativos à competência avaliada.

### 3.3.2 Exemplo de item

A figura 5, a seguir, mostra a questão do ENEM de 2012 (BRASIL, 2012). Nela, pontuam-se os seus elementos constitutivos:

**Figura 5 – Exemplo de item**



Fonte: BRASIL, 2012.

Observe que o gráfico da questão relaciona a opinião "Chato" com 12%, fazendo da alternativa B um distrator com grande potencial, uma vez que o descritor é a alternativa D.

### 3.3.3 Etapas da elaboração de itens

No afã de se conseguir a máxima qualidade na elaboração de um item, é conveniente observar alguns fatores e contemplar alguns atributos, cujos detalhes estão explicitados nas etapas seguintes.

#### a) Atenção á matriz de referência: seleção de habilidade

As matrizes curriculares são, em última instância, listas de competências e habilidades consideradas indispensáveis na formação de um cidadão com proficiência acadêmica. Trata-se de um ideário de aquisições a se perseguir na prática pedagógica, não somente no momento da avaliação. Essa matriz de referência é a pretensão conceitual do item, como juízo de aquisição do sujeito avaliado. Essa matriz fornece o passo inicial da elaboração do item. De acordo com a Matriz de Referência do ENEM, tem-se, como eixos cognitivos:

I. **Dominar linguagens (DL)**: dominar a norma culta da Língua Portuguesa e fazer uso das linguagens matemática, artística e científica e das línguas espanhola e inglesa.

II. **Compreender fenômenos (CF)**: construir e aplicar conceitos das várias áreas do conhecimento para a compreensão de fenômenos naturais, de processos histórico-geográficos, da produção tecnológica e das manifestações artísticas.

III. **Enfrentar situações-problema (SP)**: selecionar, organizar, relacionar, interpretar dados e informações representados de diferentes formas, para tomar decisões e enfrentar situações-problema.

IV. **Construir argumentação (CA)**: relacionar informações, representadas em diferentes formas, e conhecimentos disponíveis em situações concretas, para construir argumentação consistente.

V. **Elaborar propostas (EP)**: recorrer aos conhecimentos desenvolvidos na escola para elaboração de propostas de intervenção solidária na realidade, respeitando os valores humanos e considerando a diversidade sociocultural. (GLOBO, 2014, p.1).

As **competências** são em número de cinco; cada uma ligada a um conjunto de habilidades contemplando cada eixo cognitivo. Veja-se como exemplo, o quadro 1 a seguir:

Quadro 1 – Competência para área 1 de Matemática e suas tecnologias

| EIXOS COGNITIVOS<br><br>COMPETÊNCIAS DE<br>MATEMÁTICA E SUAS<br>TECNOLOGIAS  | I - Dominar a norma culta da Língua Portuguesa e fazer uso das linguagens matemática, artística e científica.                            | II - Construir e aplicar conceitos das várias áreas do conhecimento para a compreensão de fenômenos naturais, de processos histórico-geográficos, da produção tecnológica e das manifestações artísticas. | III - Selecionar, organizar, relacionar, interpretar dados e informações representados de diferentes formas, para tomar decisões e enfrentar situações-problema.                     | IV - Relacionar informações, representadas em diferentes formas, e conhecimentos disponíveis em situações concretas, para construir argumentação consistente. | V - Recorrer aos conhecimentos desenvolvidos para elaboração de propostas de intervenção solidária na realidade, respeitando os valores humanos e considerando a diversidade sociocultural. |
|--|--|---|--|---|---|
| M1<br>Construir significados e ampliar os já existentes para os números naturais, inteiros, racionais e reais.                 | H1 - Utilizar no contexto social diferentes significados e representações dos números — naturais, inteiros, racionais ou reais.          | H2 - Utilizar algum procedimento de cálculo com números naturais, inteiros, racionais ou reais.   | H3 - Resolver situação-problema com números naturais, inteiros racionais ou reais envolvendo significados da adição, subtração, multiplicação ou divisão, potenciação ou radiciação. | H4 - Avaliar a razoabilidade de um resultado numérico na construção de argumentos sobre afirmações quantitativas.   | H5 - Avaliar propostas de intervenção na realidade, utilizando conhecimentos numéricos.   |
| M2<br>Utilizar o conhecimento geométrico para realizar a leitura e a representação da realidade e agir sobre ela.              | H6 - Interpretar a localização e a movimentação de pessoas/objetos no espaço tridimensional e sua representação no espaço bidimensional. | H7 - Identificar características de polígonos ou sólidos (prismas, pirâmides, cilindros).   | H8 - Resolver situação-problema que envolva noções geométricas (ângulo, paralelismo, perpendicularismo).   | H9 - Utilizar o teorema de Pitágoras ou semelhança de triângulos na seleção de argumentos propostos como solução de problemas do cotidiano.                   |   |
| M3<br>Construir e ampliar noções de grandezas e medidas para a compreensão da realidade e a solução de problemas do cotidiano. | H10 - Estabelecer relações entre diferentes unidades de medida (comprimento, massa, capacidade, área, volume).                           | H11 - Aplicar a noção de escalas na leitura de plantas ou mapas.  | H12 - Resolver situação-problema que envolva medidas de arcos ou ângulos (grau e radiano), utilizando teorema de Pitágoras ou razão trigonométrica (seno de um ângulo agudo).        | H13 - Avaliar a razoabilidade do resultado de uma medição, na construção de um argumento consistente.   | H14 - Avaliar propostas de intervenção na realidade utilizando cálculos de perímetros, área de superfícies planas ou volume de blocos retangulares.   |

| EIXOS COGNITIVOS<br><br>COMPETÊNCIAS DE MATEMÁTICA E SUAS TECNOLOGIAS  | I - Dominar a norma culta da Língua Portuguesa e fazer uso das linguagens matemática, artística e científica. | II - Construir e aplicar conceitos das várias áreas do conhecimento para a compreensão de fenômenos naturais, de processos histórico-geográficos, da produção tecnológica e das manifestações artísticas. | III - Selecionar, organizar, relacionar, interpretar dados e informações representados de diferentes formas, para tomar decisões e enfrentar situações-problema. | IV - Relacionar informações, representadas em diferentes formas, e conhecimentos disponíveis em situações concretas, para construir argumentação consistente.   | V - Recorrer aos conhecimentos desenvolvidos para elaboração de propostas de intervenção solidária na realidade, respeitando os valores humanos e considerando a diversidade sociocultural. |
|--|---|---|--|---|---|
| <b>M4</b><br>Construir e ampliar noções de variação de grandeza para a compreensão da realidade e a solução de problemas do cotidiano.   | H15<br>Identificar leis matemáticas que expressem a relação de dependência entre duas grandezas.              | _____   | H16<br>Resolver situação-problema envolvendo a variação de grandezas direta ou inversamente proporcionais.   | H17<br>Utilizar informações expressas em forma de juros (simples ou composto) como recurso para a construção de argumentação (aumentos e descontos sucessivos). | H18<br>Avaliar propostas de intervenção na realidade, utilizando cálculos de porcentagem e/ou juros.  |
| <b>M 5</b><br>Aplicar expressões algébricas para modelar e resolver problemas, envolvendo variáveis socioeconômicas ou técnico-científicas.  | H19 - Identificar representações algébricas que expressem a relação de interdependência entre duas grandezas. | H20 - Identificar gráfico cartesiano que represente a relação de interdependência entre duas grandezas (variação linear).   | H21 - Resolver situação-problema cujos dados estejam expressos em gráfico cartesiano que mostre a variação de duas grandezas.                                    | _____   | _____   |
| <b>M6</b><br>Interpretar informações de natureza científica e social obtidas da leitura de gráficos e tabelas, realizando previsão de tendência, extrapolação, interpolação e interpretação.   | H22 - Identificar informações apresentadas em tabelas ou gráficos (de coluna, de setores e de linha).         | H23 - Utilizar informações expressas em gráficos ou tabelas para fazer inferências.   | H24 - Resolver situação-problema com dados apresentados em forma de tabela de dupla entrada ou gráfico.  | H25 - Utilizar informações expressas em gráficos ou tabelas como recurso para a construção de argumentos.   | H26 - Avaliar propostas de intervenção na realidade, utilizando informações expressas em gráficos ou tabelas.   |
| <b>M7</b><br>Compreender o caráter aleatório e não-determinístico dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas e cálculos de probabilidade, para interpretar informações de variáveis apresentadas em uma distribuição estatística. | _____   | H27 - Calcular a média aritmética de um conjunto de dados expressos em uma tabela de freqüências de dados agrupados (não em classes) ou gráficos de colunas.  | H28 - Resolver situação-problema que envolva processos de contagem ou noções de probabilidade.   | H29 - Utilizar médias aritméticas, noção de probabilidade ou conhecimentos estatísticos como recurso para a construção de argumentação.                         | H30 - Avaliar propostas de intervenção na realidade utilizando probabilidade e/ou conhecimentos estatísticos (porcentagem, gráficos, médias).   |

Fonte: INEP, 2013, p.1-2.

**b) Definição da situação-problema e elaboração do texto-base**

A definição da situação-problema e a elaboração do texto-base devem ser feitas, levando-se em consideração os seguintes fatores:

- a realidade cotidiana do público-alvo;
- o texto-base deve ser o mais interessante e atrativo possível;
- usar contextualizações mais autênticas quanto o possível, aumentando-se o grau de realidade do que está sendo avaliado; para isso, é aconselhável trabalhar com atualidades de revistas, jornais e mídia em geral.

**c) Elaborar o enunciado, fornecendo todas as informações necessárias**

A estrutura do enunciado de um item pode ser de dois tipos:

I) O enunciado do item pode solicitar que o aluno, após a leitura do texto-base, resolva a situação-problema proposta para, em seguida, identificar a alternativa que contém a resposta certa.

II) O item pode solicitar ao aluno que analise cada alternativa, individualmente, de acordo com o enunciado (que pressupõe a leitura do texto-base), para identificar a correta.

Seja qual for o tipo escolhido, é importante redigir o enunciado de maneira clara e objetiva, procurando evitar que o aluno erre o item porque não compreendeu o que estava sendo perguntado. Eis alguns procedimentos que contribuem para a redação de enunciados:

→ A compreensão do enunciado deve ser independente da leitura de todas as alternativas.

→ O enunciado deve conter apenas dados ou informações funcionais. Expressões numéricas, textos, fórmulas, figuras, tabelas etc. só podem ser incluídos se forem necessários à resolução do problema proposto. Os elementos meramente ilustrativos ou que sejam desnecessários para a resolução da situação-problema podem resultar em baixos índices de acerto do item.

→ Termos impessoais devem predominar na redação dos enunciados. Exemplos: considere-se, calcula-se, argumenta-se etc.

→ Não se aconselha empregar termos negativos, como EXCETO, INCORRETO, NÃO, ERRADO, FALSO. Deve-se preferir a formulação de itens positivos.

→ Devem-se evitar termos generalizantes, como SEMPRE, NUNCA, TODO, TOTALMENTE, ABSOLUTAMENTE, COMPLETAMENTE, SOMENTE etc.

→ O item deve testar a habilidade desenvolvida pelo aluno e não a sua capacidade de memorização.

#### **d) Elaborar as alternativas de resposta**

Na elaboração das alternativas, convém considerar os seguintes procedimentos:

→ Elaborar alternativas que tenham a mesma estrutura (paralelismo sintático e semântico) e sejam igualmente bem construídas.

→ Elaborar alternativas cujas linhas de escrita tenham a mesma extensão, de modo que o descritor não possa ser identificado por esse atributo.

→ Adotar uma abordagem homogênea de conteúdo nas alternativas, ou seja, todas devem tratar da mesma categoria, mesma espécie, mesma abrangência, mesma extensão e mesma ordem de grandeza.

→ Incluir, no descritor, todas as informações necessárias à sua identificação segura.

→ Evitar detalhes irrelevantes e conteúdos absurdos. As informações devem ser coerentes com o enunciado e com o conteúdo, devendo ser apresentadas em linguagem correta.

→ Evitar a repetição de palavras presentes no enunciado.

→ Evitar informações demasiadamente longas.

→ Apresentar alternativas que forcem o testando a recorrer ao procedimento contido na habilidade proposta, impedindo que ele acerte por exclusão.

→ Ordenar as alternativas de maneira lógica (ordem alfabética, cronológica, crescente ou decrescente). Isso, além de facilitar a leitura, impedirá a identificação do descritor pela posição.

→ Não utilizar as expressões "todas as anteriores" ou "nenhuma das anteriores".

→ Evitar que o descritor seja mais atrativo que os distratores.

#### **e) Revisão do item**

Verificar se a habilidade em foco foi, de fato, contemplada; se há apenas um descritor e se a resposta de um item é independente da resposta dos outros.

#### **f) Resolução comentada do item**

Resolver o item, comentando a resolução, procurando justificar tanto o descritor quanto os distratores.

**g) Com relação aos textos utilizados como texto-base ou como alternativas**

Dar preferência a textos oriundos de fontes primárias, originais e sem adaptações, fazendo a devida referência à fonte de onde foi extraído. Todos os textos devem vir acompanhados da referência bibliográfica, dentro dos critérios da Associação Brasileira de Normas Técnicas (ABNT). Se o texto não for integral, deve-se identificá-lo como “fragmento”. Todo texto deve conter o título, ainda que seja “fragmento”. Se o texto sofrer qualquer alteração, deve-se identificá-lo como “adaptado”. O tempo de leitura exigido do aluno deve ser levado em consideração. Aproveitar textos extensos para cobrar mais de um item e, com isso, mais de uma habilidade.

**h) Com relação aos aspectos textuais que envolvem a elaboração do item**

Verificar a redação e a apresentação do item. Seguir as regras de ortografia, gramática e sintaxe da norma padrão. Cuidar para que os conceitos, fatos, terminologia etc. tenham caráter universal. Elaborar itens com vocabulário, objetos e situações conhecidos nacionalmente.

**i) Algumas recomendações complementares**

→ Se o enunciado for uma frase incompleta a ser corretamente completada pelas alternativas, essas devem começar com letras minúsculas e terminar com o ponto apropriado para a frase (ponto final, interrogação, exclamação etc.);

→ Caso o enunciado seja uma pergunta, deve terminar com uma interrogação e as alternativas devem começar com letras maiúsculas e terminar com ponto final;

→ Com palavras ou frases curtas incompletas, cada alternativa deve começar com letra maiúscula e não apresentar pontuação no final.

→ Com relação ao tema abordado no item, deve-se evitar:

- a escolha de temas ou a citação de elementos (afirmações, situações, usos, exemplos, vocabulário, objetos, informações etc.) que suscitem polêmicas ou que possam caracterizar-se como: o viés cultural ou discriminação e preconceito em relação a gêneros, etnias, profissões, crenças, religiões etc.;
- a apologia a comportamentos e condutas em desacordo com preceitos educativos e legais, por exemplo: drogas, bebidas, aborto, crime, arma ou

incitação à violência e a danos ou destruição de bem público ou privado e da natureza;

- a propaganda de produtos;
- a abordagem de temas controversos entre os próprios especialistas da área.

### **3.3.3.1 Exemplo de item tecnicamente correto**

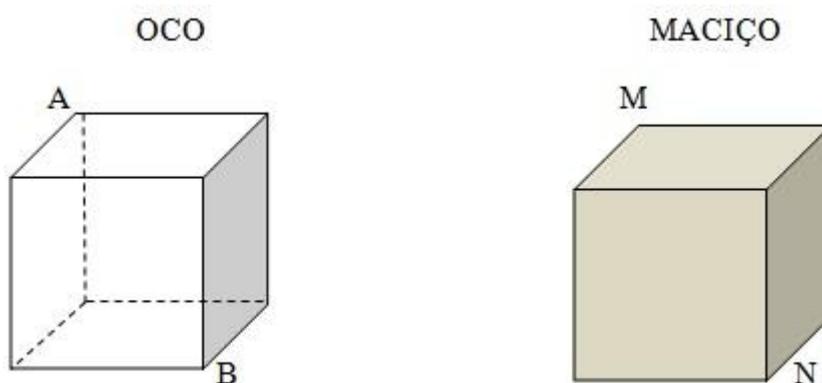
A seguir, tem-se um item que procura seguir as orientações destacadas anteriormente. Esse item em questão refere-se à área de Matemática e suas tecnologias, tendo, como eixo cognitivo, a compreensão de fenômenos e, como competência, usar o conhecimento geométrico para realizar a leitura e a representação da realidade e agir sobre ela. Ressalta-se que essa questão foi formulada pelo pesquisador.

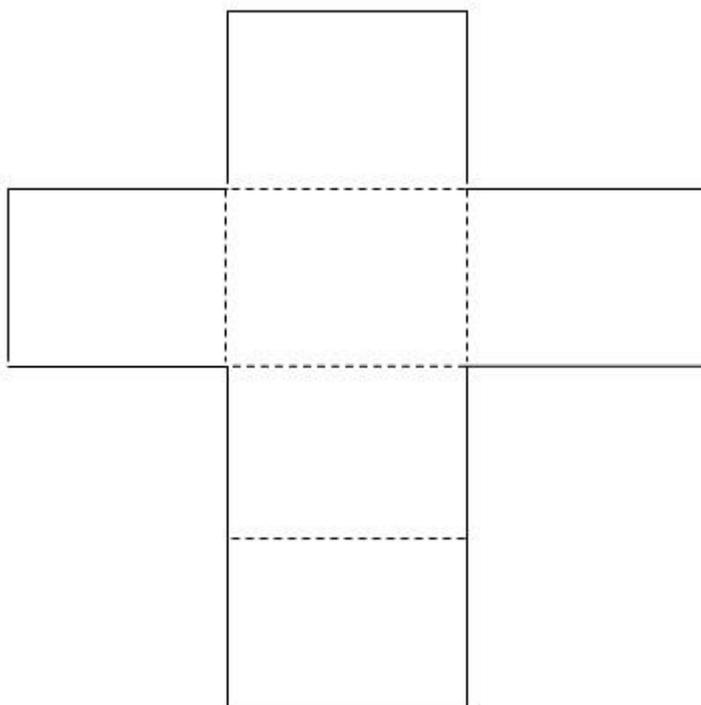
**Considere os seguintes POSTULADOS:**

*I – A eletricidade só circula num meio condutor diferente do ar.*

*II – A eletricidade desloca-se de um ponto a outro pela trajetória mais curta.*

Dois cubos de metal condutor são tais que um é maciço e o outro é oco, ambos com 3 metros de aresta, como nas figuras abaixo:





### PLANIFICAÇÃO DOS CUBOS

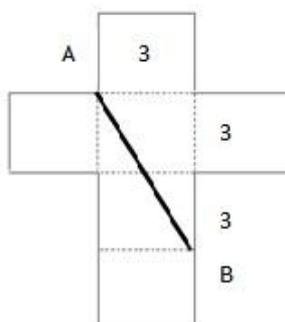
Nos vértices A e B do cubo OCO há, respectivamente, um interruptor e uma lâmpada ligada a ele. Do mesmo modo, nos vértices M e N do cubo MACIÇO, há, respectivamente, um interruptor e uma lâmpada ligada a ele. Em cada um dos casos, quando se liga o interruptor, uma corrente elétrica supostamente se desloca do interruptor para a lâmpada.

Sobre as trajetórias da corrente elétrica entre AB e entre MN, PODE-SE AFIRMAR que:

- (A) A trajetória da corrente elétrica entre AB tem o mesmo comprimento da trajetória da corrente elétrica entre MN.
- (B) A trajetória da corrente elétrica entre AB tem em torno de 1,5 m a mais que a trajetória da corrente elétrica entre MN.
- (C) A trajetória da corrente elétrica entre AB é o dobro da trajetória da corrente elétrica entre MN.
- (D) A trajetória da corrente elétrica entre AB tem 1 aresta a mais que a trajetória da corrente elétrica entre MN.
- (E) A trajetória da corrente elétrica entre AB tem 1 diagonal a mais que a trajetória da corrente elétrica entre MN.

→ **Resolução comentada:**

Como a eletricidade só circula em meio condutor, a trajetória da corrente elétrica entre MN é um segmento reto (diagonal do cubo) e medirá  $3\sqrt{3}$  m, aprox. 5,2 m. Quanto a trajetória da corrente elétrica entre AB, a menor distância assinalada nas faces do cubo, uma vez que só as faces são meios condutores, é aquela indicada na planificação do cubo (figura a seguir)



$AB = \sqrt{3^2 + 6^2} = \sqrt{45} = \text{aprox. } 6,7 \text{ m. Logo, } AB - MN = 1,5 \text{ m. Alternativa B.}$

Quanto aos distratores:

- (A) O testando que ignorar a diferença OCO/MACIÇO tende a afirmar que  $AB = MN$ .
- (C) O testando que confundir diagonal da face com aresta, em termos de tamanho pode ser levado a concluir que as diagonais da face e do sólido são congruentes.
- (D) mesmo equívoco anterior, (congruência da aresta e da diagonal do sólido).
- (E) mesmo equívoco da alternativa C (congruência da diagonal da face e da diagonal do sólido).

### 3.4 Informação do item

Uma medida bastante considerada, ao se analisar a CCI, é a informação que o item pode trazer acerca da habilidade  $\theta$ . No caso da Teoria Clássica, a habilidade é considerada constante, assim como o erro de medida. No caso da TRI, há valores de  $\theta$  que propiciam informações mais precisas sobre os parâmetros. Em geral, a função de informação do item é:

$$I_i(\theta_j) = \frac{[\frac{\partial}{\partial \theta_j} P(\theta_j)]^2}{P_i(\theta_j)Q_i(\theta_j)} \quad (23)$$

Em que:

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

O erro relativo de medida, tanto na Teoria Clássica quanto na TRI, é analiticamente medido pela dispersão ou variabilidade da habilidade e, quanto mais essa variabilidade tender a zero, mais precisos serão o parâmetros a, b e c.

Procedendo-se aos cálculos da derivada de  $P(\theta_j)$  para a versão de três parâmetros na expressão (23), tem-se:

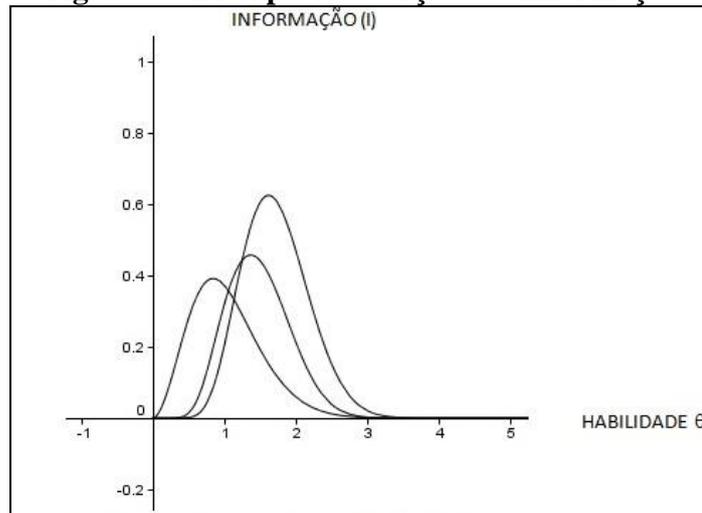
$$I_i(\theta_j) = \frac{2,89a_i^2(1-c_i)}{[c_i + \theta^{1,7}a_i(\theta-b_i)][1 + \theta^{1,7}a_i(\theta-b_i)]^2} \quad (24)$$

em que  $2,89 = D^2$  e  $D = 1,7$  (constante de adequação logística quando se considera a distribuição Normal).

Essa função mostra a influência dos parâmetros na quantidade de informação do item.

Em outras palavras, a informação é maior quando a dificuldade ( $b_i$ ) se aproxima de zero, e quanto maior for a discriminação ( $a_i$ ) do item, diminui a probabilidade do acaso ( $c_i$ ) para a habilidade  $\theta$ .

No gráfico da figura 6, estão representados 3 itens em suas funções de informação:

**Figura 6 - Exemplo de Funções de Informação**

Fonte: Elaborado pelo autor com o uso do software Geogebra

O gráfico mostra que os itens 1, 2 e 3 fornecem, individualmente, o máximo de informação quando a habilidade  $\theta$  é igual a 0,75, 1,5 e 1,8, respectivamente. Os itens 1 e 2 dão a mesma informação para a habilidade  $\theta = 1$ , Os itens 1 e 3 dão a mesma informação para a habilidade  $\theta = 1,2$  e os itens 2 e 3 dão a mesma informação para a habilidade  $\theta = 1,4$ . Entretanto, nenhum dos itens dá o total de informação possível, já que o máximo de informação é do item 3 (62%). Os itens que têm o parâmetro  $c$  negativo tendem a dar mais informação, sendo esse os casos de 2 e 3. Por outro lado, quanto maior o parâmetro  $b$ , maior é a informação.

### 3.5 Informação do teste

A informação do teste,  $I(\theta)$ , é a soma das informações de todos os itens do teste. Matematicamente, é obtida por:

$$I(\theta) = \sum_{i=1}^n I_i(\theta) \quad (25)$$

Pode-se calcular o erro relativo padrão da estimação (ERP) da habilidade, uma vez que ele é o inverso de  $I(\theta)$ , ou seja:

$$ERP(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (26)$$

## **4 ENSAIOS E EXPERIMENTOS**

A partir deste capítulo, será apresentada a aplicação de testes com as duas abordagens aqui descritas: A TCT e a TRI. Inicialmente, a TCT foi aplicada como um processo de triagem dos itens, aparentemente numa forma independente da TRI, mas, na sequência, a TRI foi aplicada a partir dos resultados da TCT, efetivando a calibragem final. Diante disso, tem-se que, apesar de serem consideradas independentes essas duas abordagens, no processo descrito, tornaram-se complementares. As amostras utilizadas são de alunos do sistema privado de ensino de Belo Horizonte, turmas de 3º Ano do Ensino Médio de 2014 (pré-teste/30 alunos da turma A e teste/ 30 alunos da turma B).

