



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática
Programa de Mestrado Profissional em
Matemática em Rede Nacional



Uma Proposta de Análise de Desempenho dos Estudantes e de Valorização da Primeira Fase da OBMEP

Ana Paula Lima Vilarinho

Brasília-DF
2015

Ana Paula Lima Vilarinho

Uma Proposta de Análise de Desempenho dos Estudantes e de Valorização da Primeira Fase da OBMEP

Dissertação apresentada ao Departamento de Matemática da Universidade de Brasília, como parte dos requisitos do Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, para obtenção do grau de Mestre.

Orientador: Prof. Dr. Mauro Luiz Rabelo

Brasília-DF
2015

Todos os direitos reservados. É proibida a reprodução total ou parcial deste trabalho sem a autorização da universidade, do autor e do orientador.

Ana Paula Lima Vilarinho graduou-se em Matemática pela Universidade de Brasília.

Dedicatória

Dedico este trabalho ao meu falecido pai, pelo exemplo que me transmitiu, de caráter, de honestidade, de dedicação e de gosto pelo estudo.

Agradecimentos

Agradeço primeiramente a Deus, por estar comigo em todos os momentos da vida e por colocar no meu caminho pessoas tão queridas e especiais.

Aos meus pais, meus exemplos de vida, pelo amor incondicional.

À minha família, irmãos, tios e primos, por acreditarem em mim e serem incentivadores do meu crescimento pessoal e profissional.

Aos meus amigos que compreenderam os momentos de ausência dedicados aos estudos.

Aos meus colegas de curso, pelo companheirismo em sala de aula e nas longas horas de estudos. Em especial aos meus amigos Douglas e Maryna, pela cumplicidade de sempre. Ao Moacyr e ao Daniel por estarem sempre disponíveis quando precisei.

Agradeço à minha colega Regiane, por estar sempre junto, em todos os momentos na elaboração desse trabalho. Obrigada pelas conversas, paciência, correções, sugestões e por todo carinho.

Agradeço a todos que contribuíram de forma atuante nesse momento privilegiado de aprendizagem: Letícia, Felipe, Edson, Cacio, Terezinha, coordenadores da OBMEP e da escola participante.

A todos os professores que me incentivaram e contribuíram na minha formação docente. Ao meu orientador Professor Dr. Mauro Luiz Rabelo, pelo exemplo profissional, por compartilhar seus conhecimentos científicos, por toda ajuda, correções e orientações neste trabalho e por, desde o início, ter acreditado no meu potencial.

Aos meus alunos, por depositarem confiança no meu trabalho, por me ensinarem um pouquinho mais a cada dia e, com essa troca de saberes, me encorajarem todos os dias a seguir nesta amada profissão.

Por fim, agradeço às pessoas que, de perto ou de longe, acreditam e torcem para o meu sucesso. Muito obrigada!

Agradecimentos institucionais

CEBRASPE - Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos;

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior;

IMPA - Instituto Nacional de Matemática Pura e Aplicada;

UnB - Universidade de Brasília;

SEDF - Secretaria de Estado de Educação do DF.

"Ninguém ignora tudo. Ninguém sabe tudo. Todos nós sabemos alguma coisa. Todos nós ignoramos alguma coisa. Por isso aprendemos sempre."(Paulo Freire)