Tanto o pré-teste quanto o teste são constituídos de 20 itens. Esses itens, até a aplicação do pré-teste, eram inéditos e todos foram elaborados dentro das técnicas apresentadas no capítulo 3, contemplando habilidades da matriz de referência do ENEM.

### **4.1 Descrição e apresentação dos elementos dos ensaios**

Os elementos utilizados no exercício descrito a partir desse capítulo são os seguintes:

#### ***4.1.1 Amostras de alunos***

Para o pré- teste, foram selecionados, por amostragem aleatória simples, 30 alunos das turmas de 3º Ano do Ensino Médio, em novembro de 2014. A tabela 1 mostra os respectivos escores desses alunos no pré- teste, que possuía 20 itens (com 5 alternativas cada um).

**Tabela 1 - Amostra para o pré-teste**

| ALUNO | ESCORE/20 |
|-------|-----------|
| 001   | 9         |
| 002   | 9         |
| 003   | 9         |
| 004   | 10        |
| 005   | 10        |
| 006   | 10        |
| 007   | 10        |
| 008   | 10        |
| 009   | 10        |
| 010   | 11        |
| 011   | 11        |
| 012   | 11        |
| 013   | 12        |
| 014   | 12        |
| 015   | 12        |
| 016   | 12        |
| 017   | 12        |
| 018   | 12        |
| 019   | 13        |
| 020   | 13        |
| 021   | 13        |
| 022   | 13        |
| 023   | 13        |
| 024   | 14        |
| 025   | 14        |
| 026   | 14        |
| 027   | 14        |
| 028   | 14        |
| 029   | 15        |
| 030   | 16        |

Fonte: Dados da pesquisa

Com base nesses resultados acima obtidos, foi possível, então, fazer a calibragem dos itens para a execução do teste. Após a aplicação do teste, os escores são apresentados na tabela seguinte (TABELA 2):

Tabela 2 – Amostra para teste

| ALUNO | ESCORE/20 |
|-------|-----------|
| 001   | 11        |
| 002   | 11        |
| 003   | 12        |
| 004   | 12        |
| 005   | 12        |
| 006   | 13        |
| 007   | 14        |
| 008   | 15        |
| 009   | 15        |
| 010   | 15        |
| 011   | 15        |
| 012   | 15        |
| 013   | 15        |
| 014   | 16        |
| 015   | 18        |
| 016   | 8         |
| 017   | 8         |
| 018   | 10        |
| 019   | 10        |
| 020   | 10        |
| 21    | 11        |
| 22    | 11        |
| 023   | 11        |
| 024   | 12        |
| 025   | 13        |
| 026   | 13        |
| 027   | 13        |
| 028   | 14        |
| 029   | 14        |
| 030   | 16        |

Fonte: Dados da pesquisa.

#### 4.1.2 Conjunto de itens para pré-teste

As habilidades contempladas nos itens do pré-teste são mostradas no quadro 2 e identificadas por item aplicado. Os itens estão apresentados integralmente no apêndice 1 deste trabalho.

| <b>Quadro 2 - Habilidades dos itens</b> |  |                         | <b>Continua</b> |
|---|--|-------------------------|-----------------|
| <b>QUESTÃO<br/>(item)</b>               | <b>HABILIDADE</b>  | <b>CONTEÚDO</b>         | <b>RESP.</b>    |
| <b>01</b>                               | <b>H1</b> - Reconhecer, no contexto social, diferentes significados e representações dos números e operações com naturais, inteiros, racionais ou reais. | Sistema de Numeração    | D               |
| <b>02</b>                               | <b>H2</b> - Identificar padrões numéricos ou princípios de contagem.   | Sequências Numéricas    | B               |
| <b>03</b>                               | <b>H3</b> - Resolver situação-problema envolvendo conhecimentos numéricos.   | Números racionais       | D               |
| <b>04</b>                               | <b>H5</b> - Avaliar propostas de intervenção na realidade utilizando conhecimentos numéricos   | Operações com naturais  | D               |
| <b>05</b>                               | <b>H12</b> - Resolver situação-problema que envolva medida de grandeza.  | Medida de tempo         | A               |
| <b>06</b>                               | <b>H4</b> - Avaliar a razoabilidade de um resultado numérico na construção de argumentos sobre afirmações quantitativas.                                 | Matemática Financeira   | C               |
| <b>07</b>                               | <b>H6</b> - Interpretar e localizar a movimentação de pessoas/objetos no espaço tridimensional e sua representação no espaço bidimensional.              | Geometria Espacial      | B               |
| <b>08</b>                               | <b>H13</b> - Avaliar o resultado de uma medição na construção de um argumento consistente.   | Geometria Analítica     | B               |
| <b>09</b>                               | <b>H7</b> - Identificar características de figuras planas ou espaciais.  | Geometria Plana         | E               |
| <b>10</b>                               | <b>H11</b> - Utilizar a noção de escalas na leitura de representação de situação do cotidiano.   | Grandezas Proporcionais | A               |

## Conclusão

| QUESTÃO<br>(item) | HABILIDADE   | CONTEÚDO                | RESP. |
|-------------------|--|-------------------------|-------|
| 11                | H20 - Interpretar gráfico cartesiano que representa relações entre grandezas.  | Leitura de Gráficos     | D     |
| 12                | Identificar a relação de dependência entre grandezas   | Função quadrática       | B     |
| 13                | H16 – Resolver situação-problema envolvendo a variação de grandezas direta ou inversamente proporcionais.                          | Divisão<br>Proporcional | A     |
| 14                | H28 – Resolver situação problema que envolva conhecimentos de estatística e probabilidade  | Mediana e Média         | D     |
| 15                | H-15 – Identificar a relação de dependência entre grandezas  | Função afim             | B     |
| 16                | H8 - Resolver situação-problema que envolva conhecimentos geométricos de espaço e forma.   | Geometria sólida        | B     |
| 17                | H24 - Utilizar informações expressas em gráficos ou tabelas para fazer inferências.  | Estatística             | D     |
| 18                | H9 - Utilizar conhecimentos geométricos de espaço e forma na seleção de argumentos propostos na solução de problemas do cotidiano. | Geometria plana         | A     |
| 19                | H28 - Resolver situação problema que envolva conhecimentos de estatística e probabilidade.   | Probabilidade           | E     |
| 20                | H23 - Avaliar propostas de intervenção na realidade utilizando conhecimentos algébricos.   | Função Quadrática       | B     |

Fonte: Elaborado pelo autor.

A tabela 3 apresenta os resultados obtidos após a aplicação e correção do pré-teste, para a amostra de alunos considerada.

**Tabela 3 - Gabarito e acertantes**

| ITEM | GABARITO | ACERTANTES |
|------|----------|------------|
| 001  | D        | 25         |
| 002  | B        | 26         |
| 003  | D        | 3          |
| 004  | D        | 15         |
| 005  | A        | 25         |
| 006  | C        | 15         |
| 007  | B        | 2          |
| 008  | B        | 17         |
| 009  | E        | 23         |
| 010  | A        | 17         |
| 011  | D        | 26         |
| 012  | B        | 15         |
| 013  | A        | 5          |
| 014  | D        | 15         |
| 015  | B        | 16         |
| 016  | B        | 13         |
| 017  | D        | 6          |
| 018  | A        | 26         |
| 019  | E        | 8          |
| 020  | B        | 6          |

**Fonte: Dados da pesquisa.**

O Índice de dificuldade de cada item é mostrado na tabela 4. A classificação de Condé dos itens por dificuldade  $I_d$  considera o item FÁCIL, se  $I_d > 0,70$ ; MÉDIO, se  $0,30 < I_d \leq 0,70$  e DIFÍCIL, se  $I_d < 0,30$ . (PASQUALI, 1996).

Tabela 4 - Índice de dificuldade por item

| ITEM | DIFIC. | CLASSIFIC. |
|------|--------|------------|
| 001  | 0,833  | FÁCIL      |
| 002  | 0,867  | FÁCIL      |
| 003  | 0,100  | DIFÍCIL    |
| 004  | 0,500  | MÉDIA      |
| 005  | 0,833  | FÁCIL      |
| 006  | 0,500  | MÉDIA      |
| 007  | 0,067  | DIFÍCIL    |
| 008  | 0,567  | MÉDIA      |
| 009  | 0,767  | FÁCIL      |
| 010  | 0,567  | MÉDIA      |
| 011  | 0,867  | FÁCIL      |
| 012  | 0,500  | MÉDIA      |
| 013  | 0,167  | DIFÍCIL    |
| 014  | 0,500  | MÉDIA      |
| 015  | 0,533  | MÉDIA      |
| 016  | 0,433  | MÉDIA      |
| 17   | 0,200  | DIFÍCIL    |
| 018  | 0,867  | FÁCIL      |
| 019  | 0,267  | DIFÍCIL    |
| 020  | 0,200  | DIFÍCIL    |

Fonte: Dados da pesquisa.

A tabela 5, por sua vez, mostra a distribuição das respostas por alternativa.

Tabela 5 - Distribuição das respostas por alternativa

| ALUNOS →  |   | 001 | 002 | 003 | 004 | 005 | 006 | 007 | 008 | 009 | 010 | 011 | 012 | 013 | 014 |     |     |
|---|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ACERTOS →   |   | 11  | 11  | 12  | 12  | 12  | 13  | 14  | 15  | 15  | 15  | 15  | 15  | 15  | 16  |     |     |
| NÚMERO<br>DE RESPOSTAS<br>POR CADA<br>ALTERNATIVA<br>DADA | A | 4   | 4   | 4   | 4   | 5   | 3   | 6   | 4   | 6   | 7   | 7   | 8   | 6   | 8   |     |     |
|   | B | 8   | 4   | 4   | 6   | 4   | 6   | 3   | 6   | 5   | 5   | 3   | 6   | 4   | 3   |     |     |
|   | C | 1   | 4   | 6   | 4   | 4   | 3   | 5   | 4   | 3   | 4   | 2   | 2   | 4   | 4   |     |     |
|   | D | 5   | 5   | 3   | 5   | 4   | 5   | 2   | 4   | 4   | 3   | 5   | 3   | 3   | 2   |     |     |
|   | E | 2   | 3   | 3   | 1   | 3   | 3   | 4   | 2   | 2   | 1   | 3   | 1   | 3   | 3   |     |     |
| ALUNOS →  |   | 015 | 016 | 017 | 018 | 019 | 020 | 021 | 022 | 023 | 024 | 025 | 026 | 027 | 028 | 029 | 030 |
| ACERTOS →   |   | 18  | 8   | 8   | 10  | 10  | 10  | 11  | 11  | 11  | 12  | 13  | 13  | 13  | 14  | 14  | 16  |
| NÚMERO<br>DE RESPOSTAS<br>POR CADA<br>ALTERNATIVA<br>DADA | A | 4   | 6   | 3   | 7   | 7   | 10  | 8   | 9   | 5   | 5   | 6   | 5   | 4   | 6   | 3   | 2   |
|   | B | 5   | 3   | 4   | 2   | 4   | 3   | 5   | 4   | 5   | 6   | 4   | 5   | 6   | 6   | 6   | 6   |
|   | C | 4   | 4   | 7   | 3   | 4   | 2   | 1   | 4   | 2   | 1   | 1   | 2   | 3   | 3   | 4   | 3   |
|   | D | 5   | 3   | 3   | 4   | 2   | 3   | 5   | 3   | 6   | 5   | 6   | 3   | 3   | 2   | 3   | 5   |
|   | E | 2   | 4   | 3   | 4   | 3   | 2   | 1   | 0   | 2   | 3   | 3   | 5   | 4   | 3   | 4   | 4   |

Fonte: Dados da pesquisa.

### 4.1.3 Primeira análise

Conforme a classificação de Condé (PASQUALI, 1996), o presente pré-teste, em nível de dificuldade, tem 40% de itens de nível médio, 30% de nível fácil e 30% de nível difícil. Por sua vez, Pasquali (1996) sugere como ideal a composição de um teste com 10% de questões muito fáceis, 20% de questões fáceis, 40% de questões médias, 20% de questões difíceis e 10% de questões muito difíceis. Nesse caso, dever-se-ia ter 2 (duas) questões muito fáceis, 4 (quatro) questões fáceis, 8 (oito) questões médias, 4 (quatro) questões difíceis e 2 (duas) questões muito difíceis.

Em acordo com Pasquali (1996), portanto, sugere-se a seguinte distribuição de dificuldade (TABELA 6). Essa mesma tabela indica, ainda, a quantidade e os itens correspondentes a cada classificação:

**Tabela 6 - Classificação do pré-teste**

| CLASSIFICAÇÃO  | PASQUALI               | Nº DE QUESTÕES | ITENS CORRESP.                          |
|----------------|------------------------|----------------|---|
| MUITO DIFÍCEIS | $0,00 < I_d \leq 0,10$ | 2              | 003 E 007                               |
| DIFÍCEIS       | $0,10 < I_d \leq 0,30$ | 4              | 013, 017, 019 e 020                     |
| MÉDIAS         | $0,30 < I_d \leq 0,70$ | 8              | 004, 006, 008, 010, 012, 014, 015 e 016 |
| FÁCEIS         | $0,70 < I_d \leq 0,90$ | 6              | 001, 002, 005, 009, 011 e 018           |
| MUITO FÁCEIS   | $0,90 < I_d \leq 1,00$ | 0              | Nenhum                                  |

Fonte: Dados da pesquisa

### 4.1.4 Medidas de posição e medidas de dispersão:

Numa distribuição com classes, os escores ficaram com as seguintes referências de posição e de dispersão identificadas na tabela 7:

Tabela 7 - Distribuição dos escores

| ACERTOS         |       | f  | Xm    | xM.f  | Fa     | f.ΔX | f.ΔX <sup>2</sup> |        |
|-----------------|-------|----|-------|-------|--------|------|-------------------|--------|
| 8               | └     | 11 | 5     | 9,50  | 47,50  | 5    | 19,00             | 72,20  |
| 11              | └     | 14 | 13    | 12,50 | 162,50 | 18   | 10,40             | 8,32   |
| 14              | └     | 17 | 11    | 15,50 | 170,50 | 29   | 24,20             | 53,24  |
| 17              | └     | 20 | 1     | 18,50 | 18,50  | 30   | 5,20              | 27,04  |
| <b>TOTAIS →</b> |       | 30 |       |       | 399,00 |      | 58,80             | 160,80 |
| Med =           | 13,30 |    | DM =  | 1,96  |        |      |                   |        |
| Mo =            | 14,43 |    | Var = | 5,55  |        |      |                   |        |
| Md =            | 13,31 |    | s =   | 2,36  |        |      |                   |        |

Fonte: Dados da pesquisa.

Tem-se que, na tabela 7,  $f$  é a frequência absoluta de uma classe genérica,  $X_m$  é o ponto médio dessa classe,  $F_a$  é a frequência acumulada até essa classe,  $Med$  é a média da distribuição,  $Mo$  é a sua moda,  $Md$  a sua mediana,  $DM$  é o seu desvio médio,  $Var$  é a sua variância,  $s$  é o seu desvio padrão e  $\Delta = |X_m - Med|$ .

Por meio da média e do desvio-padrão calculados na tabela 7, obtém-se, no cálculo da probabilidade de uma distribuição normal, a probabilidade aproximada de um aluno estar entre dois escores de cada faixa. (TABELA 8).

Tabela 8- Distribuição normal dos escores

| Acertos X     | Int. Padrão     | Cálculo     | Prob. |
|---------------|-----------------|-------------|-------|
| [9 , 11[      | [-2,46 ; -1,41[ | 0,49 - 0,42 | 0,07  |
| [11 , 13[     | [-1,41 ; -0,35[ | 0,42 - 0,14 | 0,28  |
| [13 , 15[     | [-0,35 ; 0,70[  | 0,26 + 0,14 | 0,40  |
| [15 , 17[     | [0,70 ; 1,75[   | 0,46 - 0,26 | 0,20  |
| <b>TOTAIS</b> |                 |             |       |

Fonte: Dados da pesquisa.

Dessa maneira, Considerando-se o desvio padrão de 1,9 e a média de 12,33, a distribuição normal dos escores leva à conclusão de que o aluno, para atingir um escore de, no mínimo, 65% (escore acima de 12), ele deve ser um dos 60% melhores alunos desse grupo.

#### 4.1.5 Avaliação dos parâmetros dificuldade e discriminação

A avaliação dos parâmetros de dificuldade e de discriminação dos itens passa inevitavelmente por correlações. A primeira delas é a de normalidade, que é a verificação de quanto a medida se aproxima de uma distribuição Normal e está indicada na tabela 9:

**Tabela 9 - Cálculo da correlação de normalidade**

| ITEM i | X      | X - MÉDIA | m <sub>i</sub> | REF/Z | M <sub>i</sub> | (X - MÉDIA) <sup>2</sup> | (M <sub>i</sub> - MÉDIA) <sup>2</sup> | (X - MÉDIA)(M <sub>i</sub> - MÉDIA) |
|--------|--------|-----------|----------------|-------|----------------|--------------------------|---------------------------------------|-------------------------------------|
| 1,000  | 25,000 | 9,800     | 0,034          | 0,466 | -1,825         | 96,040                   | 3,345                                 | 17,924                              |
| 2,000  | 26,000 | 10,800    | 0,083          | 0,417 | -1,385         | 116,640                  | 1,929                                 | 15,001                              |
| 3,000  | 3,000  | -12,200   | 0,132          | 0,368 | -1,115         | 148,840                  | 1,252                                 | 13,652                              |
| 4,000  | 15,000 | -0,200    | 0,181          | 0,319 | -0,910         | 0,040                    | 0,835                                 | 0,183                               |
| 5,000  | 25,000 | 9,800     | 0,230          | 0,270 | -0,740         | 96,040                   | 0,554                                 | 7,291                               |
| 6,000  | 15,000 | -0,200    | 0,279          | 0,221 | -0,585         | 0,040                    | 0,347                                 | 0,118                               |
| 7,000  | 2,000  | -13,200   | 0,328          | 0,172 | -0,455         | 174,240                  | 0,211                                 | 6,059                               |
| 8,000  | 17,000 | 1,800     | 0,377          | 0,123 | -0,315         | 3,240                    | 0,102                                 | 0,574                               |
| 9,000  | 23,000 | 7,800     | 0,426          | 0,074 | -0,185         | 60,840                   | 0,036                                 | 1,474                               |
| 10,000 | 17,000 | 1,800     | 0,475          | 0,025 | -0,065         | 3,240                    | 0,005                                 | 0,124                               |
| 11,000 | 26,000 | 10,800    | 0,525          | 0,025 | 0,065          | 116,640                  | 0,004                                 | 0,659                               |
| 12,000 | 15,000 | -0,200    | 0,574          | 0,074 | 0,185          | 0,040                    | 0,033                                 | 0,036                               |
| 13,000 | 5,000  | -10,200   | 0,623          | 0,123 | 0,415          | 104,040                  | 0,169                                 | 4,192                               |
| 14,000 | 15,000 | -0,200    | 0,672          | 0,172 | 0,445          | 0,040                    | 0,194                                 | 0,088                               |
| 15,000 | 16,000 | 0,800     | 0,721          | 0,221 | 0,555          | 0,640                    | 0,304                                 | 0,441                               |
| 16,000 | 13,000 | -2,200    | 0,770          | 0,270 | 0,745          | 4,840                    | 0,549                                 | 1,630                               |
| 17,000 | 6,000  | -9,200    | 0,819          | 0,319 | 0,915          | 84,640                   | 0,830                                 | 8,381                               |
| 18,000 | 26,000 | 10,800    | 0,868          | 0,368 | 1,115          | 116,640                  | 1,234                                 | 11,999                              |
| 19,000 | 8,000  | -7,200    | 0,917          | 0,417 | 1,390          | 51,840                   | 1,921                                 | 9,979                               |
| 20,000 | 6,000  | -9,200    | 0,966          | 0,466 | 1,835          | 84,640                   | 3,353                                 | 16,845                              |
|        |        |           |                |       | 0,004          | 1263,200                 | 17,206                                | 116,651                             |

Fonte: Dados da pesquisa.

Para o presente caso, a tabela anterior apresenta os valores calculados e, posteriormente, as fórmulas 27 e 28 fornecem uma correlação igual a 0,897, o que garante que o escore é normalmente distribuído, segundo o critério de Filliben, em 1975 (ANDRADE, TAVARES, VALLE, 2000), o qual adota a utilização da mediana ao invés da média.

Dessa maneira, o cálculo dessa correlação é proposto pela fórmula:

$$\hat{\rho} = \frac{\sum_{i=1}^n (x_i - \bar{x})(M_i - \bar{M})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (M_i - \bar{M})^2}} \quad (27)$$

Os valores de  $M_i$  são os inversos da função densidade da distribuição  $N(0,1)$  de  $m_i$ .

$$m_i = \begin{cases} 1 - m_n & \text{para } i = 1 \\ (i - 0,3175)/(n + 0,365) & \text{para } i = 2, 3, \dots, n - 1 \\ 0,5^{1/n} & \text{para } i = n \end{cases} \quad (28)$$

Um critério muito utilizado por especialistas para avaliar um item ideal diz que um item é ideal se metade da turma o acerta e a outra metade o erra (ESPÍRITO SANTO, 1978). De acordo com essa ideia, os itens de nível médio são os ideais, ou seja, na faixa de  $0,40 < I_d < 0,60$ , o que, neste presente caso, corresponde a, aproximadamente, 50% dos itens.

Outra abordagem bastante útil na análise de itens é o levantamento percentual dos que escolheram cada alternativa em cada item (SANTOS; SOARES, 2012). Esse olhar viabiliza, não somente a análise de atratividade da alternativa como distrator ou resposta correta, como, também, o padrão de resposta dos testandos.

A tabela 10 mostra esse levantamento no presente caso. Entretanto, novas considerações devem ser feitas, particularmente, no tocante à atratividade das alternativas, sobretudo os distratores. Veja-se, pois, o índice de atração de cada distrator (TAB.10), que indica, também, a média das diferenças de atratividade, em cada item, das alternativas em relação à resposta correta. Por exemplo, no item 1, a resposta correta é D, com índice de atratividade 0,833 e média das diferenças dos índices de atratividade de cada distrator em relação ao da resposta correta de 0,791:

**Tabela 10 - Percentual de cada alternativa por item**

|                     |       |       |       |       |       |       |       |       |       |       |
|---------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| <b>ITENS →</b>      | 001   | 002   | 003   | 004   | 005   | 006   | 007   | 008   | 009   | 010   |
| <b>A</b>            | 0,067 | 0,000 | 0,400 | 0,033 | 0,833 | 0,033 | 0,167 | 0,200 | 0,100 | 0,567 |
| <b>B</b>            | 0,033 | 0,867 | 0,000 | 0,233 | 0,067 | 0,133 | 0,067 | 0,567 | 0,133 | 0,000 |
| <b>C</b>            | 0,067 | 0,067 | 0,333 | 0,200 | 0,100 | 0,500 | 0,533 | 0,067 | 0,000 | 0,000 |
| <b>D</b>            | 0,833 | 0,067 | 0,100 | 0,500 | 0,000 | 0,100 | 0,167 | 0,000 | 0,000 | 0,200 |
| <b>E</b>            | 0,000 | 0,000 | 0,167 | 0,033 | 0,000 | 0,233 | 0,067 | 0,167 | 0,767 | 0,233 |
| <b>DESCRITORES</b>  | D     | B     | D     | D     | A     | C     | B     | B     | E     | A     |
| <b>ATRAT. DESC.</b> | 0,833 | 0,867 | 0,100 | 0,500 | 0,833 | 0,500 | 0,067 | 0,567 | 0,767 | 0,567 |
| <b>DIF.MÉD/ALT</b>  | 0,791 | 0,834 | 0,175 | 0,375 | 0,791 | 0,375 | 0,167 | 0,459 | 0,709 | 0,459 |
| <b>ITENS →</b>      | 011   | 012   | 013   | 014   | 015   | 016   | 017   | 018   | 019   | 020   |
| <b>A</b>            | 0,000 | 0,267 | 0,167 | 0,167 | 0,467 | 0,333 | 0,133 | 0,867 | 0,400 | 0,333 |
| <b>B</b>            | 0,133 | 0,500 | 0,167 | 0,133 | 0,533 | 0,433 | 0,500 | 0,000 | 0,000 | 0,200 |
| <b>C</b>            | 0,000 | 0,000 | 0,300 | 0,200 | 0,000 | 0,233 | 0,000 | 0,000 | 0,200 | 0,467 |
| <b>D</b>            | 0,867 | 0,000 | 0,133 | 0,500 | 0,000 | 0,000 | 0,200 | 0,000 | 0,133 | 0,000 |
| <b>E</b>            | 0,000 | 0,233 | 0,233 | 0,000 | 0,000 | 0,000 | 0,167 | 0,133 | 0,267 | 0,000 |
| <b>DESCRITORES</b>  | D     | B     | A     | D     | B     | B     | D     | A     | E     | B     |
| <b>ATRAT. DESC.</b> | 0,867 | 0,500 | 0,167 | 0,500 | 0,533 | 0,433 | 0,200 | 0,867 | 0,267 | 0,200 |
| <b>DIF.MÉD/ALT</b>  | 0,834 | 0,375 | 0,058 | 0,375 | 0,416 | 0,291 | 0,150 | 0,834 | 0,150 | 0,200 |

Fonte: Dados da pesquisa.

A tabela anterior mostra as taxas percentuais de atração de cada alternativa por item. Esses números levantam reflexões do tipo: Por que nem sempre o descritor é o mais atrativo? E isto é um fator relevante para avaliar a elaboração das alternativas dos itens.

Na tabela 10, como visto, há alternativas, em alguns itens, que não atraíram nenhum dos alunos do pré-teste. Podem-se alterar esses números, talvez de modo inesperado, mas com um relativo controle, mudando o enunciado de uma alternativa, acentuando o seu papel de descritor ou distrator. A capacidade de discriminação de um item é conceitualmente maior quanto mais próximas forem as taxas de atratividade.

A diferença média, em cada item, está indicada na última linha e o desvio global é de 0,441. Por outro lado, a maior diferença acontece nos itens 11 e 18 (0,834) e a menor, no item 13 (0,058). Em todos os itens, os descritores são mais atraentes que os distratores. Entretanto, nos itens em que há uma proximidade entre os percentuais das alternativas, o fator discriminação fica mais acentuado.

A medida apropriada para a discriminação é a Correlação Bisserial, que é dada pela fórmula:

$$\hat{\rho} = \frac{(X_p - \bar{X}_t) \cdot \hat{\theta}}{St \cdot Y} \quad (29)$$

Em que  $X_p$  é a média dos escores de quem acertou o item;  $\bar{X}_t$  é a média global dos escores;  $St$  é o desvio padrão dos escores e  $Y$  é a ordenada relativa ao índice de dificuldade na curva Normal. Assim, tem-se a correlação bisserial  $\hat{\rho}$  por item (TAB.11). Os cálculos foram feitos com o *software* BILOG-MG.

A tabela 11 mostra a correlação bisserial aplicada aos itens. A medida de correlação bisserial de um item situa-o no âmbito do teste como um todo. O coeficiente de correlação bisserial avalia como o escore total do teste se relaciona com uma variável latente supostamente responsável pelo desempenho do avaliado. Os especialistas admitem como valor mínimo de referência para a correlação bisserial 0,4, desde que seja garantida uma razoável discriminação do item, ou seja, a variância devida à resposta ao item, que é o quadrado do coeficiente bisserial, seja, no mínimo, 12% no total de escores (ESPÍRITO SANTO, 1978). O *software* BILOG- MG, que usa o algoritmo EM para calibragem dos itens, garante a calibragem, desde que o item apresente uma correlação bisserial maior do que -0,15. Nesse critério, não se inserem apenas os itens 008 e 014, os quais poderiam ser excluídos.

Tabela 11 - Correlação bisserial

|                           |       |       |       |        |        |        |       |        |       |        |       |
|---------------------------|-------|-------|-------|--------|--------|--------|-------|--------|-------|--------|-------|
| ITEM DE NÚMERO →          | 001   | 002   | 003   | 004    | 005    | 006    | 007   | 008    | 009   |        |       |
| COEF. BISSER. r →         | 0,078 | 0,310 | 0,681 | 0,398  | 0,130  | 0,435  | 0,288 | -0,181 | 0,197 |        |       |
| VALOR DE r <sup>2</sup> → | 0,006 | 0,096 | 0,464 | 0,158  | 0,017  | 0,189  | 0,083 | 0,033  | 0,039 |        |       |
| ITEM DE NÚMERO →          | 010   | 011   | 012   | 013    | 014    | 015    | 016   | 017    | 018   | 019    | 020   |
| COEF. BISSER. r →         | 0,325 | 0,633 | 0,150 | -0,043 | -0,205 | -0,052 | 0,572 | 0,455  | 0,633 | -0,037 | 0,207 |
| VALOR DE r <sup>2</sup> → | 0,106 | 0,401 | 0,023 | 0,002  | 0,042  | 0,003  | 0,327 | 0,207  | 0,401 | 0,001  | 0,043 |

Fonte: Dados da pesquisa.

Desde o início do presente exercício, tem-se usado a Teoria Clássica dos Testes para extrair informações e analisar os componentes de uma avaliação de peso. Portanto, a TCT não é dispensável nessa primeira fase, muito ao contrário; ela atua em uma relação de complementaridade juntamente com a Teoria de Resposta ao Item. Procura-se, antes de uma calibragem definitiva, executada com a TRI, diminuir ao máximo os problemas que alguns itens trazem e são revelados no pré-teste, como, por exemplo, erros nas alternativas.

A fórmula do **Coefficiente de Correlação de Pearson** é também usada para avaliar a fidedignidade do teste, verificando a correlação entre as variáveis X (percentuais de escores de metade da amostra ordenada) e Y (percentuais de escores da outra metade da amostra ordenada). Então, tem-se:

$$r = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n \cdot \sum X_i^2 - (\sum X)^2] \cdot [n \cdot \sum Y_i^2 - (\sum Y)^2]}} \quad (30)$$

A tabela 12 mostra os fatores constantes na fórmula 30 e o respectivo coeficiente de correlação  $\rho$  resultante. Há milhares de composições possíveis dos conjuntos de variáveis X e Y, com 15 elementos cada um. Ordenando-se crescentemente os escores e selecionando-se, nesta ordem, um grupo após o outro, a correlação linear é bastante boa (de 90% para cima).

**Tabela 12 - Fatores do coeficiente de Pearson**

| GRUPO I (x) | GRUPO II (Y) | XY       | X <sup>2</sup> | Y <sup>2</sup> |
|-------------|--------------|----------|----------------|----------------|
| 8,000       | 13,000       | 104,000  | 64,000         | 169,000        |
| 8,000       | 13,000       | 104,000  | 64,000         | 169,000        |
| 10,000      | 13,000       | 130,000  | 100,000        | 169,000        |
| 10,000      | 14,000       | 140,000  | 100,000        | 196,000        |
| 10,000      | 14,000       | 140,000  | 100,000        | 196,000        |
| 11,000      | 14,000       | 154,000  | 121,000        | 196,000        |
| 11,000      | 15,000       | 165,000  | 121,000        | 225,000        |
| 11,000      | 15,000       | 165,000  | 121,000        | 225,000        |
| 11,000      | 15,000       | 165,000  | 121,000        | 225,000        |
| 11,000      | 15,000       | 165,000  | 121,000        | 225,000        |
| 12,000      | 15,000       | 180,000  | 144,000        | 225,000        |
| 12,000      | 15,000       | 180,000  | 144,000        | 225,000        |
| 12,000      | 16,000       | 192,000  | 144,000        | 256,000        |
| 12,000      | 16,000       | 192,000  | 144,000        | 256,000        |
| 13,000      | 18,000       | 234,000  | 169,000        | 324,000        |
| 162,000     | 221,000      | 2410,000 | 1778,000       | 3281,000       |

Fonte: Dados da pesquisa.

Pela fórmula de Pearson, tem-se,  $r = \frac{36.150 - 35.802}{\sqrt{(426).(374)}} = 0,872$ . Mais 5 combinações

foram experimentadas e tiveram resultados similares. Então, em termos de fidedignidade, o teste, como um todo, é fidedigno. No entanto, o BILOG-MG, que faz uma estimativa individual dos itens, assinala o coeficiente de Pearson por item (TABELA 13).

**Tabela 13 - Coeficientes de Pearson (influência do item)**

| ITENS →                  | 001   | 002   | 003   | 004   | 005   | 006   | 007   | 008    | 009   |
|--------------------------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| COEFICIENTE DE PEARSON → | 0,052 | 0,196 | 0,398 | 0,317 | 0,087 | 0,347 | 0,149 | -0,143 | 0,143 |

| ITENS →                  | 010   | 011   | 012   | 013    | 014    | 015    | 016   | 017   | 018   | 019    | 020   |
|--------------------------|-------|-------|-------|--------|--------|--------|-------|-------|-------|--------|-------|
| COEFICIENTE DE PEARSON → | 0,258 | 0,401 | 0,120 | -0,029 | -0,163 | -0,041 | 0,454 | 0,318 | 0,401 | -0,027 | 0,145 |

Fonte: Dados da pesquisa.

Ainda para verificar a consistência e fidedignidade do presente teste, foi usada a fórmula 20 de Kuder-Richardson:

$$\rho_{KR} = \frac{n}{n-1} \left( 1 - \frac{\sum_{i=1}^n \theta_i (1 - \theta_i)}{S^2} \right) \quad (31)$$

Em que:

$\theta_i$  é a proporção dos alunos que acertaram o item  $i$ ;

$n$  é o número de itens do teste;

$S^2$  é a variância dos escores.

Então, tem-se, para o presente caso:  $\rho_{KR} = \frac{20}{20-1} \left(1 - \frac{3,66}{71,70}\right) = 0,998$ . Como o valor mínimo

de referência é em torno de 0,9, o teste em questão é um conjunto de dados homogêneo.

Outra abordagem para melhorar a performance do teste antes efetuar uma calibragem pela TRI, é a sugerida por Espírito Santo (1987), o qual diz que se o teste discrimina, o aluno que marca as alternativas incorretas, em pelo menos 50% das vezes, está abaixo da média. A tabela 14 mostra, no presente teste, a situação dos escores globais em relação à sua média dos escores. Observa-se que todos os alunos erraram alguma resposta; entretanto, em torno de 60% deles têm os escores abaixo da média, que é 12,33. De acordo com Espírito Santo (1987), há, no presente teste, uma quantidade de itens que têm boa discriminação. Entretanto, há alternativas com frequência nula, ou seja, sem atratividade. Um bom critério, de alta eficácia para compor um teste é dosá-lo segundo a sugestão de Pasquali (1996), sendo, no presente caso, 2 itens muito difíceis, 4 itens difíceis, 8 itens de dificuldade média, 4 itens fáceis e 2 muito fáceis e, como visto, há uma distribuição natural dos escores próxima dessa referência.

**Tabela 14 - Escore global e média de escores**

| ALUNOS | ESCORES | SITUAÇÃO | ALUNOS | ESCORES | SITUAÇÃO | ALUNOS | ESCORES | SITUAÇÃO |
|--------|---------|----------|--------|---------|----------|--------|---------|----------|
| 001    | 9       | abaixo   | 011    | 11      | abaixo   | 021    | 13      | acima    |
| 002    | 9       | abaixo   | 012    | 11      | abaixo   | 022    | 13      | acima    |
| 003    | 9       | abaixo   | 013    | 12      | abaixo   | 023    | 13      | acima    |
| 004    | 10      | abaixo   | 014    | 12      | abaixo   | 024    | 14      | acima    |
| 005    | 10      | abaixo   | 015    | 12      | abaixo   | 025    | 14      | acima    |
| 006    | 10      | abaixo   | 016    | 12      | abaixo   | 026    | 14      | acima    |
| 007    | 10      | abaixo   | 017    | 12      | abaixo   | 027    | 14      | acima    |
| 008    | 10      | abaixo   | 018    | 12      | abaixo   | 028    | 14      | acima    |
| 009    | 10      | abaixo   | 019    | 13      | acima    | 029    | 15      | acima    |
| 010    | 11      | abaixo   | 020    | 13      | acima    | 030    | 16      | acima    |

Fonte: Dados da pesquisa

#### 4.1.6 *Triagem Final dos itens*

Considerando-se os três critérios decorrentes da análise anterior; tem-se os seguintes apontamentos:

1º) Os itens com correlação bisserial a partir de 0,4 são preferidos, mas, segundo o BILOG-MG, apenas os itens com essa correlação inferior a -0,15 não são calibráveis. Nesse quesito, os itens 008 e 014 encontram-se em desacordo com a regra.

2º) O cruzamento da variância x correlação bisserial, mostra que há problemas com vários itens.

3º) A atratividade nula, que aparece na maioria dos itens, pode ser uma tendência natural da amostra ou uma característica estrutural da questão.

A atratividade nula, que aparece na maioria dos itens, parece ser mais uma tônica natural de discriminação do que uma característica estrutural da questão.

O próximo passo é a calibração final pela Teoria de Resposta ao Item, para o teste ser aplicado.

## **4.2 Aplicação da Teoria de Resposta ao Item - TRI**

A Teoria Clássica de Testes e a Teoria de Resposta ao Item não são, na verdade, dois modelos disjuntos, mas duas metodologias que se complementam, conforme mencionado anteriormente. Segundo Santos e Soares (2012), o início da Calibragem deve ser feito com as ferramentas da TCT; o que pode aumentar, em muito, a fluidez funcional da TRI.

Nos tópicos anteriores, os dados de uma amostra foram tratados com os métodos e ferramentas da TCT tendo como meta a melhoria dos resultados brutos advindos do pré-teste. Assim, nessa etapa, esses dados, agora mais filtrados, são submetidos aos passos da TRI, gerando um teste a ser aplicado a uma segunda amostra de respondentes (alunos) e esses, posteriormente, receberão uma classificação de desempenho e habilidade, ou seja, uma proficiência.

### **4.2.1 Calibragem e Reestruturação do Pré- teste**

A calibragem pelo BILOG-MG, segundo a Teoria de Resposta ao Item, excluiu os itens 008 e 014 e os parâmetros a (discriminação), b (dificuldade) e c (acaso) ficaram determinados, de acordo com a tabela 15. Nesta tabela, esses parâmetros, por item, estão nas 3ª, 4ª e 6ª colunas, respectivamente, *slope*, *threshold* e *asymptote*. Abaixo de cada parâmetro, está o erro-padrão decorrente do processo iterativo.

Tabela 15 - Parâmetros dos itens

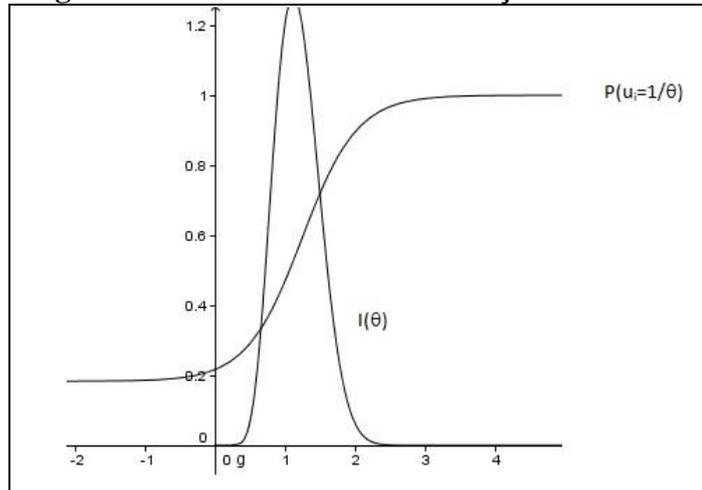
| ITEM     | INTERCEPT        | SLOPE           | THRESHOLD        | LOADING         | ASYMPTOTE       |
|----------|------------------|-----------------|------------------|-----------------|-----------------|
| ITEM0001 | 0.788<br>0.323*  | 0.569<br>0.213* | -1.386<br>0.706* | 0.494<br>0.185* | 0.222<br>0.096* |
| ITEM0002 | 1.058<br>0.360*  | 0.643<br>0.237* | -1.645<br>0.772* | 0.541<br>0.199* | 0.208<br>0.092* |
| ITEM0003 | -2.185<br>1.068* | 1.185<br>0.633* | 1.843<br>0.802*  | 0.764<br>0.408* | 0.100<br>0.047* |
| ITEM0004 | -0.442<br>0.405* | 0.817<br>0.357* | 0.541<br>0.488*  | 0.633<br>0.276* | 0.207<br>0.086* |
| ITEM0005 | 0.951<br>0.368*  | 0.742<br>0.283* | -1.282<br>0.631* | 0.596<br>0.227* | 0.204<br>0.090* |
| ITEM0006 | -0.535<br>0.444* | 0.858<br>0.387* | 0.624<br>0.505*  | 0.651<br>0.294* | 0.222<br>0.087* |
| ITEM0007 | -2.574<br>1.327* | 1.002<br>0.478* | 2.570<br>1.215*  | 0.708<br>0.338* | 0.104<br>0.045* |
| ITEM0009 | 0.698<br>0.365*  | 0.858<br>0.354* | -0.813<br>0.497* | 0.651<br>0.269* | 0.198<br>0.088* |
| ITEM0010 | -0.065<br>0.338* | 0.820<br>0.344* | 0.080<br>0.410*  | 0.634<br>0.266* | 0.184<br>0.082* |
| ITEM0011 | 1.817<br>1.267*  | 1.459<br>0.904* | -1.245<br>0.527* | 0.825<br>0.511* | 0.183<br>0.083* |
| ITEM0012 | -0.720<br>0.505* | 0.802<br>0.356* | 0.897<br>0.623*  | 0.626<br>0.278* | 0.255<br>0.091* |
| ITEM0013 | -2.182<br>1.085* | 0.668<br>0.279* | 3.265<br>1.874*  | 0.556<br>0.232* | 0.169<br>0.059* |
| ITEM0015 | -0.397<br>0.390* | 0.699<br>0.288* | 0.569<br>0.551*  | 0.573<br>0.236* | 0.230<br>0.091* |
| ITEM0016 | -0.541<br>0.427* | 1.045<br>0.521* | 0.518<br>0.393*  | 0.723<br>0.360* | 0.163<br>0.074* |
| ITEM0017 | -1.915<br>0.970* | 0.894<br>0.412* | 2.142<br>1.125*  | 0.666<br>0.307* | 0.165<br>0.064* |
| ITEM0018 | 1.817<br>1.267*  | 1.459<br>0.904* | -1.245<br>0.527* | 0.825<br>0.511* | 0.183<br>0.083* |
| ITEM0019 | -1.561<br>0.748* | 0.600<br>0.235* | 2.602<br>1.341*  | 0.515<br>0.201* | 0.203<br>0.074* |
| ITEM0020 | -2.034           | 0.947           | 2.148            | 0.688           | 0.168           |

Fonte: Dados da pesquisa.

De acordo com a tabela 15, os itens que mais discriminam são os de números 3, 7, 11, 16 e 18. Os itens identificados como os de maior dificuldade são os de números 13, 19 e 20 e os mais fáceis foram o 1 e o 2. Os mais prováveis de serem acertados com opção aleatória (por acaso) são os de números 12 e 15.

Além disto, o BILOG-MG apresenta a Curva Característica do Item - CCI e a curva de informação para todos os itens. As figuras 7, 8, 9 e 10 exemplificam essas curvas. A figura 7 apresenta a CCI do item 11, que representa um dos que mais discriminam<sup>2</sup>.

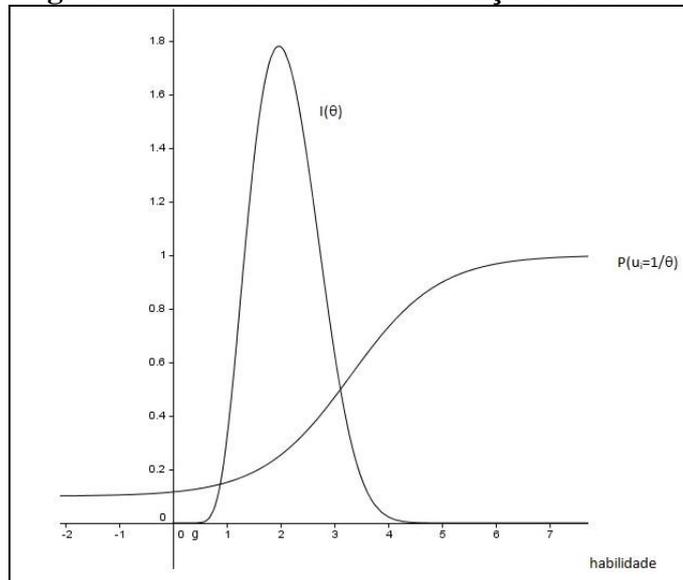
**Figura 7 - CCI e curva de informação do item 11**



Fonte: Dados da pesquisa.

A figura 8 apresenta a CCI do item 13, que representa um dos itens de maior dificuldade:

**Figura 8 - CCI e curva de informação do item 13**

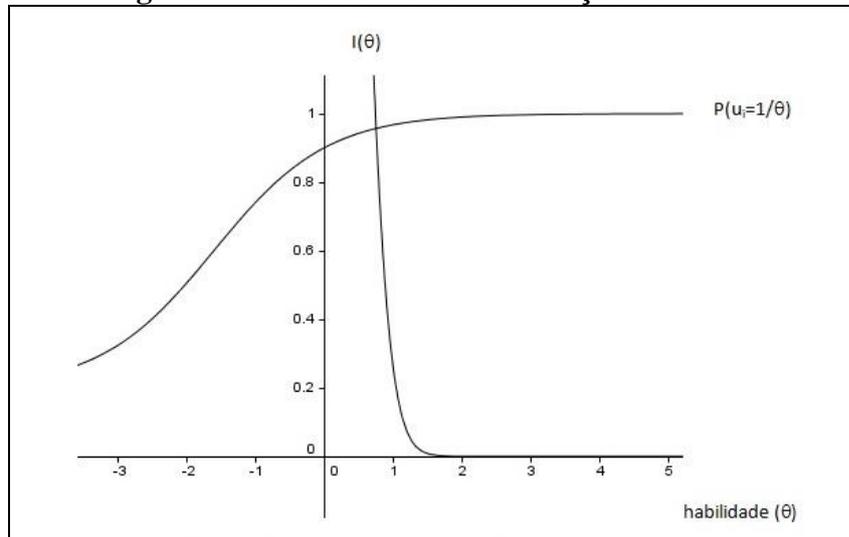


Fonte: Dados da pesquisa.

<sup>2</sup> As CCI de todos os itens encontra-se no apêndice 2 dessa dissertação.

A figura 9, por sua vez, apresenta a CCI do item 2, considerado como o de maior facilidade.

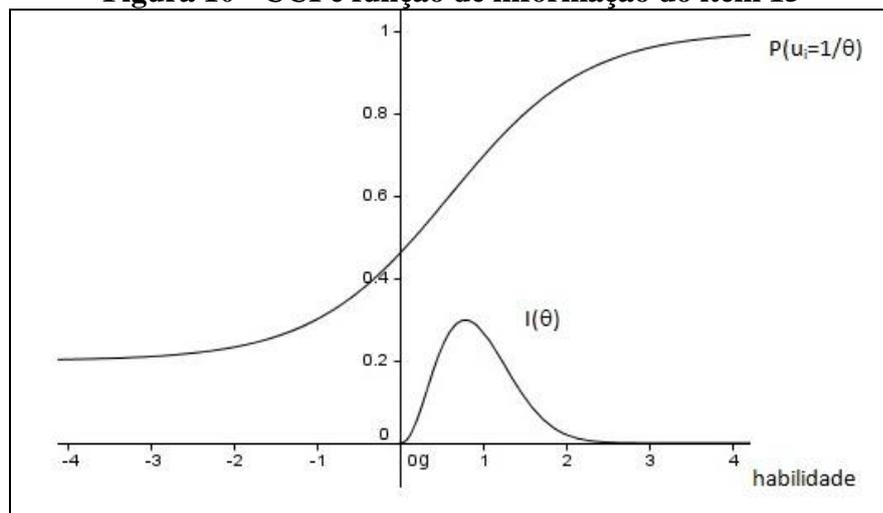
**Figura 9 - CCI e curva de informação do item 2**



Fonte: Dados da pesquisa.

Finalizando, tem-se a figura 10, que representa a curva CCI do item que possui a maior probabilidade de ser acertado por acaso, no caso, o 15.

**Figura 10 - CCI e função de informação do item 15**



Fonte: Dados da pesquisa.

### 4.3 Aplicação final do teste

A tabela 16, a seguir, mostra o desempenho absoluto do grupo (30 alunos) submetido ao teste calibrado.

**Tabela 16 - Escore bruto da amostra no teste**

| ALUNO Nº | ESCORE/20 |
|----------|-----------|
| 001      | 8         |
| 002      | 8         |
| 003      | 10        |
| 004      | 10        |
| 005      | 10        |
| 006      | 11        |
| 007      | 11        |
| 008      | 11        |
| 009      | 11        |
| 010      | 11        |
| 011      | 12        |
| 012      | 12        |
| 013      | 12        |
| 014      | 12        |
| 015      | 13        |
| 016      | 13        |
| 017      | 13        |
| 018      | 13        |
| 019      | 14        |
| 020      | 14        |
| 021      | 14        |
| 022      | 15        |
| 023      | 15        |
| 024      | 15        |
| 025      | 15        |
| 026      | 15        |
| 027      | 15        |
| 028      | 16        |
| 029      | 16        |
| 030      | 18        |

Fonte: Dados da pesquisa.

A distribuição dos escores é mostrada na tabela 17. Por esta distribuição, foram calculadas as medidas de posição e as medidas de dispersão. O coeficiente de dispersão, no caso, foi de 19% (coeficiente de variação =  $2,42/12,77 = 0,189...$  indicação de razoavelmente homogêneos segundo uma sugestão de classificação: entre 10 e 20%), conforme será visto mais adiante.

**Tabela 17 - Distribuição dos escores brutos**

| ACERTOS (X)     | F  | f.X | \Delta X | f. \Delta X | f. \Delta X ^2 |
|-----------------|----|-----|----------|-------------|----------------|
| 8               | 2  | 16  | 4,77     | 9,54        | 45,51          |
| 10              | 3  | 30  | 2,77     | 8,31        | 23,02          |
| 11              | 5  | 55  | 1,77     | 8,85        | 15,66          |
| 12              | 4  | 48  | 0,77     | 3,08        | 2,37           |
| 13              | 4  | 52  | 0,23     | 0,92        | 0,212          |
| 14              | 3  | 42  | 1,23     | 3,69        | 4,54           |
| 15              | 6  | 90  | 2,23     | 13,38       | 29,84          |
| 16              | 2  | 32  | 3,23     | 6,46        | 20,87          |
| 18              | 1  | 18  | 5,23     | 5,23        | 27,35          |
| <b>TOTAIS →</b> | 30 | 383 |          | 59,46       | 169,37         |

Fonte: Dados da pesquisa.

Tem-se, a partir das informações apresentadas na tabela, as seguintes medidas:

→ Média = 12,77

→ Mediana = 13

→ Moda = 15

→ Desvio médio = 1,98

→ Variância = 5,84

→ Desvio padrão = 2,42

→ Coeficiente de dispersão = 19%

A tabela seguinte mostra as respostas dos alunos e o gabarito. Essas respostas, como no pré-teste, foram codificadas, em 1, 2, 3, 4 e 5, respectivamente, para A, B, C, D e E. Os percentuais de incidência das alternativas no gabarito foram 22,2% para A, 33,3% para B, 5,6% para C, 27,8% para D e 11,1% para E.