Resumo

A Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP) tem, cada vez mais, ganhado espaço no cenário nacional pelo alcance de sua aplicação e pela proposta de incentivar os estudantes brasileiros a desenvolver o gosto pela aprendizagem da Matemática na educação básica. Em 2014, mais de 18 milhões de estudantes foram inscritos para participar da primeira fase em todo o país. Esse cenário motivou o desenvolvimento deste trabalho, com o intuito de compreender melhor em que medida a prova da primeira fase dessa competição pode servir de instrumento para dar *feedback* aos estudantes e docentes, em termos da aprendizagem de Matemática, e aos elaboradores da prova, acerca da qualidade do instrumento de avaliação aplicado. O estudo foi feito considerando uma amostra composta de 534 estudantes de oitavos e nonos anos de uma escola pública de uma cidade satélite do Distrito Federal, que fizeram a Primeira Fase da OBMEP em 2014. Para o tratamento dos dados, contou-se com o apoio do Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos (CEBRASPE), utilizando-se os softwares *SPSS* e *BILOG-MG*. Para a análise psicométrica dos itens, lançou-se mão da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI). Adotou-se o modelo logístico de 3 parâmetros da TRI, que descreve a probabilidade de um estudante acertar determinado item do teste em função de sua proficiência e dos parâmetros de dificuldade, discriminação e acerto ao acaso do respectivo item. Os resultados mostraram que a prova apresentou alto grau de dificuldade para o grupo pesquisado e com quantidade elevada de itens com baixo poder discriminativo. Observou-se, também, significativa divergência nos resultados em relação aos estudantes selecionados para a Segunda Fase da OBMEP, caso o critério utilizado fosse embasado unicamente na TRI, e não na TCT em conjunto com outros critérios subjetivos definidos pelas escolas. Ao final da pesquisa, foram criadas propostas de modelos de boletim de desempenho para todos os participantes da OBMEP, atitude que pode contribuir para maior envolvimento e motivação dos alunos e das escolas com a Olimpíada, gerando benefícios no processo de aprendizagem da Matemática dos participantes.

Palavras-chaves: OBMEP. Aprendizagem de Matemática. Teoria de Resposta ao Item (TRI).

Abstract

The Brazilian Mathematical Olympiad of Public Schools (OBMEP) has been gaining, each day, more space in national scenario due to the extent of its application and its proposal of stimulating Brazilian students in order to develop taste in learning Mathematics in basic education. In 2014, more than 18 million students were enrolled for its first round all over the country. This scenario motivates the development of this research in order to better comprehend in what ways the first phase exam of this competition can be used as a feedback instrument to students and teachers, in terms of Mathematics learning, and to exam developers, in terms of quality of the applied instrument evaluation. The study was made based on a sample of 534 eighth and ninth grade students of a public school in a satellite city of the Federal District of Brazil that took the first phase of the exam (OBMEP) in 2014. For data treatment, this research counted with the support of the Brazilian Center of Research in Evaluation and Selection and Events Promotion (CEBRASPE), and the softwares *SPSS* and *BLOG-MG*. For the items psychometric analysis the Classical Test Theory (CTT) and the Item Response Theory (IRT) were also utilized. This research adopted the 3 parameters logistic model of IRT, which describes the probability of a student to succeed in a certain item according to its proficiency and the difficulty parameters, discrimination and random guess of the respective item. The results have shown that the exam presented a high level of difficulty for the researched group and with a high quantity of items with low discriminative power. It was also observed a meaningful divergence of the results regarding the selected students for the second phase of the OBMEP, if the chosen criteria were based solely on the IRT analysis and not on the CTT in conjunction with other subjective criteria defined by the schools. By the end of the research, performance report models were created for all the OBMEP participants, attitude that can contribute for better involvement and motivation of students and schools with the Olympiad, generating benefits in the Mathematics learning process of participants.

Keywords: OBMEP. Mathematics learning. Item Response Theory (IRT).

Sumário

1	Olimpíadas Brasileiras de Matemática	18
1.1	A Olimpíada Brasileira de Matemática (OBM)	18
1.2	A Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP) . . .	19
2	Análise Psicométrica de Itens	27
2.1	Teoria Clássica dos Testes (TCT)	27
2.2	Teoria de Resposta ao Item (TRI)	31
3	Material e Métodos	40
3.1	Coleta de Dados	40
3.2	Programas Computacionais Utilizados	41
4	Aplicação da TCT e da TRI na Primeira Fase da OBMEP de 2014	45
4.1	Visão Geral da Prova	45
4.2	Análise Psicométrica e Pedagógica dos Itens	49
4.3	Observações Gerais	81
5	Considerações Finais	82
5.1	O Uso da TRI na Seleção de Estudantes para a Segunda Fase da OBMEP	82
5.2	Sugestões de Aprimoramento do Processo	85