Tabela 18 - Gabarito e respostas dos alunos

| GABARITO | D                    | B | D | D | A | C | B | E | A | D | B | A | B | B | D | A | E | B |
|----------|----------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ALUNOS   | RESPOSTAS DOS ALUNOS |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 001      | D                    | B | C | D | B | B | B | D | B | D | B | E | B | E | E | E | A | B |
| 002      | D                    | B | D | E | A | C | E | E | A | C | A | E | C | A | C | B | D | B |
| 003      | D                    | B | D | A | A | C | E | E | A | D | A | C | A | A | D | D | C | B |
| 004      | D                    | B | C | D | C | C | B | E | A | C | A | C | A | B | D | D | B | B |
| 005      | D                    | B | C | A | C | C | B | E | A | C | A | B | A | A | D | A | E | B |
| 006      | D                    | B | D | A | C | C | B | E | C | C | A | B | A | B | D | A | E | B |
| 007      | D                    | B | D | A | C | C | B | E | A | C | A | B | A | A | D | A | E | B |
| 008      | D                    | B | D | D | C | C | B | E | A | C | B | D | A | A | D | A | A | C |
| 009      | D                    | B | B | D | A | C | B | E | C | D | A | E | B | A | D | E | A | B |
| 010      | D                    | B | D | C | A | C | E | E | C | C | A | A | A | A | D | A | E | B |
| 011      | D                    | B | A | D | A | C | E | E | C | C | B | A | A | A | D | A | E | B |
| 012      | D                    | B | B | D | A | C | B | E | C | D | A | C | A | A | D | A | E | B |
| 013      | D                    | B | A | A | A | C | E | E | A | D | A | A | B | B | D | D | D | B |
| 014      | D                    | B | D | D | C | C | B | E | C | C | B | A | B | B | E | D | D | B |
| 015      | D                    | B | C | B | C | C | B | E | A | C | B | A | B | B | D | D | D | B |
| 016      | D                    | B | C | D | C | C | B | E | A | C | B | A | B | A | D | A | E | D |
| 017      | D                    | B | D | D | A | A | B | E | B | D | B | C | B | B | B | A | B | B |
| 018      | C                    | B | D | A | A | C | E | E | A | D | B | A | B | B | D | A | A | E |
| 019      | D                    | B | D | D | A | C | B | E | A | C | B | E | B | A | D | A | A | B |
| 020      | D                    | A | D | D | C | C | C | E | A | D | B | A | B | B | D | D | E | B |
| 021      | D                    | A | B | D | C | C | B | E | A | D | A | A | B | B | D | A | E | B |
| 022      | D                    | A | D | D | A | C | B | E | A | D | B | A | A | A | D | A | E | B |
| 023      | D                    | A | B | B | A | C | B | E | A | D | B | A | B | B | D | A | E | B |
| 024      | D                    | B | D | D | A | C | B | E | A | D | A | B | B | B | D | D | E | B |
| 025      | D                    | B | C | D | C | C | E | E | A | D | B | A | B | B | D | A | E | B |
| 026      | D                    | B | D | D | C | C | B | E | C | C | B | A | B | B | D | A | E | B |
| 027      | E                    | B | C | D | A | C | B | E | A | D | B | A | B | B | D | D | E | B |
| 028      | D                    | B | A | A | A | C | B | E | A | D | B | A | B | B | D | A | E | B |
| 029      | D                    | B | D | A | A | C | B | E | C | D | B | A | B | B | D | A | E | B |
| 030      | D                    | B | D | D | A | C | B | E | A | D | B | A | B | B | D | A | E | B |

Fonte: Dados da pesquisa.

Usando o BILOG-MG, com o script abaixo, classificou-se, então, o grupo amostral por habilidade:

```

>GLOBAL DFName = 'C:\Users\supergel\Desktop\BILOG\TFINAL.DAT',
    NPArm = 3,
    SAVe;
>SAVE SCOrE = 'TFINAL.SCO';
>LENGTH NITems = (18);
>INPUT NTOtal = 18,
    NIDchar = 4,
    KFName = 'TFINAL.KEY';
>ITEMS ;
>TEST1 TNAme = 'TEST0001',
    INUmber = (1(1)18);
(4A1, 18A1)
>CALIB ACCel = 1.0000,
    TPRior,
    GPRior;
>SCORE RSCtype = 1,
    LOCation = (500.0000);

```

Na tabela 19, encontram-se a habilidade e o erro padrão de cada um dos alunos que realizaram o teste:

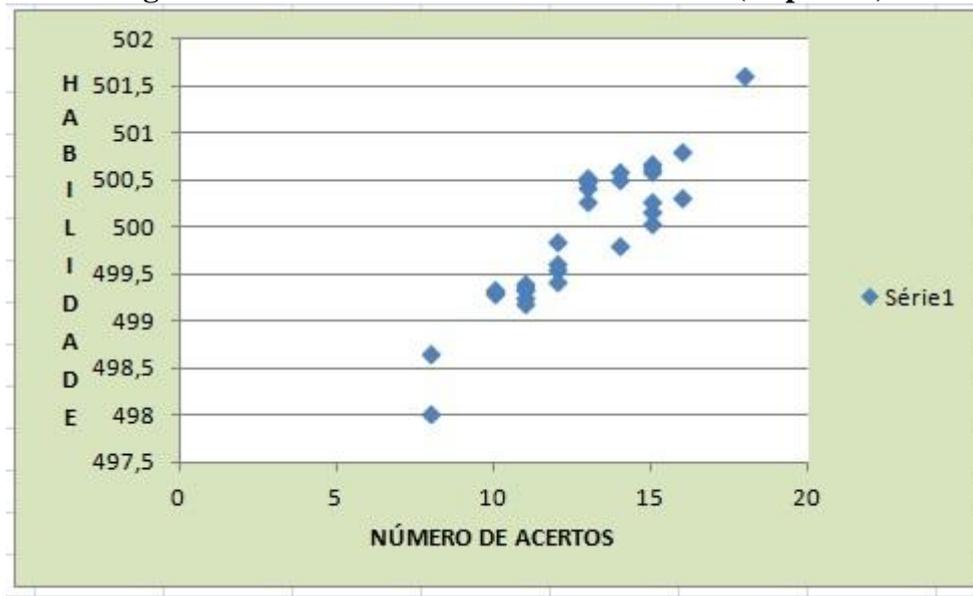
**Tabela19 - Classificação dos alunos por habilidade**

| ALUNO Nº          | HABILIDADE $\theta$    | ERRO PADRÃO   |
|-------------------|------------------------|---------------|
| 001               | 498,0253               | 0,6960        |
| 002               | 498,6536               | 0,5240        |
| 003               | 499,3141               | 0,4619        |
| 004               | 499,3007               | 0,4587        |
| 005               | 499,3454               | 0,4396.       |
| 006               | 499,2454               | 0,4798        |
| 007               | 499,3454               | 0,4396        |
| 008               | 499,4086               | 0,4138        |
| 009               | 499,3541               | 0,4757        |
| 010               | 499,1921               | 0,4982        |
| 011               | 499,5536               | 0,4780        |
| 012               | 499,4146               | 0,4366        |
| 013               | 499,6114               | 0,4989        |
| 014               | 499,8473               | 0,5777        |
| 015               | 500,4192               | 0,3715        |
| 016               | 500,2702               | 0,4118        |
| 017               | 500,4881               | 0,4380        |
| 018               | 500,5248               | 0,4044        |
| 019               | 499,8025               | 0,4830        |
| 020               | 500,5010               | 0,4140        |
| 021               | 500,5851               | 0,4325        |
| 022               | 500,0299               | 0,4842        |
| 023               | 500,6038               | 0,4842        |
| 024               | 500,1601               | 0,4625        |
| 025               | 500,6778               | 0,4622        |
| 026               | 500,6289               | 0,4412        |
| 027               | 500,2718               | 0,5463        |
| 028               | 500,3101               | 0,5477        |
| 029               | 500,8077               | 0,4929        |
| 030               | 501,6038               | 0,6049        |
| <b>MÉDIA REAL</b> | <b>DESVIO PADRÃO →</b> | <b>0,7625</b> |

Fonte: Dados da pesquisa.

Nesta tabela, considera-se como valor de referência 500 pontos (média) e cada unidade de desvio padrão, 100 pontos. Observe-se que o número de itens certos (número de acertos) não determina funcionalmente a habilidade; por isso, alunos com mesma quantidade de acertos podem ter diferentes notas (habilidade). Entretanto, há uma correlação entre o número de acertos e a nota da ordem de 91,33%, como se vê no gráfico de dispersão da figura 11.

**Figura 11 - Número de acertos x habilidade (dispersão)**



Fonte: Dados da pesquisa

Em razão desta correlação, o ajuste pela técnica de regressão linear identifica a tendência assinalada no gráfico da figura 12. A expressão dessa correlação é  $\theta(X) = (0,2809)X + 496,3232$ . Essa reta representa uma valorização equitativa dos itens, partindo-se de um valor de referência igual a 496,3232, cada item valendo 0,2809.

**Figura 12 - Número de acertos x habilidade (tendência)**



Fonte: Dados da pesquisa.

A partir dos resultados obtidos, elaborou-se outro produto dessa pesquisa, além das questões formuladas para o pré-teste aplicado aos alunos e que será descrito a seguir:

#### 4.4 O produto: um banco de itens com planilhas

Toda a metodologia que avalia itens na Teoria Clássica de Testes - TCT é otimizada, de modo eficaz, com a utilização de planilhas EXCEL. Essas planilhas permitem associar matrizes com termos de expressões matemáticas, célula a célula, simplificando as totalizações e automatizando as conclusões. Como essa teoria constitui-se de fórmulas, tanto de Estatística Descritiva quanto dos processos correlativos, as planilhas são os instrumentos fundamentais de execução desses processos de avaliação.

Por outro lado, a relevância primária do presente trabalho dissertativo é a utilização da Teoria de Resposta ao Item - TRI, complementando a Teoria Clássica dos Testes - TCT no efetivo exercício de avaliar, por professores e escolas, garantindo-se, assim, a qualidade no acompanhamento de seus aprendizes e implementando medidas eficazes de construção do conhecimento. Portanto, integrando-se um instrumento como a planilha aos modelos de avaliação e classificação citados, tem-se uma das inúmeras possibilidades para um novo aplicativo, o **Planum** é uma delas.

O Planum é um banco de itens submetidos à Teoria de Resposta ao Item calibrados. Cada bateria de até 20 itens pode ser armazenada no Planum diretamente, se forem calibrados ou depois de passarem pelo Filtro e/ou por qualquer *software* que execute a calibragem pela TRI. O Filtro é um programa estruturado com planilhas, que submete o grupo de itens a uma triagem, facilitando a calibragem pela TRI.

A análise executada pelo filtro conta com dois tipos de instrumentos, que são os instrumentos de atributos estruturais do grupo como um teste (tais como o equilíbrio e a consistência) e os instrumentos que avaliam cada item, considerando os parâmetros expressivos de dificuldade, discriminação e de acerto aleatório. O quadro 3, a seguir, mostra os quesitos avaliados e a metodologia correspondente de avaliação. Os seis primeiros quesitos referem-se à consistência do teste, ao seu equilíbrio e ao seu grau de normalidade. Os quatro últimos quesitos são relativos aos itens individualmente.

**Quadro 3 – Quesitos avaliados pelo filtro**

| QUESITO AVALIADO                              | METODOLOGIA DE AVALIAÇÃO  |
|---|---|
| Erro, Resultado Global e Discriminação        | Levantamento de dados relativos (ERROS E MÉDIAS)                |
| Equilíbrio dos níveis de dificuldade no teste | Levantamento percentual (20% fáceis, 60% médias e 20% difíceis) |
| Grau de Normalidade                           | Fórmula de Correlação de Filliben (1975)                        |
| Potencial de discriminação do item            | Coefficiente de Correlação de Pearson                           |
| Homogeneidade e Consistência do teste         | Fórmula 20 de Curt-Richardson                                   |
| Confiabilidade do Teste                       | Teste de SPEARMAN BROWN   |
| Potencial de discriminação do item            | Processo de Correlação D  |
| Potencial de discriminação do item            | Coefficiente de Correlação Bisserial                            |
| Variância individual dos itens                | Constatação por cálculo   |
| Atratividade nula de Alternativas             | Constatação por cálculo   |

Fonte: Elaborado pelo pesquisador.

As conclusões do diagnóstico e as recomendações de procedimento são dadas em células adjacentes da tabela anterior, sendo que o procedimento recomendado para cada item (MANTER, MUDAR OU EXCLUIR) está associado ao seu percentual de correção em todo o diagnóstico: MANTER (75% ou mais), MUDAR (maior de 50% e menor do que 75%) e EXCLUIR (abaixo de 50%).

No Planum, os itens são armazenados e identificados pelo Código de Referência do Item - CRI, que possui 5 caracteres; sendo os três primeiros algarismos compondo o número do item e os últimos são letras abreviando os nomes dos ramos do conhecimento matemático (Aritmética - AT, Álgebra - AL, Geometria Plana - GP, Geometria Espacial - GE, Geometria Analítica - GA e Estatística - ES). Assim, por exemplo, 012ES é a questão de número 12 de Estatística. Entretanto, na pesquisa ao banco, pode-se entrar pelo conteúdo específico (Funções afim, por exemplo), com a habilidade (H1 a H30), que o item procura avaliar ou pelo ramo do conhecimento matemático (Aritmética, Álgebra, Geometria Plana, Geometria Espacial, Geometria Analítica e Estatística).

Todos os comandos do Planum são executados com funções elementares do EXCEL associadas em sintaxes convenientes. Além de Planilhas, há arquivos de dados do WORD, PDF e de extensão JPEG. O Planum ainda traz todos os itens resolvidos e apresenta cada item pesquisado com todos os seus atributos (resposta, nível de dificuldade, procedência, ano, percentual de marcação de cada alternativa por item no pré-teste, parâmetro de dificuldade, parâmetro de discriminação e probabilidade de acerto por acaso).

## 5 CONSIDERAÇÕES FINAIS

O conhecimento matemático é um patrimônio que cresce dia a dia com a pesquisa científica; patrimônio esse construído pela mente coletiva do gênio humano, que se mostra em regime de complementaridade a cada novidade descoberta. Exemplo disto: a Teoria Clássica dos Testes - TCT sendo complementada, enquanto instrumento de medição psicométrica, pela Teoria de Resposta ao Item - TRI. Dessa maneira, a TCT não perdeu sua importância com o desenvolvimento dos modelos logísticos que constituem a TRI, mas foi a raiz de todo o processo de pesquisa que gerou a TRI, ao antecipar um critério de triagem de itens, uma pré-calibragem, antes de serem submetidos ao ajuste minucioso da TRI.

As duas teorias, como processos isolados, apresentam uma diferença fundamental de base: enquanto a TCT se baseia no teste e seus escores brutos, a TRI se baseia nos itens que compõem o teste. Portanto, a TCT como metodologia exclusiva de avaliação tem um caráter sumário, ou seja, o avaliado sabe ou não sabe um conceito, domina ou não uma habilidade. Por seu lado, a TRI considera o aprendizado ou domínio parcial e, além disto, legitima essa ideia verificando a coerência das respostas através dos parâmetros discriminação, dificuldade e incidência do acaso.

Para se garantir como processo fidedigno de avaliação, a TRI deve contar com a unidimensionalidade do objeto de avaliação, ou seja, avalia-se uma única competência por item. Isso se deve ao postulado de que existe uma habilidade, um traço latente, imprescindível à compreensão de um conceito, ao domínio de uma competência. Esse traço não é aparente e, por isso, não pode ser objetivamente avaliado. Daí a importância fundamental do item, a qual deve ser elaborado com essência, ou seja, com capacidade real de discriminação, pois é a referência de todo o processo. Assim, é preciso que suas alternativas de respostas não sejam apenas pretextos, mas, no mínimo, meias verdades que contemplem o conhecimento parcial e, por isso, exerçam atração no avaliado, sejam distratores ou descritores.

Portanto, a questão inicialmente levantada sobre a fidedignidade da TRI na sua atribuição de medir habilidade, uma proficiência, está onde sempre esteve: no campo pedagógico e não em uma fórmula mágica desenvolvida e executada no computador, por um *software*. A mente que elabora um item é a de um professor e não de uma máquina, pois os aspectos que fundamentam uma avaliação passam necessariamente pelo fator humano.

Nas avaliações aqui apresentadas, a amostra de 30 alunos que qualificaram o pré-teste teve mais qualidade do que os outros 30 alunos submetidos ao teste final, pois na mesma escala de 100 com média 500, o resultado global do pré-teste foi melhor. As notas extremas

do pré-teste foram 346,7746 e 640,5448, enquanto no teste os extremos foram 498,0253 e 501,6038; em ambos os contextos, média em torno de 500 confirmada. Observe que a amplitude amostral de proficiência foi de 293,77 no pré-teste e de 3,5785 no teste, mostrando o pré-teste uma diversidade maior de proficiências. Isso é uma indicação de que a calibragem teve como referência uma amostra de alunos com alto domínio das habilidades envolvidas nos itens.

O produto dessa dissertação consiste em um aplicativo denominado Planum. A ideia desse aplicativo é suprir uma carência que, na maioria dos casos, nem é percebida, dada a prática comum nos procedimentos pedagógicos ligados ao estabelecimento de parâmetros avaliativos de aquisição no campo da aprendizagem. Construir um instrumento de avaliação que, de fato, avalie com presteza e coerência a eficiência de um trabalho de formação em suas duas interfaces - a escola e o aprendiz - é uma missão que quase sempre foi subestimada em sua importância. Com o desenvolvimento da informática, a Teoria de Resposta ao Item - TRI, pôde se tornar o modelo apropriado de avaliação quantitativa, posto que contempla as variáveis qualitativas, graças à natureza dos métodos estatísticos. Além disso, valoriza-se a Teoria Clássica dos Testes - TCT, estabelecendo-se uma relação de complementaridade entre as duas teorias.

## REFERÊNCIAS

- ANDRADE, Dalton Francisco de; TAVARES, Heliton Ribeiro; VALLE, Raquel da Cunha. **Teoria da Resposta ao Item: Conceitos e Aplicações**. São Paulo: SINAPE, 2000.
- ANGOFF, W.H. Scales, norms and equivalent scores. In: R.L. Thorndike (Ed.). **Educational measurement**. 2.ed. Washington, DC, 1971. p.508-600.
- BRASIL. INEP. **Prova ENEM 2012: domingo**. 2012. Disponível em: [http://download.inep.gov.br/educacao\\_basica/enem/provas/2012/caderno\\_enem2012\\_dom\\_ci\\_nza.pdf](http://download.inep.gov.br/educacao_basica/enem/provas/2012/caderno_enem2012_dom_ci_nza.pdf). Acesso em: 10 dez. 2015.
- BRASIL. Ministério da Educação e Cultura. **ENEM: a segunda maior prova de acesso ao Ensino Superior do mundo**. 2015. Disponível em: <http://www.brasil.gov.br/educacao/2015/10/enem-a-segunda-maior-prova-de-acesso-ao-ensino-superior-do-mundo>. Acesso em: 13 jan. 2016.
- COIMBRA, Carlos A. Q. **Teoria da Medição de Variáveis Latentes: História e Utilização Atual**. Brasília: MAST/MCT, 2006.
- CURRAN, P. J, BOLLEN, K. A. The best of both worlds: Combining autoregressive and latent curve models. In: COLLINS, L. M; SAYER, A. G, (Ed.) **New methods for the analysis of change**. Washington, DC: American Psychological Association Press; 2001. p.105-136.
- ESPÍRITO SANTO, Alexandre do. **Correlação Bisserial**. 1978. Disponível em: [www.uel.br/revistas/uel/index.php/seminario/article/viewFile/.../5627](http://www.uel.br/revistas/uel/index.php/seminario/article/viewFile/.../5627). Acesso em: 9 set. 2015.
- GLOBO. G1. **Matriz de referência do ENEM**. 2014. Disponível em: [g1.globo.com/jornalhoje/download/0,,5840-1,00.doc](http://g1.globo.com/jornalhoje/download/0,,5840-1,00.doc). Acesso em: 10 nov. 2015.
- HAMBLETON, R.K., JONES, R.W. & ROGERS, H.J. Influence of item parameter estimation errors in test development. **Journal of Educational Measurement**, v.30, n.2, p. 143-155. 1993.
- INEP. **Matriz competência: Matemática Ensino Médio**. 2013. Disponível em: [http://download.inep.gov.br/educacao\\_basica/enceja/matriz\\_competencia/Mat\\_Mat\\_Tec\\_EM.pdf](http://download.inep.gov.br/educacao_basica/enceja/matriz_competencia/Mat_Mat_Tec_EM.pdf). Acesso em: 20 dez. 2015.
- KLEIN, Ruben. **Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB)**. 2009. Disponível em: <http://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/38>. Acesso em: 13 dez. 2015.
- MARTINS, Gilberto de Andrade. **Estatística Geral e Aplicada**. 3. ed. São Paulo: Ed. Atlas, 2006.
- PASQUALI, Luiz. (Org.) **Teoria e métodos de medida em ciências do comportamento**. Brasília: INEP, 1996. 432p.

PASQUALI, Luiz. **Psicometria: teoria dos testes na psicologia e na educação**. 5.ed. Rio de Janeiro: Vozes, 2013.

PASQUALI, Luiz; PRIMI, Ricardo. Fundamentos da Teoria de Resposta ao Item – TRI. In: **Avaliação Psicológica**. v.2, n.2, p. 99-110. 2003.

RENZO, Oswald. **Algumas Considerações Sobre a Existência do Traço Latente na Construção de Provas Psicométricas**. São Paulo: Moytica, Eficiência Empresarial, 2008.

SANTOS, Marcos Antônio da Cunha; SOARES, José Francisco. Estimativas Não Paramétricas das Curvas Características dos Itens em Testes Educacionais. **Est. Aval. Educ.**, São Paulo, v. 23, n. 52, p. 130-144, maio/ago. 2012.

VIANNA, Heraldo Marelim. **Avaliações nacionais em larga escala**. São Paulo: DPE, 2003.

VIANNA, Heraldo Marelim. **Provas e testes no concurso vestibular**. Disponível em: <http://www.fcc.org.br/pesquisa/publicacoes/es/artigos/110.pdf>. Acesso em: 09 ago. 2014.

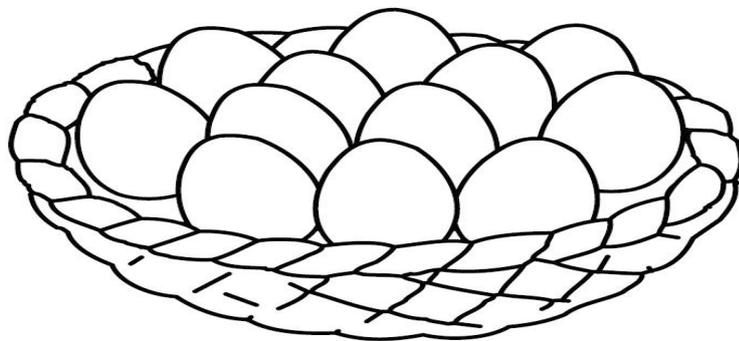
## APÊNDICES

### Apêndice 1 - Conjunto de itens utilizado no pré-teste

#### QUESTÃO 1:

Um procedimento rural, ainda adotado em parte, é a contagem de ovos de 12 em 12 (dúzias). Então, 1 dúzia equivale a 12 ovos e 12 dúzias recebe o nome de grossa, portanto 1 grossa são 12 dúzias, ou seja, 144 ovos. O motivo dessa adoção é a facilidade com a qual se pode dividir por 2, 3, 4, 6 e 12. Por outro lado, se um comerciante vende ovos por unidade, ele pode achar mais fácil contar em dezenas e aí teríamos 1 dezena equivale a 10 ovos e 10 dezenas equivale a 1 centena. Então, se um fazendeiro conta os ovos de uma cesta, de 12 em 12, e registra 154, ele mesmo entenderá que há na cesta 1 grossa, 5 dúzias e 4 unidades, ou seja, 208 ovos. A ideia seria análoga, se o fazendeiro contasse de 10 em 10. Portanto, se, na contagem de 10 em 10, o fazendeiro registrasse 171 ovos, certamente, contando em dúzias, teria o registro

- (A) 117.
- (B) 711.
- (C) 145.
- (D) 123.
- (E) 132.

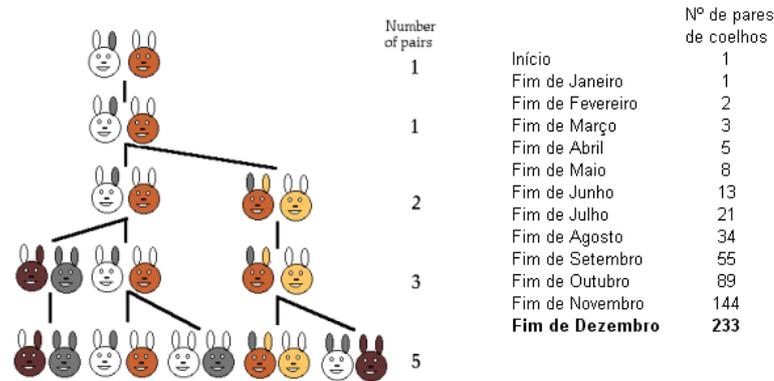


(guiadomarketing.com – disponível em 29/03/2013)

#### QUESTÃO 2:

Em 1202, o matemático Leonardo de Pisa, também conhecido como Fibonacci, publicou seu livro de Aritmética e Álgebra de nome *Liber Abaci*, ou *livro do ábaco*. Nesse livro, Fibonacci propôs o seguinte problema: *Suponha que um casal de coelhos só se reproduz a partir de dois meses de idade e que, em cada vez, gera apenas um casal de coelhos. Começando-se uma criação com um casal recém nascido de coelhos, qual será o número de casais de coelhos mês a mês?* Então surgiu a sequência:

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ...



(Disponível em: [web.wducom.pt](http://web.wducom.pt). Acesso em: 29/03/2013)

Conhecida como *Sequência de Fibonacci*. Se ao problema de Fibonacci acrescentarmos a condição “as coelhas dessa espécie morrem fatalmente com o nascimento das crias e os coelhos viúvos não mais se reproduzem”, teríamos modificada a sequência de Fibonacci. Nesse caso o quinto termo da nova sequência representaria

- (A) 3 coelhos 2 coelhas.
- (B) 5 coelhos e 2 coelhas.
- (C) 5 coelhos e 3 coelhas.
- (D) 2 casais de coelhos.
- (E) 5 coelhos e 1 casal.

### QUESTÃO 3:

Sabemos que a razão entre o comprimento de uma circunferência e o seu diâmetro é o número  $\pi$ , aproximadamente igual a 3,14. Fazendo uma pesquisa sobre esse número, um cientista incumbiu o seu assessor de medir várias rodas e seus diâmetros e, com uma calculadora, dividir as medidas citadas, atribuindo o valor encontrado a  $\pi$ . Então, cada uma das centenas de roda seria marcada com um adesivo contendo o respectivo valor da razão encontrada. O assessor foi instruído a aproximar todas as medidas para o inteiro mais próximo, antes de digitá-las na calculadora. Dentre todas as rodas medidas, a menor apresentou, na calculadora,



**QUESTÃO 5:**

Numa pista retilínea de prova automobilística, um cronômetro marca o tempo, em horas, dos carros e com o comprimento da pista, determinam a velocidade média desses veículos. O comprimento da pista é de 132.000 m. Se o tempo de um carro é de 1 hora e 39 minutos, então sua velocidade média, em km/h, é

(Disponível em: [esporte.hsw.uol.com.br](http://esporte.hsw.uol.com.br). Acesso em: 29/03/2013)

- (A) 80.
- (B) 95.
- (C) 97.
- (D) 100.
- (E) 102.

**QUESTÃO 6 :**

Seu Ataliba, quando mais jovem, entrou para um programa de aposentadoria que funcionava assim: O associado entra com um capital que seja proporcional ao tempo em que deseja se aposentar e o banco se encarrega de aplicar esse dinheiro durante esse tempo, da maneira mais rentável possível. A tabela abaixo dá alguns parâmetros desse programa.

| FAIXA DE APLICAÇÃO EM REAIS | TAXA DE JURO POR PERÍODO DE TEMPO | TEMPO ATÉ SE APOSENTAR EM ANOS | VALOR MENSAL A RECEBER   | TEMPO DE RECEBIMENTO EM ANOS |
|-----------------------------|-----------------------------------|--------------------------------|--------------------------|------------------------------|
| 10.000 A<br>15.000          | 100% AO TRIÊNIO<br>(3 ANOS)       | 12                             | 2.000,00 A<br>3.000,00   | 10                           |
| 15.001 A<br>20.000          | 100% AO BIÊNIO<br>(2 ANOS)        | 12                             | 4.000,00 A<br>5.330,00   | 20                           |
| 20.001 A<br>25.000          | 200% AO TRIÊNIO<br>(3 ANOS)       | 12                             | 6.750,00 A<br>8.440,00   | 20                           |
| 25.001 A<br>30.000          | 200% AO BIÊNIO<br>(2 ANOS)        | 12                             | 12.000,00 A<br>16.000,00 | VITALÍCIO                    |

Agora, 12 anos depois de se associar ao plano de aposentadoria, Seu Ataliba percebeu que o capital que havia aplicado nesse tempo rendeu 80 vezes o seu valor. Então, É CORRETO afirmar que Seu Ataliba

- (A) aplicou R\$ 12.000,00 e vai receber durante 10 anos.
- (B) aplicou R\$ 18.000,00 e vai receber durante 20 anos.
- (C) aplicou R\$ 22.000,00 e vai receber durante 20 anos.
- (D) aplicou R\$ 25.000,00 e vai receber durante 10 anos.
- (E) aplicou R\$ 29.000,00 e vai receber para o resto de sua vida.

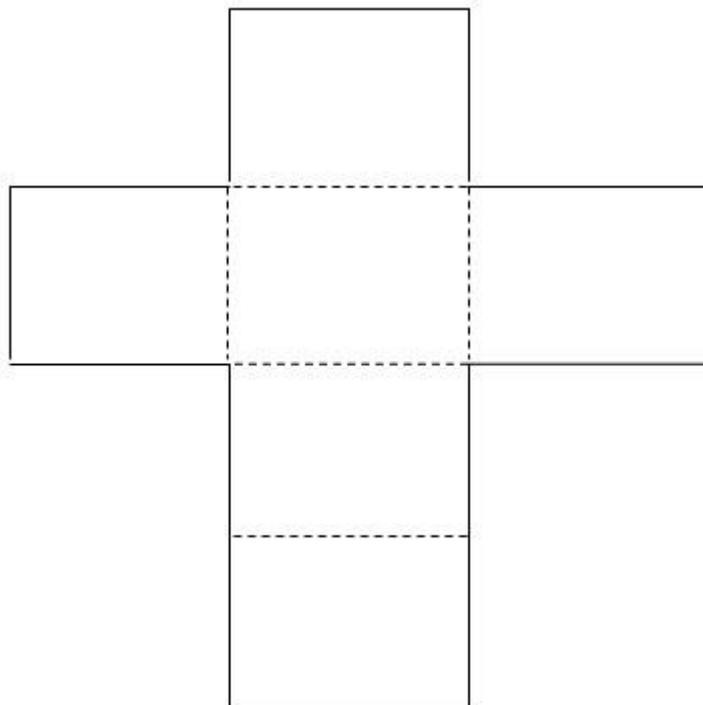
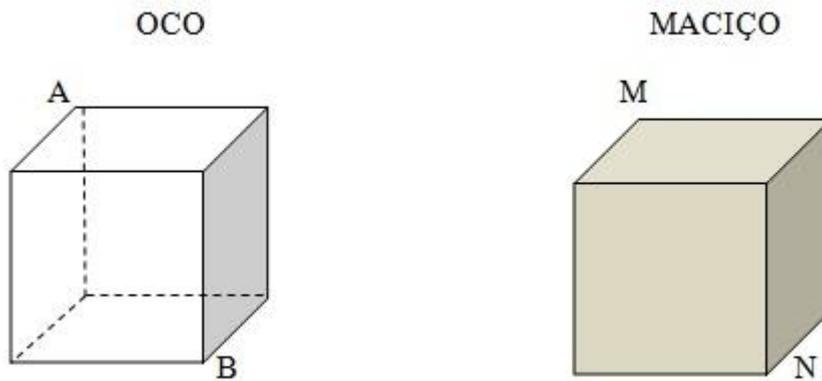
### QUESTÃO 7:

Considere os seguintes POSTULADOS:

*I – A eletricidade só circula num meio condutor diferente do ar.*

*II – A eletricidade desloca-se de um ponto a outro pela trajetória mais curta.*

Dois cubos de metal condutor são tais que um é maciço e o outro é oco, ambos com 3 metros de aresta, como nas figuras abaixo:



### PLANIFICAÇÃO DOS CUBOS

Nos vértices A e B do cubo OCO há, respectivamente, um interruptor e uma lâmpada ligada a ele. Do mesmo modo, nos vértices M e N do cubo MACIÇO, há, respectivamente, um interruptor e uma lâmpada ligada a ele. Em cada um dos casos, quando se liga o interruptor, uma corrente elétrica supostamente se desloca do interruptor para a lâmpada.

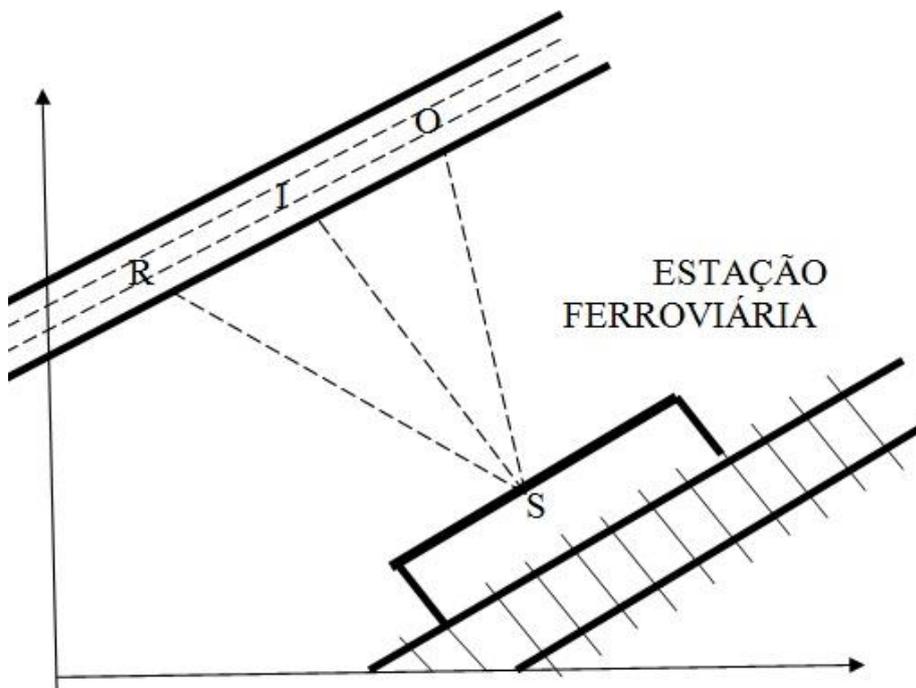
Sobre as trajetórias AB e MN, PODE-SE AFIRMAR que

- (A) AB tem o mesmo comprimento de MN.
- (B) AB tem em torno de 1,5 m a mais que MN.

- (C) AB tem o dobro do comprimento de MN.
- (D) AB tem 1 aresta a mais que MN.
- (E) AB tem 1 diagonal a mais que MN.

### QUESTÃO 8:

Observe a figura:



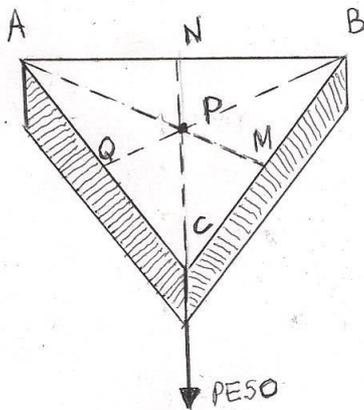
A figura representa um trecho de uma cidade com rio navegável, por onde sai toda a produção de alimentos que chega pela estação ferroviária. Uma das três linhas tracejadas é o trajeto da produção, via transporte rodoviário, da estação até o porto na margem do rio. Em relação a um sistema ortogonal de eixos, mostrado na figura, sabe-se que as distâncias dos pontos **R** e **O** até o eixo horizontal são, respectivamente, 4 km e 7 km; as distâncias desses pontos até o eixo vertical são, respectivamente, 1 km e 5 km. Sabe-se ainda que a saída **S** da produção na estação ferroviária está a 3 km do eixo horizontal e a 8 km do eixo vertical. Se cada

quilômetro de transporte rodoviário fica por R\$ 450,00, o trajeto MAIS ECONÔMICO fica por

- (A) R\$ 3.150,00.
- (B) R\$ 2.250,00.
- (C) R\$ 1.750,00.
- (D) R\$ 1.150,00.
- (E) R\$ 900,00.

### QUESTÃO 9:

Uma plataforma de concreto tem o formato de um triângulo como o da figura abaixo.

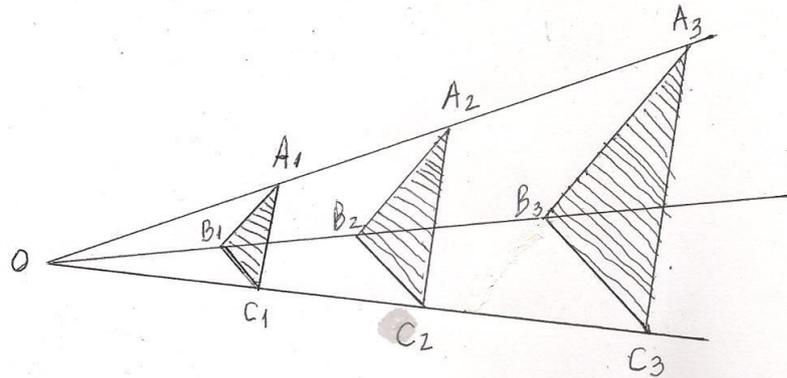


O triângulo da figura tem, além do lado AB de 50 cm, dois lados de 40 cm cada um. Essa plataforma será sustentada no ar por uma coluna fina, perpendicular a ela, na direção do ponto P, que é o centro de gravidade da plataforma. Por outro lado, o ponto P é o ponto de interseção das três linhas tracejadas. Sobre essas linhas tracejadas, **É CORRETO** afirmar-se que

- (A) elas formam ângulos retos com os lados do triângulo.
- (B) elas formam entre si apenas ângulos menores do que  $90^\circ$ .
- (C) cada uma delas divide ao meio um ângulo do triângulo.
- (D) cada uma divide a outra na razão  $\frac{1}{4}$ .
- (E) cada uma divide um lado do triângulo ao meio.

**QUESTÃO 10:**

Há um processo geométrico de ampliação ou redução de figuras chamado de **HOMOTETIA**. Segundo esse processo, a partir de um ponto de fuga, traçam-se retas pelos pontos da figura e sobre essas retas, marcam-se os novos pontos à direita (ampliando) ou à esquerda (reduzindo) dos pontos originais, todos a uma mesma distância desses originais.



Na figura acima, as distâncias  $OA_1$ ,  $A_1A_2$  e  $A_2A_3$  são, respectivamente iguais a 10 cm, 5 cm e 12 cm. Se a área do triângulo de vértice  $A_2$  é  $225 \text{ cm}^2$ , então, as áreas dos triângulos de vértices  $A_1$  e  $A_3$ , em centímetros quadrados, são, respectivamente,

- (A) 100 e 729.
- (B) 50 e 360.
- (C) 30 e 320.
- (D) 30 e 325.
- (E) 27 e 249.

**QUESTÃO 11:**

A Fundação Seade e o Dieese divulgam sistematicamente os resultados da Pesquisa de Emprego e Desemprego. O gráfico a seguir mostra as taxas de desemprego relativas a janeiro, fevereiro e março de 2013 (barras nesta ordem) em algumas regiões metropolitanas do Brasil:



O maior coeficiente de empregabilidade da Região sudeste foi de

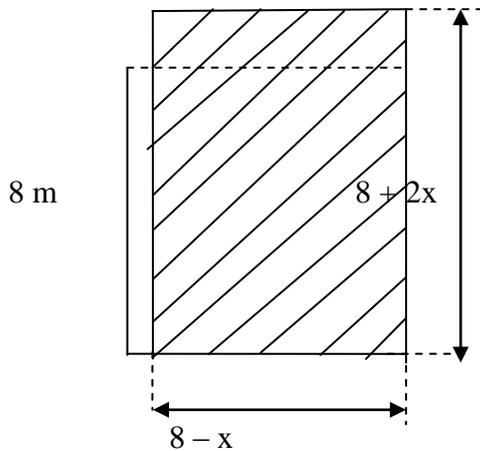
- (A) São Paulo, janeiro de 2013.
- (B) São Paulo, fevereiro de 2013.
- (C) Belo Horizonte, março de 2013.
- (D) Belo Horizonte, janeiro de 2013.
- (E) Salvador, março de 2013.

**QUESTÃO 12:**

Um jogo de festa junina é proposto nos seguintes moldes:

Num campo quadrado de 8 m de lado, um coelho é solto e o jogador deve pegá-lo o mais rápido possível. O número de pontos obtidos é a área do terreno dividida pelo tempo em minutos, uma vez que cada jogador pode alterar os lados do quadrado, tirando um pedaço na base e somando o dobro desse pedaço na altura. Por exemplo, um jogador pode tirar 1 m na base e somar 2 m na altura; sua área será de  $7 \times 10 = 70 \text{ m}^2$ .

Pode-se concluir que se dois jogadores levaram o mesmo tempo para pegar o coelho, o jogador que venceu, com certeza, deslocou-se numa área



- A)  $1 \text{ m}^2$  menor do que a do quadrado inicial.
- B)  $1,5 \text{ m}^2$  menor do que a do quadrado inicial.
- C)  $2,0 \text{ m}^2$  menor do que a do quadrado inicial.
- D)  $6,0 \text{ m}^2$  menor do que a do quadrado inicial.
- E)  $8,0 \text{ m}^2$  maior do que a do quadrado inicial.

### QUESTÃO 13:

A TRI – Teoria de Resposta ao Item- critério de quantificação e atribuição de proficiência, usada pelo ENEM, provoca um efeito de proporcionalidade inversa do valor da questão em relação ao número de candidatos que a acertaram. Por isso se diz que mais vale a questão menos acertada pelo universo de candidatos. Suponha, em outro contexto, que uma prova tenha valor total de 73 pontos e seja constituída de 5 questões, sendo o valor de cada questão inversamente proporcional ao percentual de acertos desta questão pelo grupo de candidatos. Veja a tabela de acertos:

| QUESTÃO | 1   | 2   | 3   | 4   | 5   |
|---------|-----|-----|-----|-----|-----|
| ACERTOS | 20% | 35% | 70% | 80% | 40% |

Dentre os candidatos que fizeram a prova, destaquemos os quatro do quadro abaixo:

| NOME DO CANDIDATO | QUESTÕES CERTAS POR CANDIDATO |   |   |   |   |
|-------------------|-------------------------------|---|---|---|---|
|                   | 1                             | 2 | 3 | 4 | 5 |
| Mauro             |                               |   |   |   |   |
| Adriana           |                               |   |   |   |   |
| Inez              |                               |   |   |   |   |
| Paulo             |                               |   |   |   |   |

MARQUE a alternativa que classifica corretamente os quatro candidatos, de acordo com o critério exposto.

- A) Adriana –Paulo – Mauro – Inez.
- B) Adriana –Paulo – Inez – Mauro.
- C) Paulo – Mauro – Inez – Adriana.
- D) Mauro – Adriana – Inez – Paulo.
- E) Mauro – Adriana – Paulo \_ Inez.

**QUESTÃO 14:**

Um pequeno empreendedor tem um negócio onde emprega 5 funcionários, cujos salários, em reais, são 700, 710, 820, 1400 e 1600. Na época do reajuste salarial, o empresário achou justo conceder como reajuste a diferença percentual entre a mediana e a média dos salários de até então. MARQUE a alternativa que mostra o novo custo, em reais, da folha de pagamento do negócio.

- A) 4.100,00.
- B) 5.230,00.
- C) 6.228,00.
- D) 6.668,00.
- E) 7.228,00.

**QUESTÃO 15:**

Uma companhia de energia exclusivamente rural fornece energia ao preço fixo por kWh consumido, mas cobra uma taxa fixa de manutenção de linhas no valor de R\$ 25,00. A autenticação bancária de uma das contas dessa companhia registra um preço pago de R\$ 247,50, sendo que o consumo registrado na conta foi de 500 kWh. Como a companhia cobrou 10% de multa por atraso, Pode-se garantir que o preço do kWh consumido foi

- A) R\$ 0,35.
- B) R\$ 0,40.
- C) R\$ 0,45.
- D) R\$ 0,50.
- E) R\$ 0,55.

**QUESTÃO 16:**

Um vendedor de praia embala fatias de melão, som casca e sem sementes e inclui no preço de venda o papel transparente com o qual embala cada fatia. Veja o quadro abaixo.



|   |                    |
|---|--------------------|
| Raio de cada melão .....                                      | 9 cm               |
| Medida angular de cada fatia .....                            | $72^\circ$         |
| Preço de custo de cada melão ...                              | R\$ 3,80           |
| Preço de cada centímetro quadrado de papel transparente ..... | R\$ 0,005          |
| Lucro do vendedor .....                                       | R\$ 3,60 por melão |

Se o raio do miolo de sementes é 4 cm e

- A) 2,60.
- B) 3,50.
- C) 3,90.
- D) 4,20.
- E) 4,50.

**QUESTÃO 17:**

A tabela abaixo mostra as faixas salariais de 32 funcionários de uma empresa:

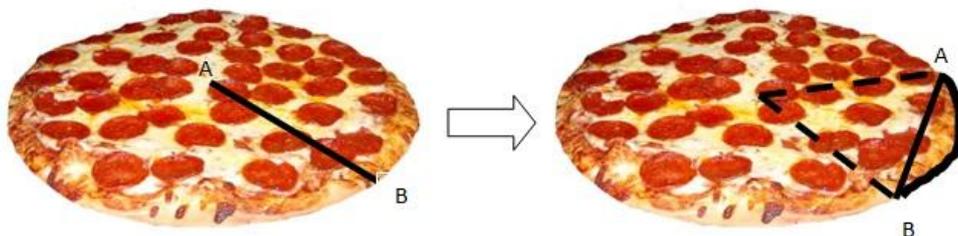
| FAIXAS (R\$)    | FREQUÊNCIA ACUMULADA DESCRITIVA              |
|-----------------|--|
| [800 ; 920[     | 10 funcionários ganham menos de 920 reais.   |
| [920 ; 1.040[   | 18 funcionários ganham menos de 1.040 reais. |
| [1.040 ; 1.160[ | 27 funcionários ganham menos de 1,160 reais. |
| [1.160 ; 1.280[ | 32 funcionários ganham menos de 1.280 reais. |
| <b>TOTAL</b>    | <b>32 FUNCIONÁRIOS</b>                       |

Sérgio e Raquel são os responsáveis únicos pela renda familiar. Se os 32 funcionários desta empresa forem colocados em fila, por ordem crescente de salário, Raquel e Sérgio ficariam, respectivamente, na 21<sup>a</sup> e 30<sup>a</sup> posição. Neste caso, a renda familiar em questão é

- A) R\$ 1.080,00.
- B) R\$ 1.232,00.
- C) R\$ 2.100,00.
- D) R\$ 2.312,00.
- E) R\$ 2.428,00.

**QUESTÃO 18:**

Um *pizzaiolo* tem uma técnica para fatiar uma pizza. Veja a figura:



Com um barbante, ele mede o raio AB e, em seguida, marca o tamanho da fatia com o raio AB; repete a demarcação um certo número de vezes em torno da pizza até circundá-la por completo. Então, a pizza fica dividida em

- A) 6 fatias iguais.
- B) 6 fatias.
- C) 7 fatias.
- D) 8 fatias.
- E) 8 fatias iguais.

### QUESTÃO 19:

Bruno e Carla jogam *cara ou coroa*, segundo as seguintes regras:

- Cada um lança uma moeda 4 vezes seguidas, decidindo-se quem começa no *par ou ímpar*.
- Vence cada partida aquele que conseguir primeiro uma sequência de três resultados iguais ou três resultados alternados.

Analise cada proposição abaixo e, em seguida, marque a alternativa que assinala a(s) proposição(ões) verdadeira(s).

I - A probabilidade de Bruno vencer na primeira partida é de 31,25%.

II- Se Carla ganhou o *par ou ímpar*, ela tem  $\frac{1}{4}$  de probabilidade de ganhar na primeira partida.

III - Se Carla não ganhou o *par ou ímpar* e os dois primeiros lançamentos de Bruno foram, respectivamente, coroa e cara, então, ele tem 75% de probabilidade para não vencer o jogo na primeira partida.

- A) Todas as alternativas.
- B) Apenas as alternativas I e II.
- C) Apenas as alternativas I e III.
- D) Apenas as alternativas II e III.
- E) Apenas uma das alternativas.

**QUESTÃO 20:**

A Prefeitura de Bonitópolis, num projeto de urbanização, distribuiu lotes, com a seguinte condição:

**Lotes retangulares que possam ser cercados com 84m de muro.**

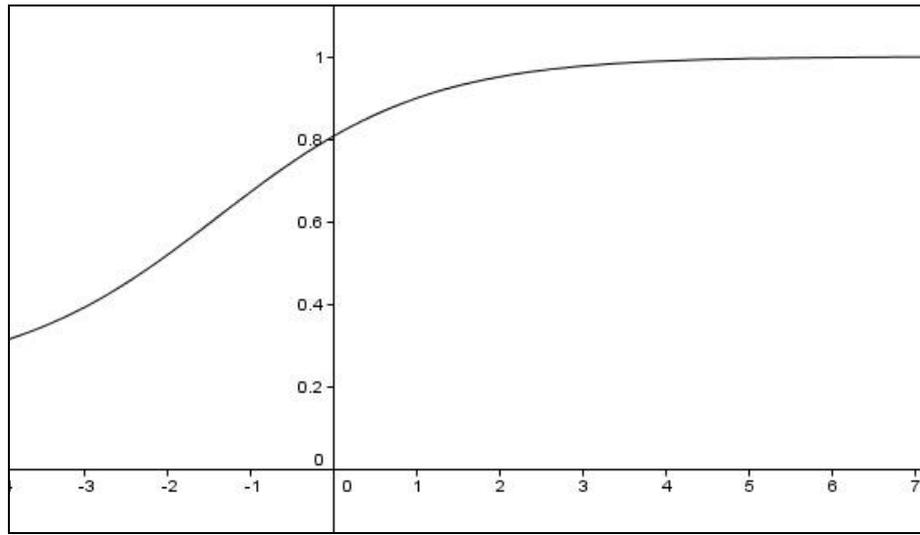
Nessas condições, o lote de maior área possível é um retângulo de lados medindo

- A) 12m por 30m.
- B) 20m por 22m.
- C) 21m por 21m.
- D) 25m por 26m.
- E) 36m por 15 m.