Lista de Tabelas

1	Estrutura e Pontuação da OBM.	19
2	Classificação das escolas em grupos de acordo com o número de alunos inscritos na Primeira Fase da OBMEP em cada nível.	21
3	Número de escolas e estudantes inscritos e percentual de municípios brasileiros participantes da Primeira Fase da OBMEP nos anos de 2005 a 2014.	24
4	Informações sobre a OBM e OBMEP.	26
5	Classificação e percentual esperado para os índices de dificuldade na TCT.	27
6	Classificação dos itens de acordo com o poder de discriminação na TCT.	28
7	Distribuição esperada e classificação do item em relação ao seu nível de dificuldade.	35
8	Classificação do item quanto ao seu poder de discriminação.	36
9	Classificação do item na escala (500, 100) em relação a sua dificuldade.	38
10	Classificação do item na escala (500, 100) quanto ao seu poder de discriminação.	38
11	Visão geral da prova da Primeira Fase da OBMEP de 2014.	45
12	Distribuição dos itens em relação ao parâmetro dificuldade, segundo a TCT.	46
13	Distribuição dos itens em relação a discriminação, pela TCT.	46
14	Relação dos estudantes classificados para Segunda Fase da OBMEP pelo método da TCT e que divergem dos estudantes classificados pela TRI.	82
15	Comparação de escores obtidos pelos estudantes selecionados na prova com todos os itens e na prova considerando apenas os 11 itens que possibilitaram o processamento dos dados, via TRI.	84
16	Médias de Proficiência de alunos do 9º ano em Matemática na Prova Brasil - 2013.	87

Lista de Figuras

1	Análise Gráfica do Item	29
2	Análise Gráfica do Item	30
3	Curva Característica do Item	34
4	CCI de um item com $a = 0$	36
5	CCI de um item com $a = 0,5$	36
6	CCI de um item com $a = 1$	37
7	CCI de um item com $a = 1,5$	37
8	Tela de entrada do software <i>BILOG-MG</i>	42
9	Resultados encontrados na fase inicial, no arquivo com extensão .PH1 . . .	43
10	Tela de entrada do software <i>BILOG-MG</i> após a retirada de alguns itens . .	43
11	Modelo de Boletim de desempenho pela TCT	47
12	Modelo de Boletim de desempenho pela TRI	48

Introdução

A avaliação de larga escala no Brasil foi implementada no final dos anos 80, porém, seu desenvolvimento e expansão ocorreu ao longo da década de 90, após a criação do Sistema de Avaliação da Educação Básica (SAEB). Os indicadores de desempenhos e as informações oferecidas pelas avaliações que fazem parte do SAEB retratam a realidade da qualidade da educação básica brasileira, fornecendo informações essenciais para subsidiar formulação e reformulação de políticas públicas, definindo prioridades de intervenção, auxiliando na alocação de recursos técnicos e financeiros que contribuem para redução das desigualdades e possíveis diferenças regionais existentes na educação básica do país.

O monitoramento da qualidade da educação brasileira tem revelado profundas deficiências no aprendizado de Matemática. Segundos dados publicados pelo portal *qedu.com.br*, os últimos resultados de um importante indicador da qualidade do ensino brasileiro, a Prova Brasil de 2013, revelaram que apenas aproximadamente 11% dos estudantes de 9º ano aprenderam o que se considera adequado em Matemática e, além disso, apontaram para uma diferença de desempenho muito significativa entre os estudantes dos diversos estados brasileiros. Isso sinaliza a existência de um problema grave de eficiência do ensino básico de Matemática oferecido pelas escolas brasileiras, de modo geral.