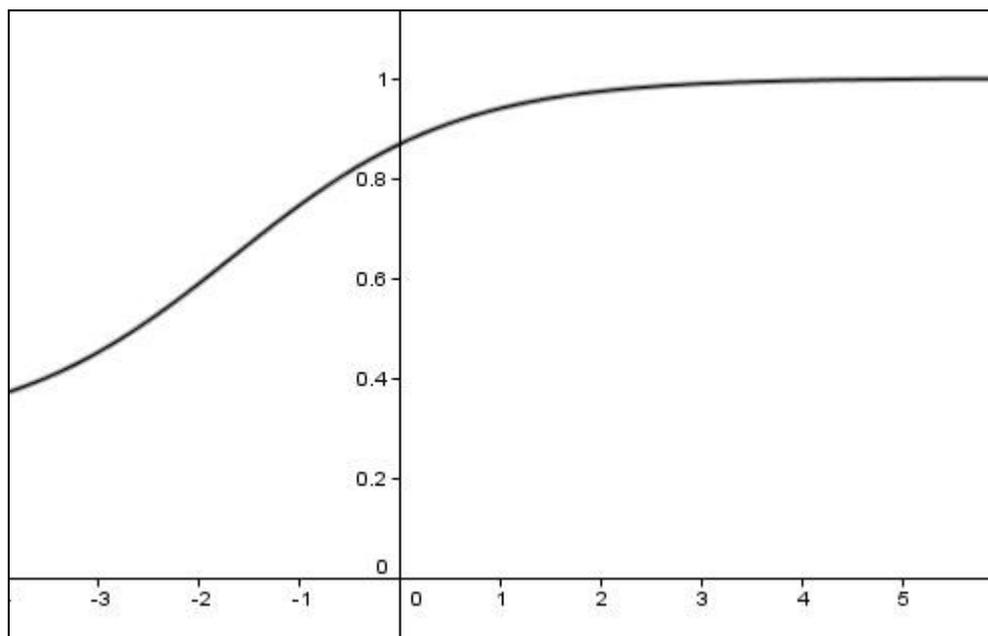
## Apêndice 2 - Curvas características dos itens (pré-teste)

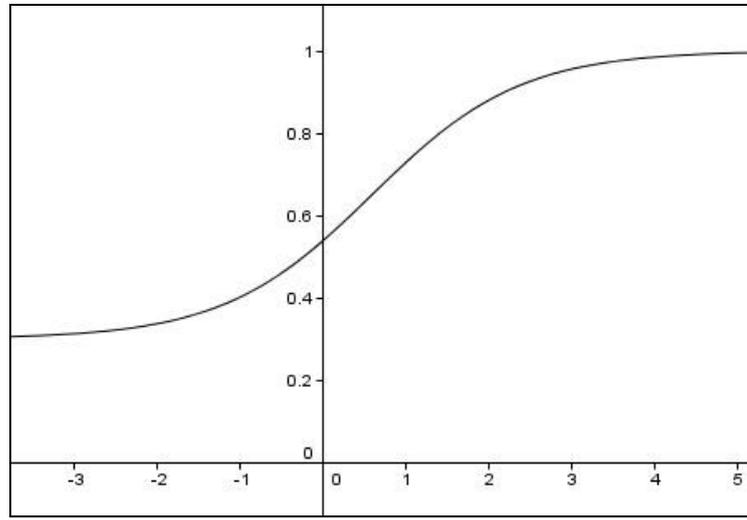
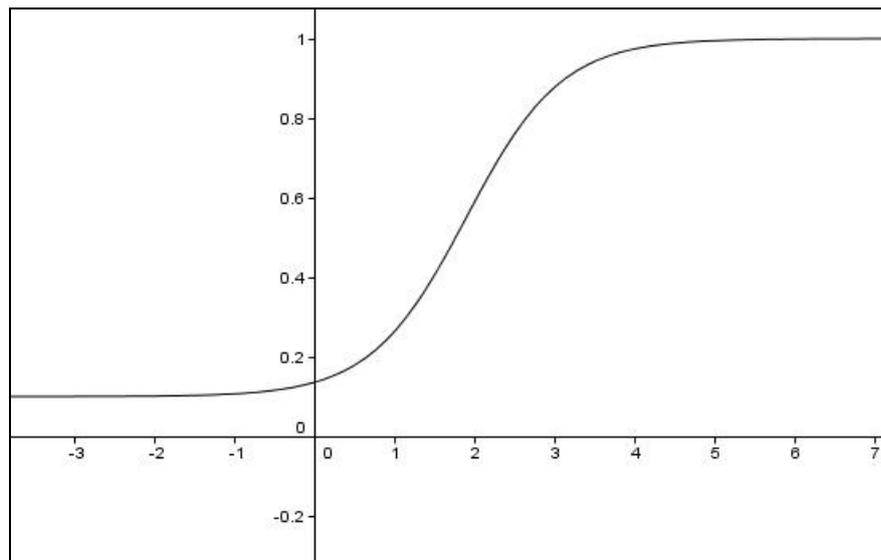
Todas as curvas seguintes foram traçadas com o *software* GEOGEBRA. O eixo horizontal representa a habilidade ( $\theta$ ) e o eixo vertical a probabilidade de acertos [ $P(u=1|\theta)$ ].

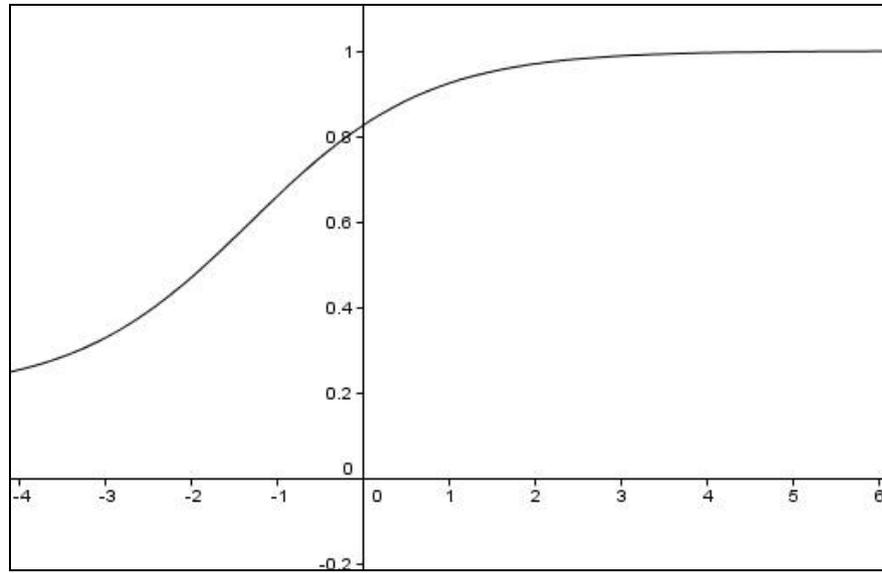
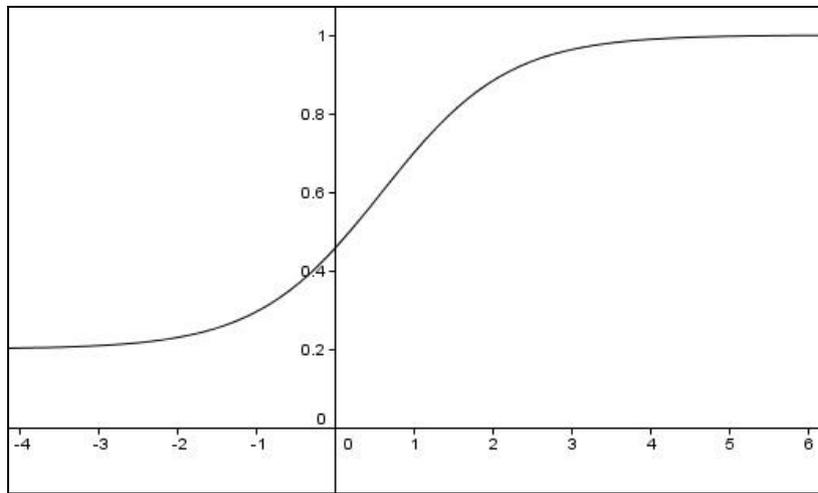
### ITEM 1



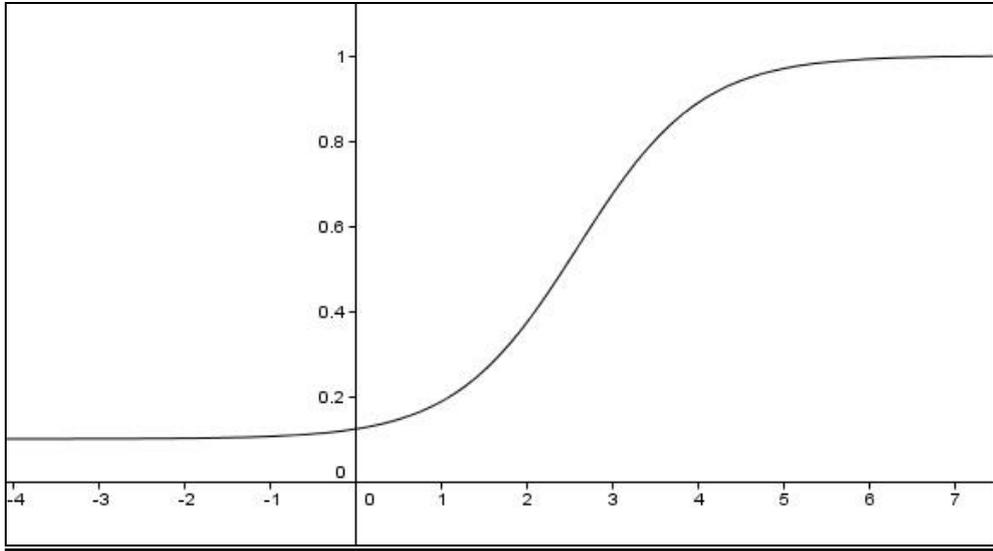
### ITEM 2



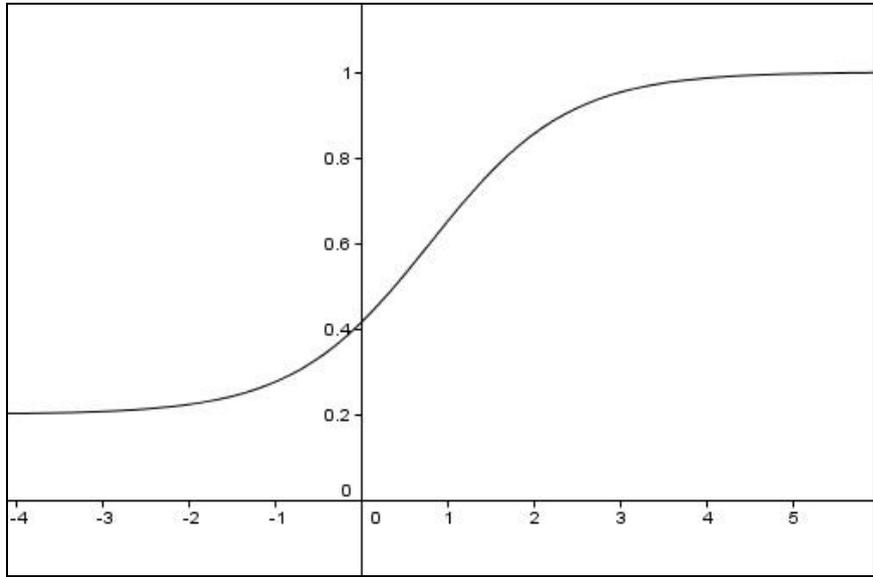
**ITEM 3:****ITEM 4:**

**ITEM 5:****ITEM 6:**

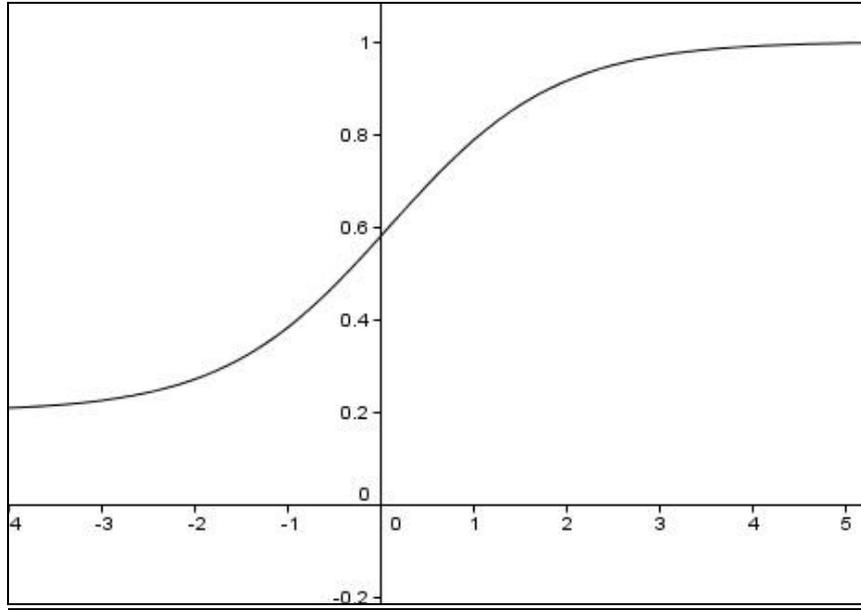
**ITEM7:**



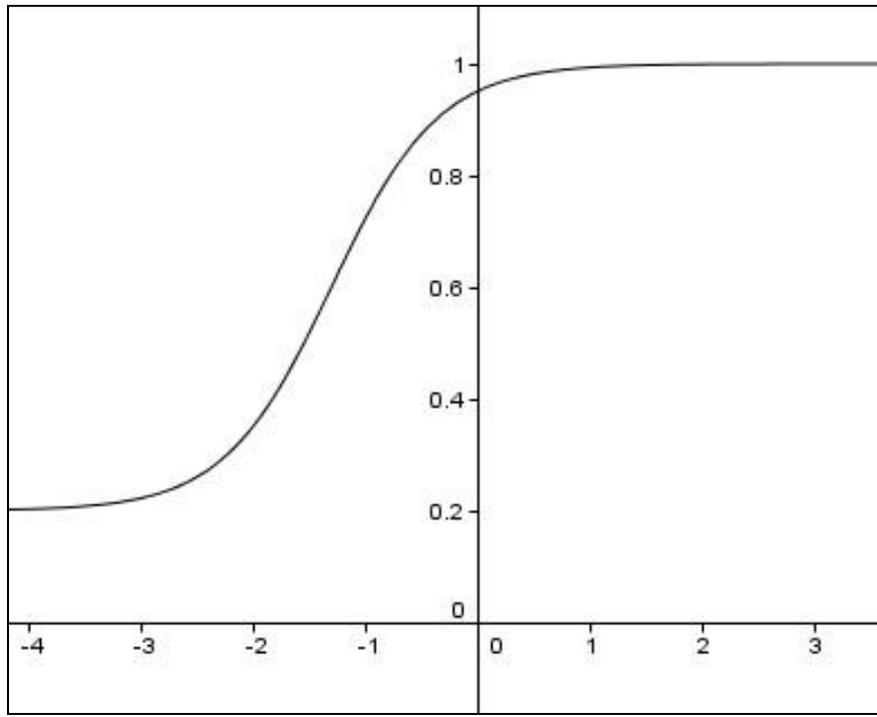
**ITEM9:**



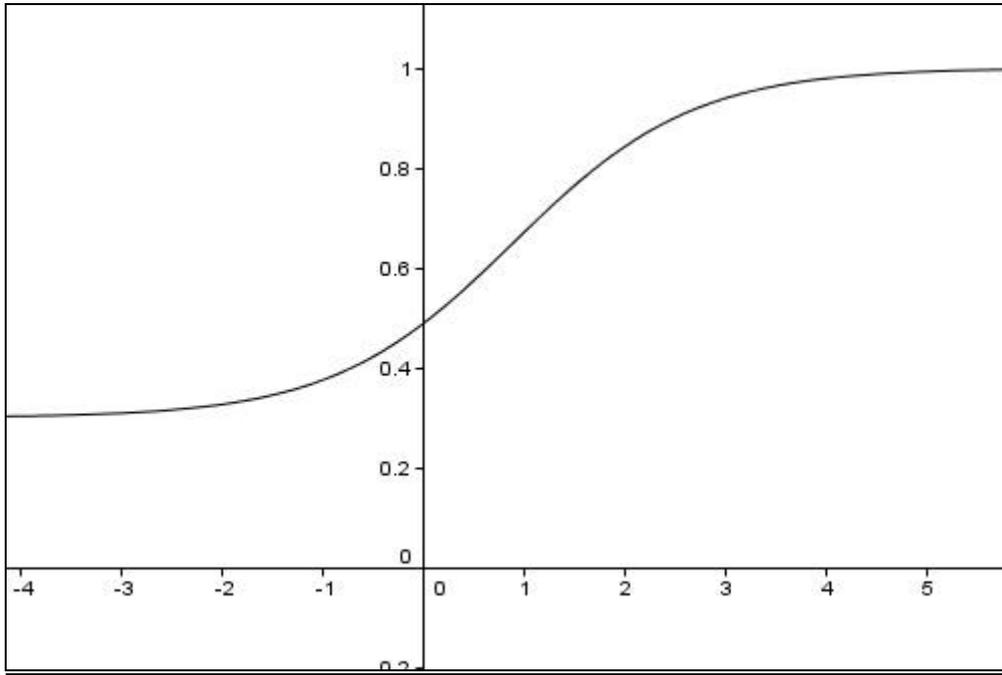
**ITEM 10:**



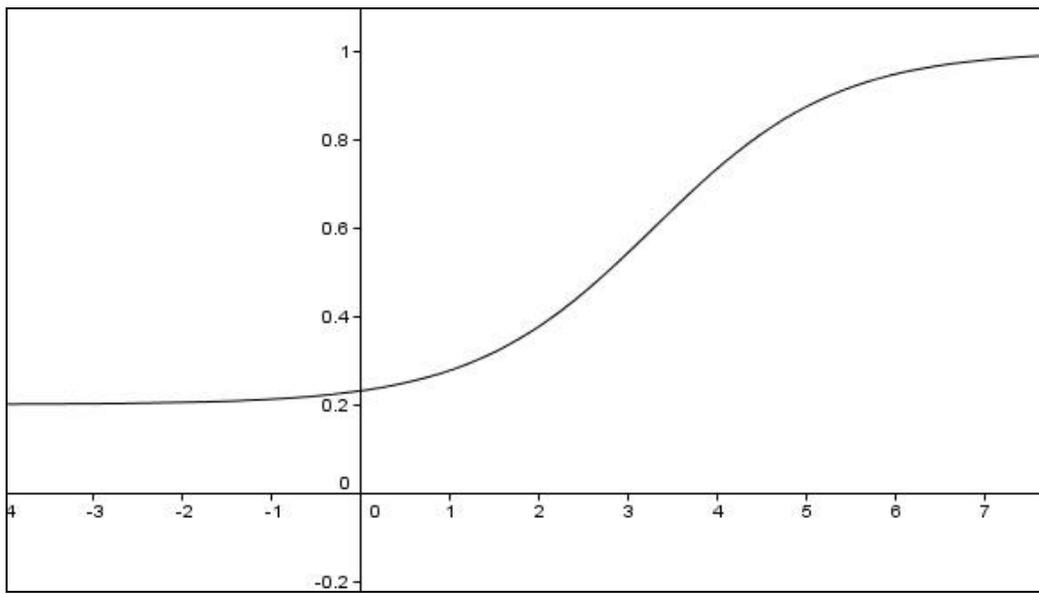
**ITEM 11:**



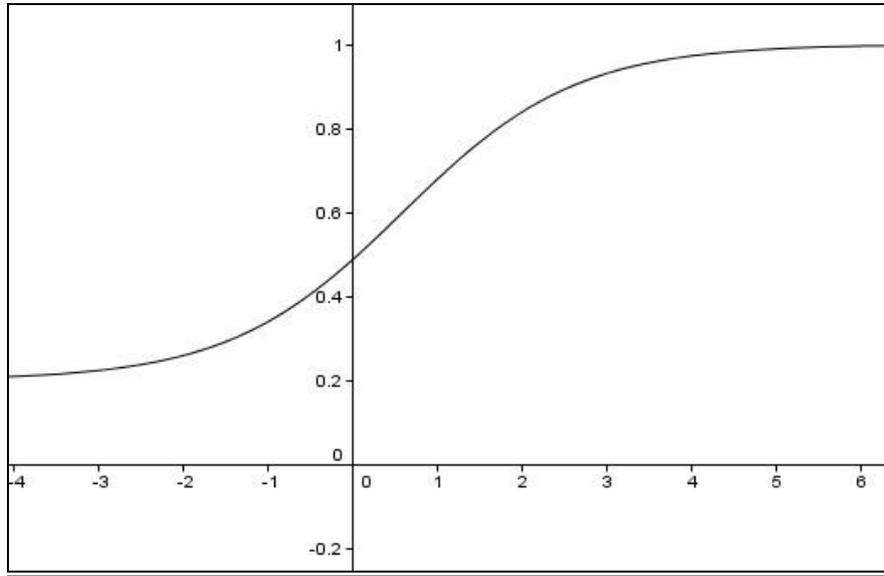
**ITEM 12:**



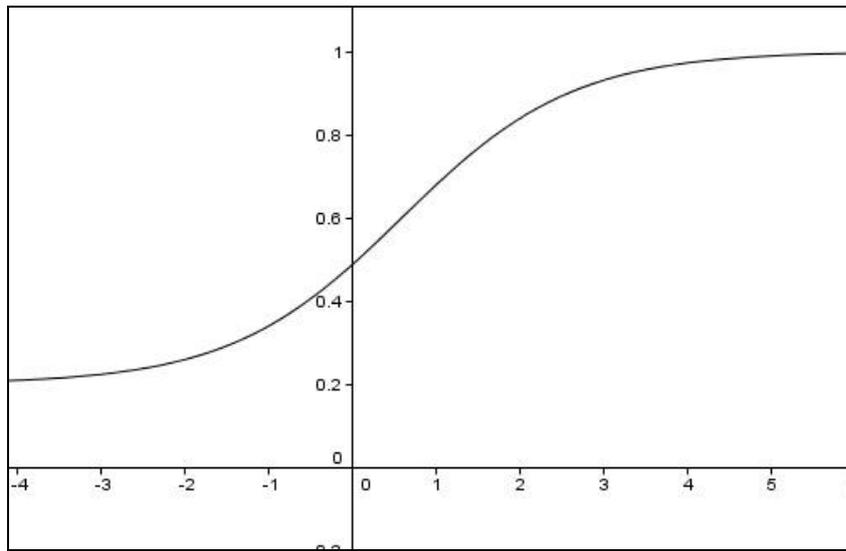
**ITEM 13:**



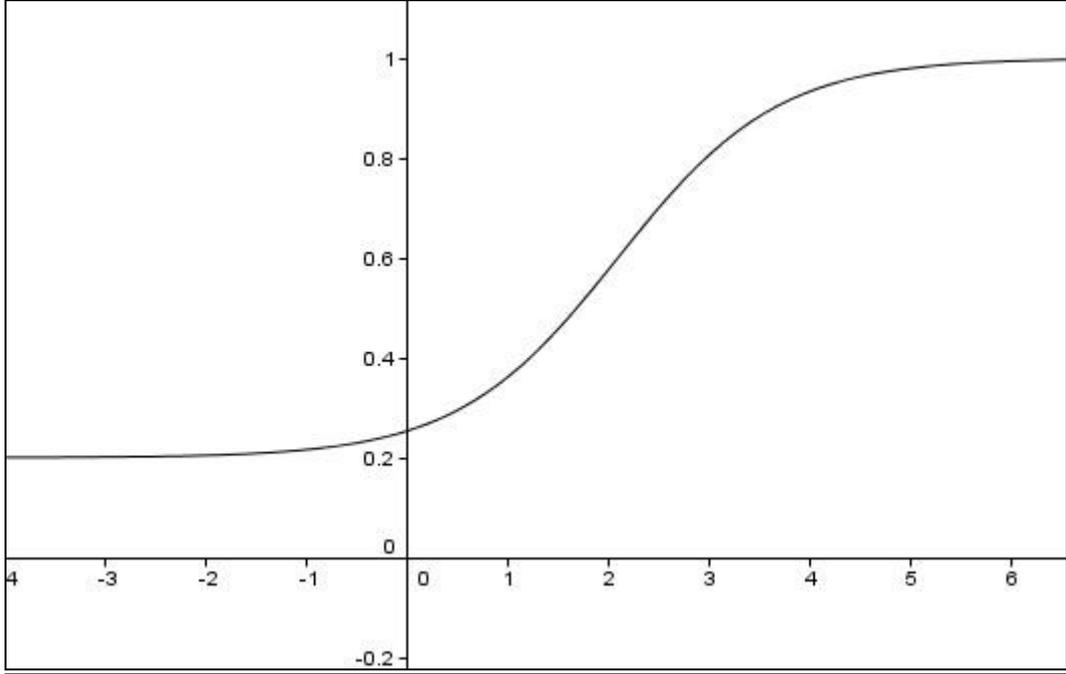
**ITEM 15:**



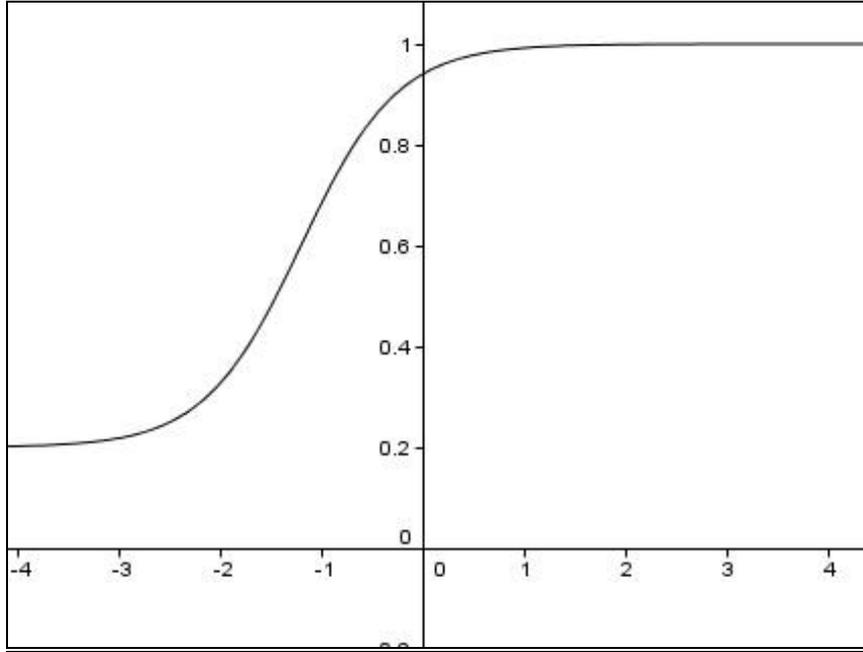
**ITEM 16:**



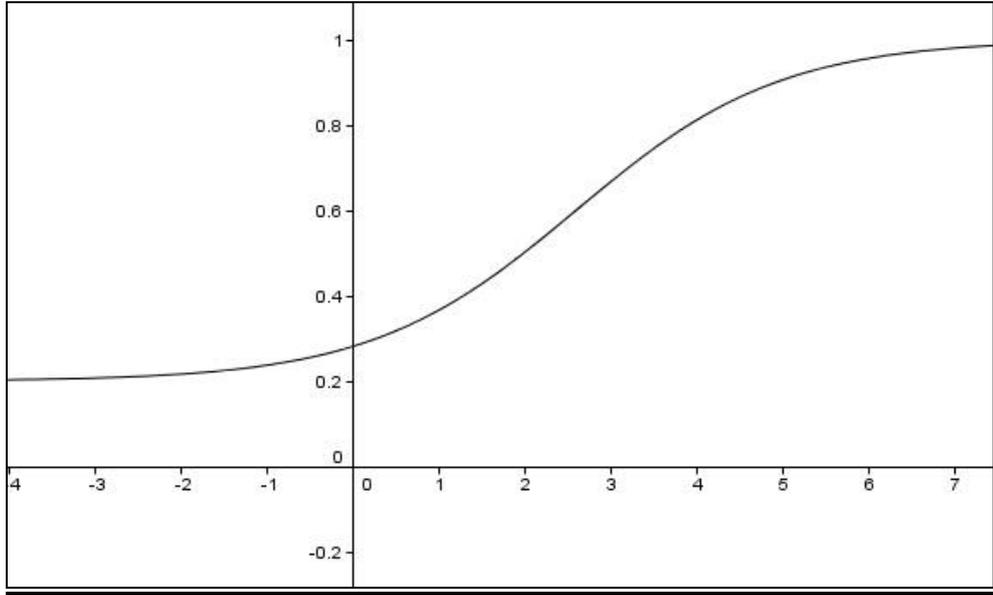
**ITEM 17:**



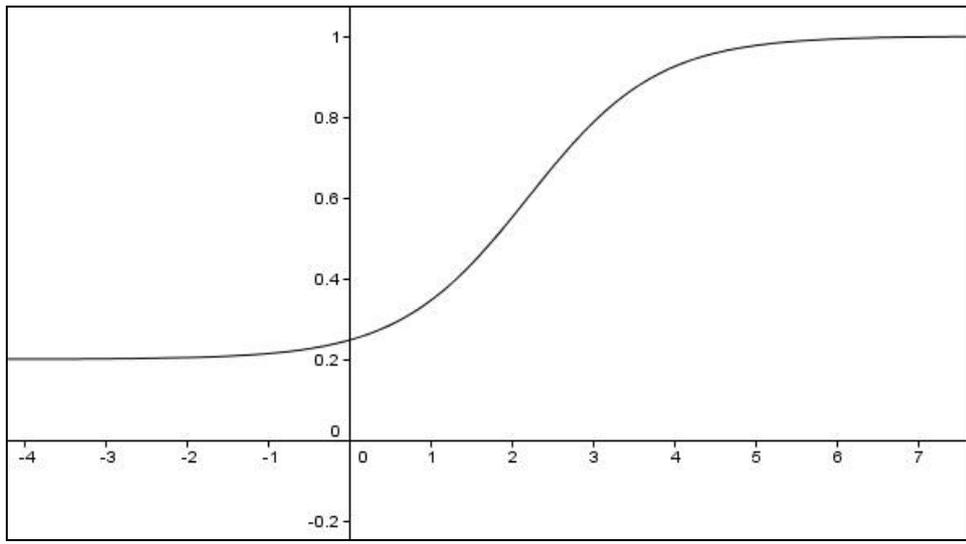
**ITEM 18:**



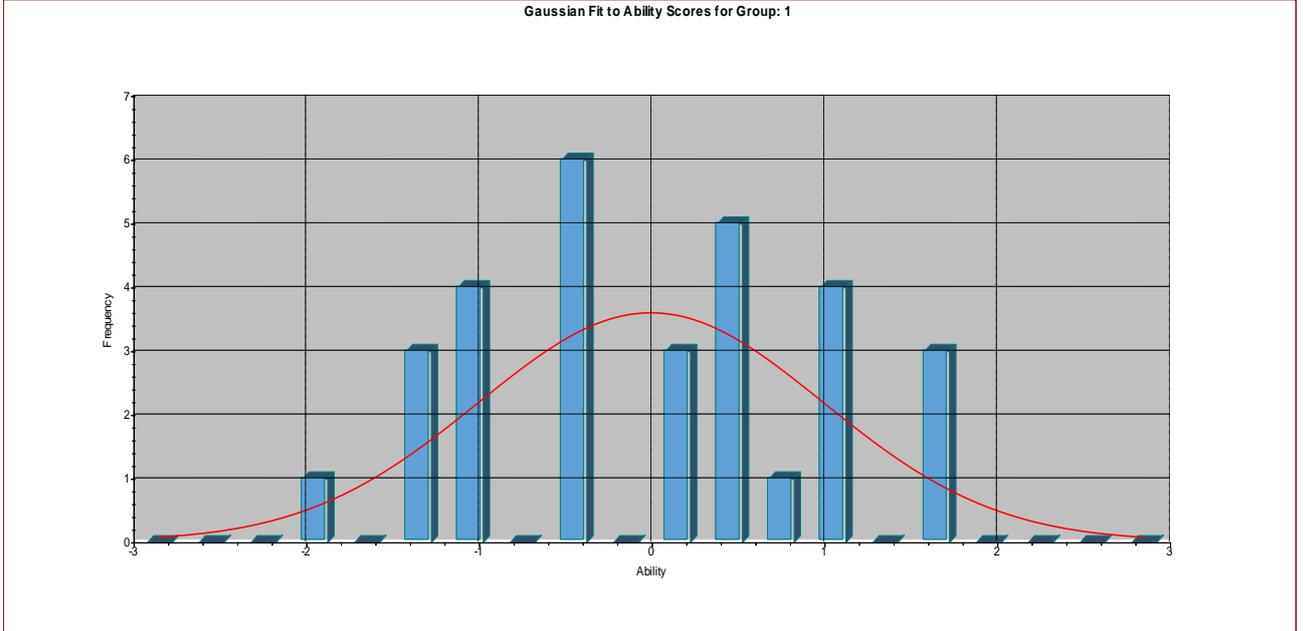
**ITEM 19:**



**ITEM 20:**



### Apêndice 3 - Histograma das habilidades



## Apêndice 4 - Base teórica do BILOG-MG

### 1- Script (sintaxe) do teste:

```

>GLOBAL DFName = 'C:\Users\Rogério\Desktop\TRIBILOG\BILOG\CONFER.DAT',
    NPArm = 3,
    SAVe;
>SAVE SCORe = 'CONFER.SCO';
>LENGTH NITems = (20);
>INPUT NTOtal = 20,
    NIDchar = 4,
    KFName = 'CONFER.KEY';
>ITEMS ;
>TEST1 TNAmE = 'MAT1',
    INUmber = (1(1)20);
(4A1, 20A1)
>CALIB ACCel = 1.0000,
    TPRior,
    GPRior;
>SCORE RSCtype = 1,
    LOCation = (500.0000),
    SCAle = (100.0000);

```

### 1- Procedimentos de estimação:

O BILOG-MG estima inicialmente os parâmetros dos itens e, posteriormente, as habilidades dos respondentes. Para a primeira estimativa, o programa está equipado para usar o processo MVM - Máxima Verossimilhança Marginal e o processo Bayesiano MMP - Maximização Marginal a Posteriori. As equações provenientes desses processos são resolvidas pelo Algoritmo EM ou pelo Método "Scoring" de Fisher. A estimação da habilidades pode ser feita por EAP - Esperança a posteriori, por EMV -Estimação por Máxima Verossimilhança ou ou MAP - Maximização a Posteriori. As equações oriundas desses processos são resolvidas, respectivamente, pelos processos iterativos de Newton-Raphson, Newton - Gauss e Quadratura de Pontos. Os problemas de teste perfeito (acerto

total) e teste nulo (erro total), atribuídos ao Algoritmo EM, são contornados com a atribuição de meio ponto na questão mais difícil e meio ponto na questão mais fácil, respectivamente.

## 2- Função de Verossimilhança – Estimação dos parâmetros dos itens

Estimar parâmetros por máxima verossimilhança, com base numa amostra, é descobrir, ente todas as distribuições possíveis com aquela amostra, qual a que tem maior probabilidade de estar associada a ela.

A função genérica de verossimilhança é

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta) \quad (1)$$

É necessário determinar a função específica para o presente caso; é do tipo  $f(\Theta/\zeta)$  em que  $\Theta$  é a habilidade conhecida e  $\zeta$  é o vetor de parâmetros dos itens ( $a_i$ ,  $b_i$ ,  $c_i$ ). Para isso, o BILOG-MG usa o *Método da Quadratura de Gauss*, segundo o qual, uma função  $g(\Theta/\eta)$ , contínua com integral finita, pode ser substituída por outra, de tal modo que ao invés de se integrar  $g(\Theta/\eta)$ , calcula-se a área de um certo número de retângulos, cujos pontos médios são os pontos de quadratura. Enfim, a função de verossimilhança pode ser escrita, nos moldes de quadratura:

$$g_j^* = \frac{P(u_j | \bar{\theta}_k; \zeta) A_k}{\sum_{k=1}^q P(u_j | \bar{\theta}_k; \zeta) A_k} \quad (2)$$

Daí, as equações de estimação dos parâmetros  $a_i$ ,  $b_i$  e  $c_i$ :

$$a_i : D(1 - c_i) \sum_{j=1}^n \sum_{k=1}^q r_j [(u_{ji} - P_{ki})(\bar{\theta}_k - b_i) w_{ki}] g_j^*(\bar{\theta}_k) = 0 \quad (3)$$

$$b_i : -D(1 - c_i) \sum_{j=1}^n \sum_{k=1}^q r_j [(u_{ji} - P_{ki}) w_{ki}] g_j^*(\bar{\theta}_k) = 0 \quad (4)$$

$$c_i : \sum_{j=1}^n \sum_{k=1}^q r_j [(u_{ji} - P_{ki}) \frac{w_{ji}}{P_{ki}^*}] g_j^*(\bar{\theta}_k) = 0 \quad (5)$$

### 3- Estimaco bayesiana das habilidades:

A funo de verossimilhana  

$$g^*(\Theta_j) \cong g(\theta_j | u_j; \zeta; \eta) \propto P(u_j; \zeta) g(\theta_j | \eta) \quad (6)$$

Por estimaco da Moda da posteriori e aplicando logaritmo, tem-se

$$\log g^*(\Theta_j) = \text{constante} + \log P(u_j; \Theta_j; \zeta) + \log g(\theta_j | \eta) \quad (7)$$

derivando para maximizar, tem-se

$$\partial \log g^*(\Theta_j) = \frac{\partial \log P(u_j | \Theta_j; \zeta)}{\partial \Theta_j} + \frac{\partial \log g(\theta_j | \eta)}{\partial \Theta_j} \quad (8)$$

$$\begin{aligned} \text{Pela independ ncia local, tem-se } \log P(u_j | \Theta_j; \zeta) &= \log[\prod_{i=1}^I P(u_i | \zeta; \Theta_j)] = \\ &= \sum_{i=1}^I \log P(u_{ji} | \zeta; \Theta_j) \quad (9) \end{aligned}$$

Substituindo-se 9 em 8 e lembrando-se que  $P(u_{ji} | \zeta; \Theta_j) = P_{ji}^{u_{ji}} \cdot Q_{ji}^{1-u_{ji}}$ ,

Tem-se a equao para a estimaco das habilidades:

$$\theta_j : D \sum_{i=1}^I a_i(1 - c_i)(u_{ji} - P_{ji})W_{ji} - \frac{(\theta_j - \mu)}{\sigma^2} = 0. \quad (10)$$

### Apêndice 5 - codificação numérica das respostas:

As respostas dos alunos, assim como o gabarito, entram no BILOG-MG codificadas de um dos modos a seguir:

#### **gab 42441320514214241652**

001 22414132511131315423

002 11331132531111125244

003 22123144531314215144

004 22241113533313214542

005 12114423524142315151

006 32421333242312115132

007 22341123542112154421

009 21442231512445242255

010 33442231415442231115

011 43331231523552431452

012 12411234513115221455

013 34241324334115144552

015 22341214515213124245

016 32231324554233122142

017 32231224515314254234

018 22441215515235451214

019 22441424514335122255

020 22441544115224423342

021 35445454524214214352

022 34231324514354224251

023 12411314514312522142

024 12241424514235221154

025 12441324514334221224

026 12321454512214224152

027 12521324554215424131

028 32241354515214224451

029 32441324514214523251

030 32411324514244222132

CADA LINHA TEM O SEGUINTE FORMATO:

→AS TRÊS PRIMEIRAS COLUNAS FORMAM O NÚMERO

DO ALUNO COM TRÊS DÍGITOS: 001, 002, ..., 019, 020.

→A QUARTA COLUNA É O ESPAÇO EM BRANCO.

→ CADA COLUNA SEGUINTE É A RESPOSTA DE UM ITEM

DE ACORDO COM O CÓDIGO A = 1, B= 2, C = 3, .....

→A PRIMEIRA LINHA EQUIVALE AO GABARITO.

01 01100001110000000000  
02 00001001100010010000  
03 01000000100011100100  
04 01011000100010101001  
05 01000010101000000110  
06 01101101000010000101  
07 01011011100010001000  
09 10110001110000100010  
10 00110001010000100100  
11 00001001100000000011  
12 01101000110010110010  
13 00011111001010001011  
15 01011001110110011000  
16 01001111101100010101  
17 11001011110011101000  
18 01111000110100000000  
19 01111010111000010010  
20 11111001010101010001  
21 00110001101111101011  
22 10001111111001111010  
23 01101101111011010101  
24 11011011111100110110  
25 11111111111001110000  
26 11001001110111111111  
27 11001111101110011100  
28 01011101110111111010  
29 11111110111111010010  
30 1110111111110111010

CADA LINHA TEM O SEGUINTE FORMATO:

→ AS DUAS PRIMEIRAS COLUNAS FORMAM O NÚMERO DO ALUNO COM DOIS DÍGITOS: 01, 02, ..., 19, 20.

→ A TERCEIRA COLUNA É O ESPAÇO EM BRANCO.

→ CADA COLUNA SEGUINTE É A RESPOSTA DE UM ITEM DE ACORDO COM O CÓDIGO CERTA = 1, ERRADA = 0.

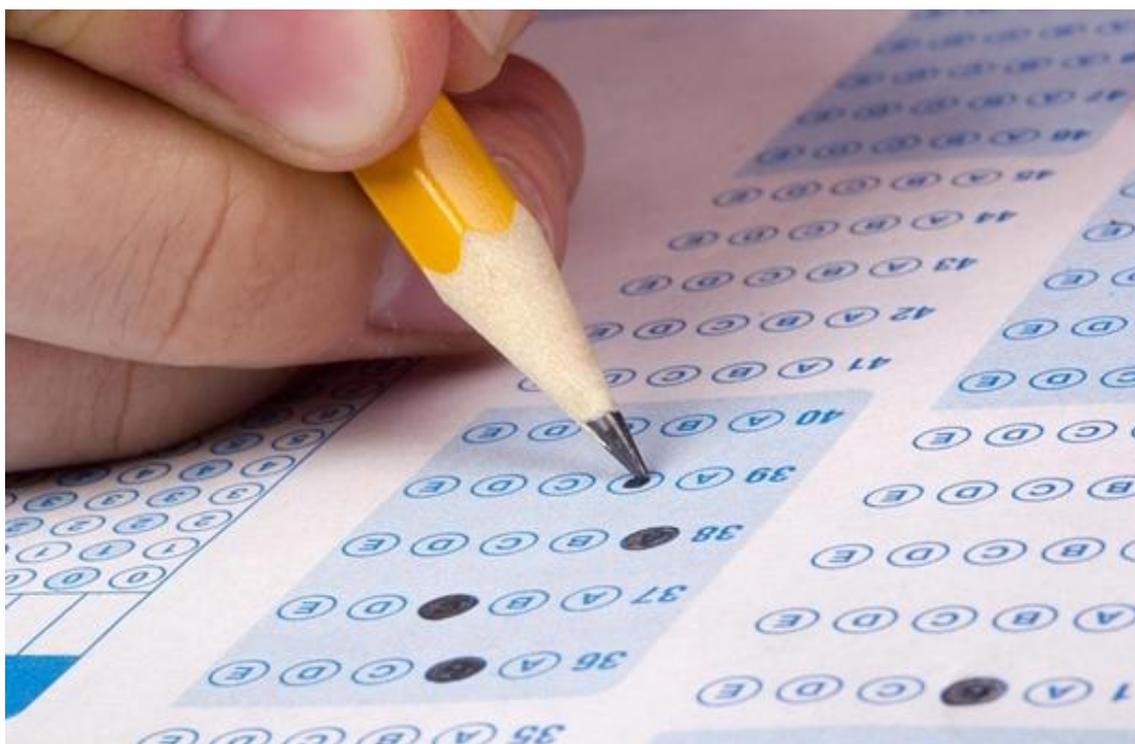
→ O GABARITO É 11111111111111111111

## Apêndice 6 - Produto Banco de Itens Planum

Programa de Pós-Graduação em Ensino de Ciências e Matemática  
da Pontifícia Universidade Católica de Minas Gerais.

# Planum

BANCO DE ITENS POR PLANILHAS



# **Planilium**

BANCO DE ITENS POR PLANILHAS

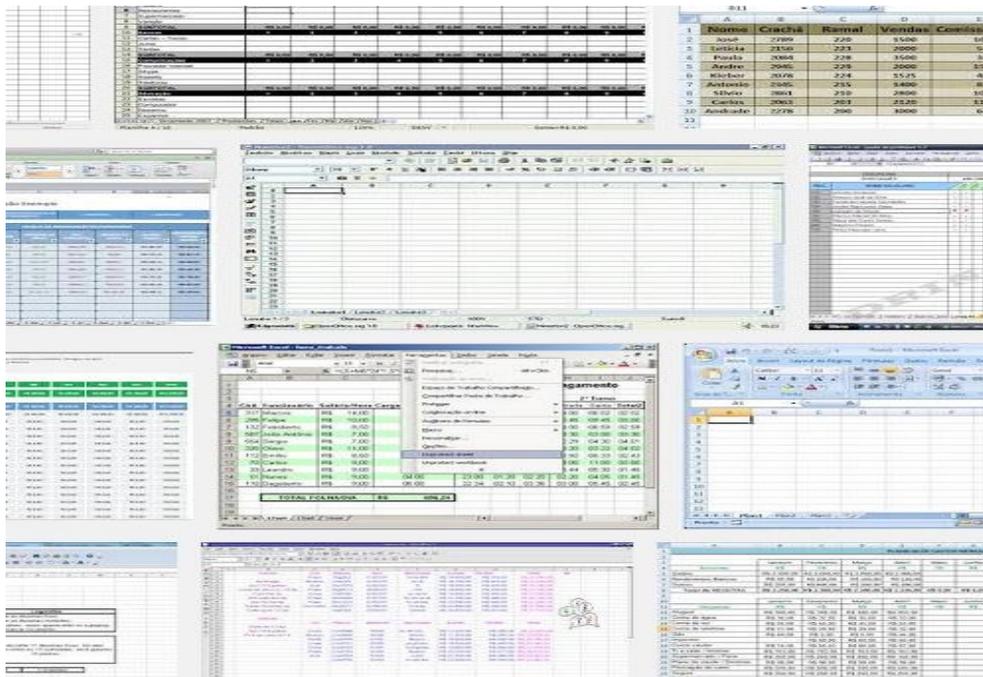
Rogério Rodrigues

Tânia Fernandes Bogutchi

Eliane Scheid Gazire

Produto integrante da avaliação do Programa de Pós-Graduação em  
Ensino de Ciências e Matemática da Pontifícia Universidade Católica  
de Minas Gerais.

Belo Horizonte  
2015



*"A imaginação é mais importante que a inteligência"*  
*(Albert Einstein)*

## SUMÁRIO

|  |            |
|--|------------|
| <b>1- APRESENTAÇÃO .....</b>   | <b>116</b> |
| <b>2- DINÂMICA GERAL DO PLANUM .....</b>   | <b>118</b> |
| <b>3- DIAGRAMA OPERACIONAL DO PLANUM .....</b>                                     | <b>119</b> |
| <b>4- A BASE TEÓRICA DA PLANILHA FILTRO.TCT: .....</b>                             | <b>121</b> |
| <b>4.1 Coleta de resultados do Pré-teste e primeiras medidas e apurações .....</b> | <b>121</b> |
| <b>4.2 Correlação de Pearson (r) .....</b>   | <b>122</b> |
| <b>4.3 Fórmula de Spearman Brown .....</b>   | <b>122</b> |
| <b>4.4 Diferença D.....</b>  | <b>122</b> |
| <b>4.5 Correlação Bisserial .....</b>  | <b>122</b> |
| <b>4.6 Correlação Bisserial por ponto.....</b>                                     | <b>123</b> |
| <b>4.7 Considerações Complementares .....</b>                                      | <b>123</b> |
| <i>4.7.1 Correlação bisserial e a variância do item .....</i>                      | <i>123</i> |
| <i>4.7.2 Atratividade das alternativas de cada item .....</i>                      | <i>123</i> |
| <i>4.7.3 Resposta errada e escore inferior à média.....</i>                        | <i>123</i> |
| <b>5- ARQUIVOS INDEPENDENTES.....</b>  | <b>124</b> |
| <b>5.1 Habilidades.....</b>  | <b>124</b> |
| <b>5.2 Itens e Resoluções .....</b>  | <b>124</b> |
| <b>5.3 PICS .....</b>  | <b>124</b> |
| <b>6- ROTEIRO FUNCIONAL DO PLANUM .....</b>  | <b>125</b> |
| <b>6.1 Início .....</b>  | <b>125</b> |
| <b>6.2 Primeiro menu .....</b>   | <b>126</b> |
| <b>6.2.1 Pesquisa e seleção de itens .....</b>                                     | <b>127</b> |
| 6.2.1.1 Seleção por ramo da matemática .....                                       | 128        |
| 6.2.1.2 Seleção por conteúdo específico.....                                       | 129        |
| 6.2.1.3 Seleção por habilidade.....  | 129        |
| <b>6.2.2 Conhecer atributos do banco .....</b>                                     | <b>130</b> |
| <b>6.2.3 Inserir novos itens .....</b>   | <b>130</b> |
| 6.2.3.1 A Planilha INSERIR ITENS .....   | 130        |
| 6.2.3.2 Inserir Itens Não calibrados .....   | 131        |
| <b>7- LEITURAS SUGERIDAS.....</b>  | <b>131</b> |

## 1- APRESENTAÇÃO

O aplicativo que ora se apresenta é um banco de itens calibrados para compor uma avaliação em Matemática, considerando-se, como referência, a Matriz do ENEM - Exame Nacional do Ensino Médio. Como o elemento instrumental predominante nos cálculos de medidas e no registro organizado é objeto do Excel, optou-se por um banco articulado com planilhas do Excel nas funções básicas de cadastro (inclusão de itens), calibragem, armazenagem, Pesquisa e demandas afins.

A ideia desse aplicativo é suprir uma carência que, na maioria dos casos, nem é percebida, dada a prática comum nos procedimentos pedagógicos ligados ao estabelecimento de parâmetros avaliativos de aquisição no campo da aprendizagem. Construir um instrumento de avaliação que, de fato, avalie com presteza e coerência a eficiência de um trabalho de formação em suas duas interfaces - a escola e o aprendiz- é uma missão que quase sempre foi subestimada em sua importância. Com o desenvolvimento da informática, a Teoria de Resposta ao Item - TRI, pôde se tornar um modelo apropriado de avaliação quantitativa, posto que contempla as variáveis qualitativas, graças à natureza dos métodos estatísticos. Além disso, valoriza-se a Teoria Clássica dos Testes - TCT, estabelecendo-se uma relação de complementaridade entre as duas teorias.

A trajetória do item no Planum começa com o cadastro, passa por uma avaliação prévia com instrumentos da Teoria Clássica dos Testes, submete-se às recomendações empíricas de pesquisadores e, então, com o uso de algum software especializado, que contemple a TRI - Teoria de Resposta ao Item, como o R ou o BILOG-MG, passa por uma calibragem, antes de ser incorporado ao acervo do banco. O Planum foi pensado para receber testes com até 20 itens por vez; caso algum item não seja considerado fidedigno, ele poderá ser substituído, excluído ou modificado antes de submetido à TRI ou depois disso.

O Planum conta com 3 pastas de arquivos básicos de dados, referentes aos itens que possui em seu acervo, a saber:

-Pasta *Resoluções*, com todos os itens separados individualmente em arquivos PDF e as respectivas resoluções em arquivo Word.

-Pasta *Habilidades*, com a matriz de referência do ENEM e cada habilidade separada individualmente em arquivo de imagem jpg.

-Pasta *Pics*, com a visualização individual de cada item em arquivo de imagem jpg.

A pasta de trabalho *Planum.Geral* compreende um conjunto de 18 planilhas do Excel, sendo 6 de acesso geral (*Alfa, Beta, Gama, Início, Entrada e Seleção e Pesquisa*), 5 de informações sobre os itens (*conteúdo, habilidades, psicometria, atratividade e procedência*), 2 de apoio ao procedimento de busca (*Pce e Phab*) e 5 dos ramos da matemática, que são de registro centralizado de todas as características dos itens; essas planilhas detalham atributos dos itens como procedência (origem e ano), nível de dificuldade (difícil, média ou fácil), resposta e dados psicométricos como dificuldade, discriminação e possibilidades do acaso, além das atratividades das alternativas em cada item; são elas: *Aritmética, Álgebra, Geometria Plana, Geometria Espacial, Geometria Analítica, Estatística*. Independente da pasta *Planum.Geral*, há a planilha *Filtro.tct*, que executa, de acordo com a Teoria Clássica dos Testes, uma triagem de itens que serão submetidos à Teoria de Resposta ao Item. essa planilha calcula medidas da Estatística Descritiva, determina correlações e analisa a fidedignidade e a consistência dos testes.

No *Planum*, a identidade do item é o Código de Referência do Item - CRI, constituído de três algarismos e duas letras. As duas letras referem-se ao ramo de origem da questão: Aritmética (AT), Álgebra (AL), Geometria Plana (GP), Geometria Espacial (GE), Geometria Analítica (GA) e Estatística (ES). Os três algarismos numeram os itens de cada ramo; assim, tem-se, por exemplo, 003ES, questão 3 de Estatística.

Espera-se que o *Planum* se revele de notável relevância, principalmente para os professores preocupados em construir o conhecimento com seus alunos, avaliando suas aquisições de maneira transparente e real.

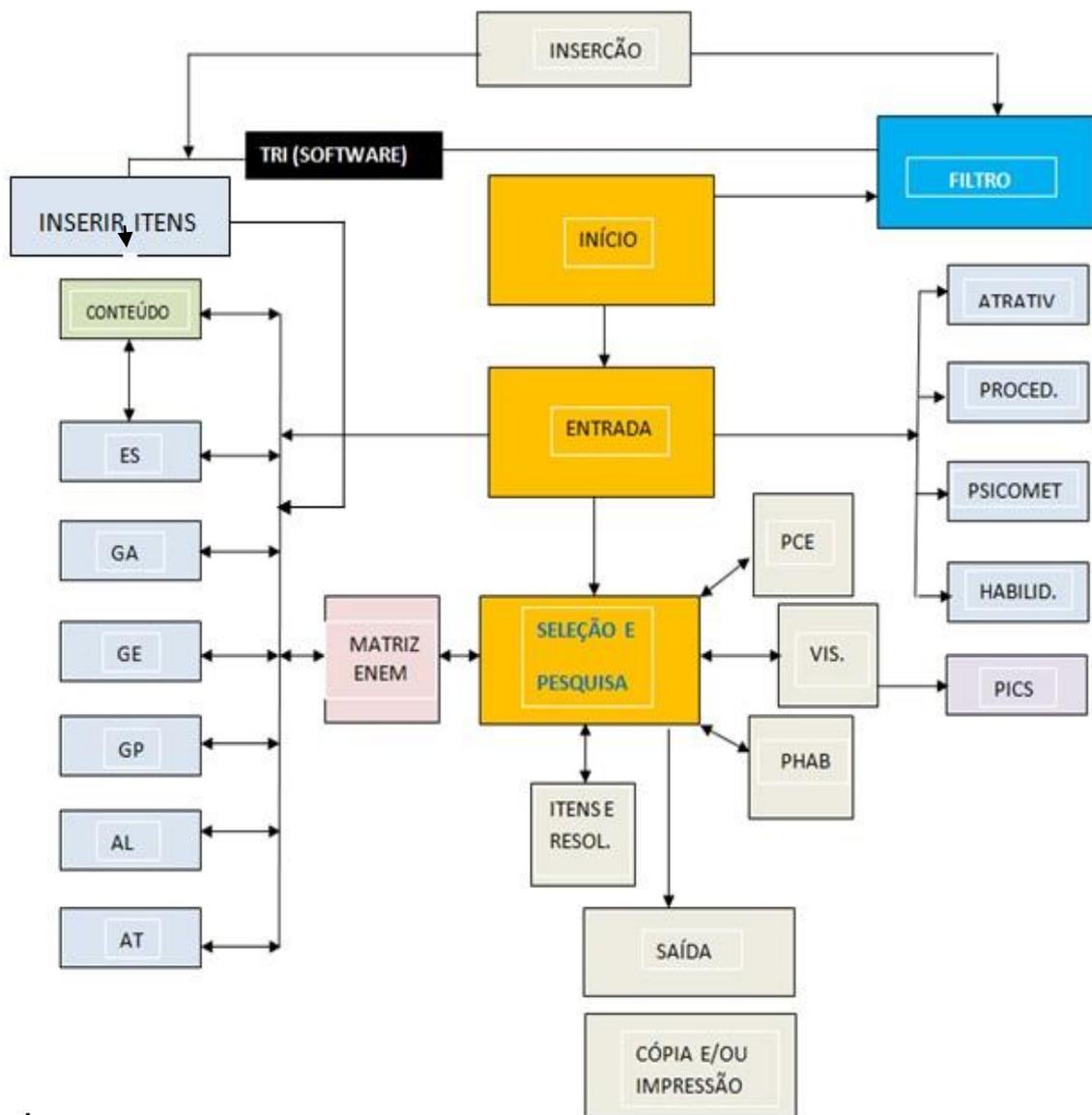
## 2- DINÂMICA GERAL DO PLANUM

Um item entra no Planum em um dos dois estados: calibrado por algum software ou apenas submetido ao pré-teste. Nesse último caso, ele será pré-calibrado numa planilha chamada *Filtro.tct*; depois disto, fará a calibragem definitiva no BILOG -MG ou em outro software que execute a Teoria de Resposta ao Item. No primeiro caso, ou seja, se o item já foi calibrado, então ele já entra direto para o acervo do Planum.

O acesso primário ao Planum é feito na planilha *Início*, depois de se passar por uma abertura explicativa com três planilhas auxiliares (*Alfa*, *Beta* e *Gama*). Na planilha *Início*, o usuário pode optar por conhecer alguns atributos aplicados aos itens, que são os arquivos de dados *Habilidades*, *Atratividade*, *Conteúdos*, *Procedência* e *Psicometria*, além dos arquivos de ramos da Matemática: *Aritmética*, *Álgebra*, *Geometria Plana*, *Geometria Espacial*, *Geometria Analítica* e *Estatística*.

Outra opção para o usuário, na planilha *Início*, é a seleção de itens, atribuição precípua do Planum. Neste caso, o usuário poderá, na sequência escolher itens por ramo da Matemática, por conteúdo específico ou por habilidade. A terceira opção disponível na planilha *Início* é usar o *Filtro.tct* para pré- calibrar itens. Assim sendo, deve-se entrar com um cartão de respostas de um pré-teste já aplicado. as respostas deverão estar no formato binário, ou seja, 1 para respostas corretas e 0 para as incorretas. É necessário um mínimo de 10 itens.

### 3- DIAGRAMA OPERACIONAL DO PLANUM



|

A planilha Filtro.tct tem a função de analisar previamente os itens agrupados em testes e avaliações, antes que sejam submetidos à Teoria de Resposta ao Item. Esse trabalho, feito com as ferramentas da Teoria Clássica dos Testes, é uma triagem dos melhores itens, mostrando que as duas teorias se complementam. Nessa função, a planilha Filtro.tct procura ponderações sobre o nível dos itens e a contribuição de cada um deles para o equilíbrio da avaliação. Então, essa planilha estrutura distribuições das respostas de um pré- teste, calcula as medidas tradicionais da Estatística Descritiva e avalia a capacidade de discriminação dos itens, assim como sua fidedignidade. Para isso, verifica o grau de correlação dos itens por métodos diversos como Pearson, Kuder - Richardson, Bisserial, Normalidade. A planilha Filtro.tct ainda considera algumas ponderações empíricas de pesquisadores em avaliação escolar. Nesse mister, as fórmulas e processos são estruturados em forma de tabela. A consulta à Tabela de Distribuição Normal Padronizada é feita dentro da própria planilha, sendo um setor dela.

As planilhas dos ramos da Matemática, entenda-se, Aritmética, Álgebra, Geometria Plana, Geometria Espacial, Geometria Analítica e Estatística detém todo o acervo de itens, organizados em seu devidos escaninhos. As demandas de consulta são atendidas pelo fluxo a partir dessas planilhas. Os atributos de cada item, assim como seus parâmetros psicométricos estão arquivados nessas planilhas. Enfim, todo o histórico de cada item está ali armazenado.

#### 4- A BASE TEÓRICA DA PLANILHA FILTRO.TCT:

Envolvida com as funções de medir e avaliar as potencialidades do item, a planilha FILTRO calcula as medidas de posição e de dispersão, depois da inserção de dispositivos lógicos e fórmulas na sintaxe própria do Excel. Assim, tem-se:

##### 4.1 Coleta de resultados do Pré-teste e primeiras medidas e apurações

Um teste tem acesso ao Planum na forma de uma planilha com o gabarito e as respostas dos alunos em duas versões; uma delas é na forma original das alternativas, ou seja, A,B,C, D e E. A outra forma é numérica do tipo 0 (erro) ou 1 (acerto). Em ambos os casos, cada linha tem as respostas de um aluno. A partir desses dados, são calculados:

- a) O escore de cada aluno
- b) O número de acertos por item
- c) o nível de dificuldade de cada item

$$C_{\text{DIFICULDADE}} = (\text{N}^{\circ} \text{ de alunos que acertaram o item}) / (\text{N}^{\circ} \text{ de alunos que fizeram o item})$$

- d) O número de alunos que marcaram, em cada item, cada alternativa
- e) Distribuições e medidas

Os escores brutos são, inicialmente, organizados em uma distribuição por variável e suas medidas de posição e de dispersão são calculadas , utilizando-se a comodidade da própria tabela de distribuição, que tem o seguinte formato:

| ESCORE (x) | f | f.x | F <sub>a</sub> | f. $ x - \bar{x} $ | f. $(x - \bar{x})^2$ |
|------------|---|-----|----------------|--------------------|----------------------|
|            |   |     |                |                    |                      |
|            |   |     |                |                    |                      |

$$\bar{x} = \frac{\sum f \cdot x}{\sum f}, \text{ Mo} = x / f_x \text{ é máximo}, \text{ Md} = x / \text{em } \{x_1,$$

$$x_2, \dots, x_n\}, X = X_{(1+n)/2}.$$

$$D_{\text{médio}} = \frac{\sum f \cdot |\Delta x|}{\sum f}, S^2 = \frac{\sum f \cdot (\Delta x)^2}{\sum f - 1}, S = \sqrt{S}, C_{\text{disp}} = \frac{S}{\bar{x}}$$

Os escores são distribuídos em classe, considerando-se o valor de cada item inversamente proporcional à sua dificuldade. Isso faz dos escores variáveis contínuas, ideais para uma distribuição com classe e para a distribuição normal padronizada, também articulada pela FILTRO. A tabela de probabilidades da Normal Padronizada é uma das ferramentas integrantes da planilha.

#### 4.2 Correlação de Pearson (r)

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum (x_i - \bar{x})^2)(\sum (y_i - \bar{y})^2)}}$$

#### 4.3 Fórmula de Spearman Brown

$$\alpha_{XY} = \frac{N \cdot r_{XY}}{1 + (n - 1)r_{XY}}$$

#### 4.4 Diferença D

É a diferença entre o nível de dificuldade calculado para os 27% maiores escores e aquele calculado para os 27% menores..

#### 4.5 Correlação Bisserial

$$\rho_{biss} = \frac{(\bar{X}_t - \bar{X}_p) \cdot \hat{\theta}}{s.Y}$$

#### 4.6 Correlação Bisserial por ponto

$$\rho_{bp} = \frac{\bar{X}_c - \bar{X}_e}{s} \sqrt{\frac{\theta}{1 - \theta}}$$

#### 4.7 Considerações Complementares

##### 4.7.1 Correlação bisserial e a variância do item

Para cada item,  $(Cbi)^2 \cong s^2 \geq (x - \bar{x})^2$  (verificação)

##### 4.7.2 Atratividade das alternativas de cada item

Equilíbrio entre os distratores e o descritor na atratividade, ou seja, na preferência geral dos alunos.

##### 4.7.3 Resposta errada e escore inferior à média

Em geral, o aluno que tem respostas erradas, em 50% dos casos, pelo menos, tem escore inferior à média da turma (verificação).

## **5- ARQUIVOS INDEPENDENTES**

Apesar de haver uma rede considerável de ligações entre os arquivos do Planum, os arquivos de dados como o Pics (arquivos jpeg), Habilidades e Itens e Resoluções são independentes, mas por uma questão de organização de fluxo, são acessados pelas vias gerais do aplicativo (Início, Entrada e Seleção e Pesquisa).

### **5.1 Habilidades**

Refere-se ao conjunto das 30 habilidades que norteiam a formulação dos itens que compõem o exame nacional (ENEM): H1 a H30. Nesse arquivo, estão relacionados os itens do banco e suas respectivas habilidades.

### **5.2 Itens e Resoluções**

Neste arquivo estão os itens do banco e suas resoluções detalhadas.

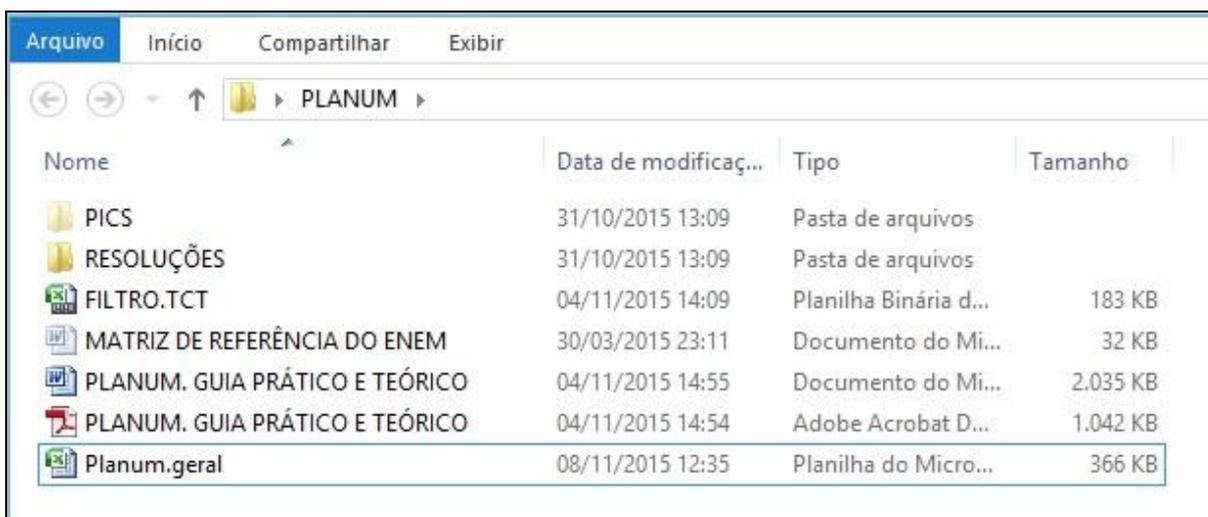
### **5.3 PICS**

Arquivos individuais dos itens para visualização (formato Jpeg).

## 6- ROTEIRO FUNCIONAL DO PLANUM

### 6.1 Início

Na pasta PLANUM, clicar no arquivo Planum.geral:



Após o clique, virá uma sequência de três telas de apresentação:





**Planum**  
SÉRIE DE RESPOSTAS ONLINE

Rogério Rodrigues

Produto integrante da avaliação do Programa de Mestrado em Educação Matemática da Pontifícia Universidade Católica de Minas Gerais.



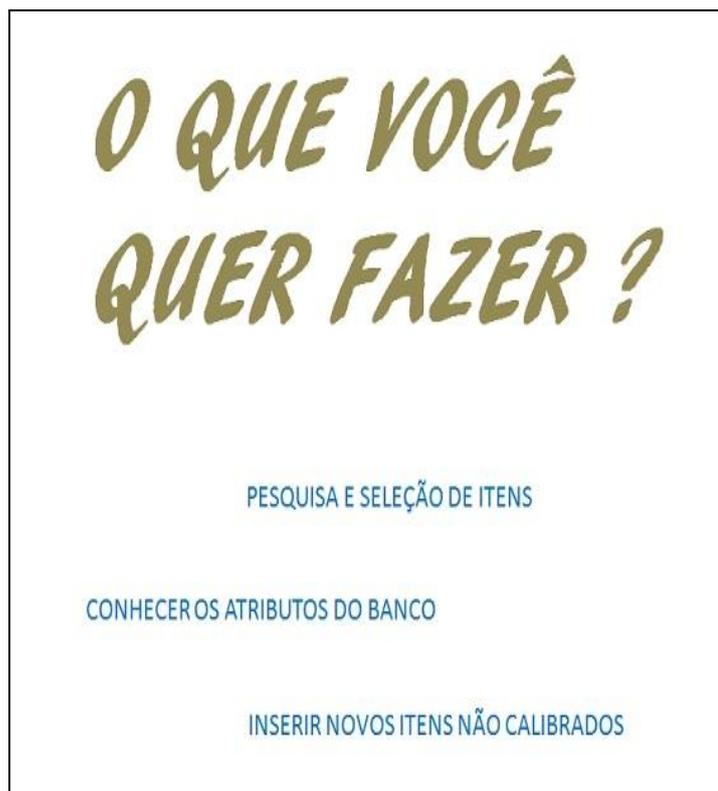
**Planum**  
SÉRIE DE RESPOSTAS ONLINE

Rogério Rodrigues

O presente aplicativo tem como atribuição principal a apresentação de itens de Problematiza (alinhado segundo a TRT – Teoria de Resposta ao Item). Nessa perspectiva, ele oferece toda a informação sobre os itens, tais como origem, resposta, nível de dificuldade, procedimento e informações parâmetros como os parâmetros de discriminação, dificuldade, probabilidade das alternativas no pré-teste e respostas de validade em alguns. Além disso, há a possibilidade de se fazer uma programação para, antes de submetermos à TRT uma PRC (APRENDIZAGEM) através do sistema e resultados da TRT – Teoria Clássica dos Testes.

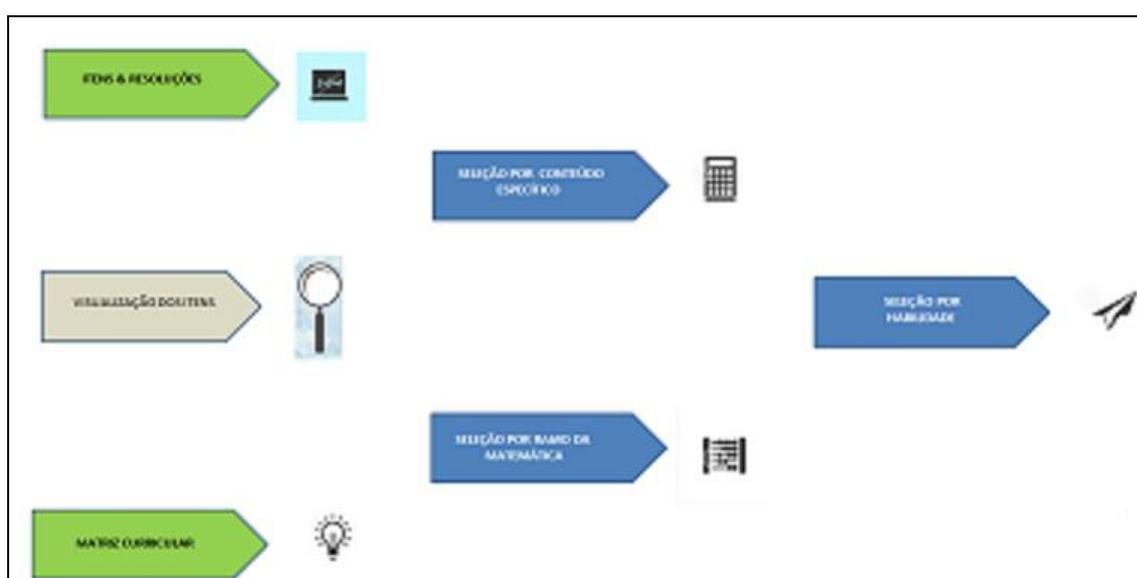
## 6.2 Primeiro menu

Clicando-se na seta à direita aparecerá a seguinte tela:



Clicando-se em *PESQUISA E SELEÇÃO DE ITENS*, abre-se o menu

### **6.2.1 Pesquisa e seleção de itens**









### 6.2.3.2 Inserir Itens Não calibrados

Como o acervo definitivo é de itens calibrados segundo a TRI - Teoria da Resposta ao Item, nessa opção abre-se um conjunto de quatro planilhas de pré-calibragem. Usando os critérios da TCT - Teoria Clássica de Testes, uma triagem prévia dos itens respondidos no pré-teste prepara-os para a TRI, segundo sugestão de SANTOS & SOARES (2012). Essas planilhas são: *Cálculo*, *Diagnóstico*, *Nova e Resultado*. As duas primeiras são memórias de cálculos, a terceira é de introdução de novos itens a serem analisados e a última é um gabarito-rascunho de respostas da reciclagem, essa emite pareceres conclusivos sobre os itens analisados.

## 7- LEITURAS SUGERIDAS

ANDRADE, Dalton Francisco de; TAVARES, Heliton Ribeiro; VALLE, Raquel da Cunha. **Teoria da Resposta ao Item: Conceitos e Aplicações**. São Paulo: SINAPE, 2000.

ANGOFF, W.H. Scales, norms and equivalent scores. In: R.L. Thorndike (Ed.). **Educational measurement**. 2.ed. Washington, DC, 1971. p.508-600.

BRASIL. INEP. **Prova ENEM 2012 - domingo**. 2012. Disponível em: [http://download.inep.gov.br/educacao\\_basica/enem/provas/2012/caderno\\_enem2012\\_dom\\_cinza.pdf](http://download.inep.gov.br/educacao_basica/enem/provas/2012/caderno_enem2012_dom_cinza.pdf). Acesso em: 10 dez. 2015.

BRASIL. Ministério da Educação e Cultura. **ENEM: a segunda maior prova de acesso ao Ensino Superior do mundo**. 2015. Disponível em: <http://www.brasil.gov.br/educacao/2015/10/enem-a-segunda-maior-prova-de-acesso-ao-ensino-superior-do-mundo>. Acesso em: 13 jan. 2016.

COIMBRA, Carlos A. Q. **Teoria da Medição de Variáveis Latentes: História e Utilização Atual**. Brasília: MAST/MCT, 2006.

CURRAN, P. J, BOLLEN, K. A. The best of both worlds: Combining autoregressive and latent curve models. In: COLLINS, L. M; SAYER, A. G, (Ed.) **New methods for the analysis of change**. Washington, DC: American Psychological Association Press; 2001. p.105-136.

ESPÍRITO SANTO, Alexandre do. **Correlação Bisserial**. 1978. Disponível em: [www.uel.br/revistas/uel/index.php/seminario/article/viewFile/.../5627](http://www.uel.br/revistas/uel/index.php/seminario/article/viewFile/.../5627). Acesso em: 9 set. 2015.

GLOBO. G1. **Matriz de referência do ENEM**. 2014. Disponível em: [g1.globo.com/jornalhoje/download/0,,5840-1,00.doc](http://g1.globo.com/jornalhoje/download/0,,5840-1,00.doc). Acesso em: 10 nov. 2015.

HAMBLETON, R.K., JONES, R.W. & ROGERS, H.J. Influence of item parameter estimation errors in test development. **Journal of Educational Measurement**, v.30, n.2, p. 143-155. 1993.

INEP. **Matriz competência: Matemática Ensino Médio**. 2013. Disponível em: [http://download.inep.gov.br/educacao\\_basica/encejeja/matriz\\_competencia/Mat\\_Mat\\_Tec\\_EM.pdf](http://download.inep.gov.br/educacao_basica/encejeja/matriz_competencia/Mat_Mat_Tec_EM.pdf). Acesso em: 20 dez. 2015.

KLEIN, Ruben. **Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB)**. 2009. Disponível em: <http://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/38>. Acesso em: 13 dez. 2015.

MARTINS, Gilberto de Andrade. **Estatística Geral e Aplicada**. 3. ed. São Paulo: Ed. Atlas, 2006.

PASQUALI, Luiz. (Org.) **Teoria e métodos de medida em ciências do comportamento**. Brasília: INEP, 1996. 432p.

PASQUALI, Luiz. **Psicometria: teoria dos testes na psicologia e na educação**. 5.ed. Rio de Janeiro: Vozes, 2013.

PASQUALI, Luiz; PRIMI, Ricardo. Fundamentos da Teoria de Resposta ao Item – TRI. In: **Avaliação Psicológica**. v.2, n.2, p. 99-110. 2003.

RENZO, Oswald. **Algumas Considerações Sobre a Existência do Traço Latente na Construção de Provas Psicométricas**. São Paulo: Moytica, Eficiência Empresarial, 2008.

SANTOS, Marcos Antônio da Cunha; SOARES, José Francisco. Estimativas Não Paramétricas das Curvas Características dos Itens em Testes Educacionais. **Est. Aval. Educ.**, São Paulo, v. 23, n. 52, p. 130-144, maio/ago. 2012.

VIANNA, Heraldo Marelim. **Avaliações nacionais em larga escala**. São Paulo: DPE, 2003.

VIANNA, Heraldo Marelim. **Provas e testes no concurso vestibular**. Disponível em: <http://www.fcc.org.br/pesquisa/publicacoes/es/artigos/110.pdf>. Acesso em: 09 ago. 2014.