A expansão da utilização das avaliações de larga escala nas últimas décadas levaram a um aumento na preocupação com a metodologia utilizada para obtenção dos resultados. Aprofundou-se, também, a discussão a respeito da qualidade e da validade dos instrumentos utilizados, como a “definição” de princípios básicos para elaboração de itens que cumpram finalidade educacional avaliativa. Além disso, houve a ampliação e reestruturação de métodos e estatísticas utilizadas para obtenção das informações acerca do aprendizado, do desempenho dos estudantes e da interpretação pedagógica dos resultados.

No que tange à avaliação educacional de larga escala, é imprescindível a construção e utilização de uma matriz de referência que dê suporte para a elaboração de itens, definindo objetivos e finalidades de cada item do instrumento, e possibilitando a interpretação qualitativa dos resultados.

Em 1995, o SAEB foi reestruturado e passou a utilizar a Teoria de Respostas ao Item (TRI) para orientar a construção do banco de itens, a aplicação dos testes e a análise dos resultados. Essa modificação possibilitou a comparabilidade dos resultados ao longo do tempo e entre as séries avaliadas, facilitando o acompanhamento da evolução dos desempenhos dos estudantes do Brasil como um todo, de cada região, de cada estado e de cada município. Além das informações quantitativas, dados oriundos da TRI permitiram a interpretação qualitativa dos resultados, isto é, possibilitaram saber o que os estudantes realmente aprenderam em cada uma das áreas avaliadas, de acordo com seu nível de proficiência, permitindo, assim, melhor compreensão dos resultados pelos professores e gestores para definição de metas e medidas interventivas necessárias para criação de estratégias de ensino, que proporcionem o aprendizado pleno e, portanto, contribuam para o aumento da qualidade da educação.

Após essa breve contextualização acerca da experiência brasileira com o SAEB, compartilho a minha motivação para chegar ao objeto de estudo deste trabalho. Após cursar a disciplina “Avaliação Educacional” no Mestrado Profissional em Matemática - PROF-MAT, na qual houve melhor compreensão a respeito das concepções e finalidades das

avaliações educacionais de larga escala, dos fundamentos da Teoria Clássica dos Testes e da Teoria de Resposta ao Item e da metodologia de construção de itens, também chamado de “engenharia de construção de itens” (RABELO, 2013, p.177), constatou-se a necessidade de se refletir sobre as qualidades e as finalidades dos processos avaliativos no Brasil. Como tinha vivenciado a aplicação da primeira fase da Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP) na escola em que atuo, surgiu naturalmente a curiosidade de investigar em detalhes o instrumento, sua adequabilidade para o alcance dos objetivos da Olimpíada, o suporte teórico-metodológico de sua elaboração e o desempenho dos estudantes, a partir dos conhecimentos teóricos adquiridos ao longo da disciplina.

Desse modo, como ponto de partida, escolhi fazer uma análise da prova da Primeira Fase da OBMEP, bem como das respostas que os alunos de oitavo e nono ano do Ensino Fundamental de uma escola pública de uma determinada Região Administrativa do Distrito Federal apresentaram durante a realização do exame.

Mais especificamente, será realizado um estudo de caso envolvendo estudantes participantes da Olimpíada, que poderá sinalizar para os organizadores da OBMEP, algumas possibilidades de aperfeiçoamento da Primeira Fase no que diz respeito à elaboração do instrumento, e ao tratamento dos resultados. Além disso, pretende-se sugerir subsídios aos docentes e retorno a cada estudante participante em relação à sua aprendizagem e ao seu desempenho perante aos demais.

O trabalho foi norteado pelos seguintes questionamentos:

- A avaliação está coerente com os propósitos estabelecidos?
- Os itens estão formulados de acordo com as sugestões da “engenharia de itens”?
- O processo de seleção dos alunos para a Segunda Fase da OBMEP está realmente selecionando os alunos com melhor desempenho?
- Qual retorno é dado aos estudantes não selecionados para a Segunda Fase da OBMEP?
- Se a prova utilizasse fundamentos da TRI para a seleção dos estudantes para a Segunda Fase, os resultados sofreriam alterações?
- Qual o nível de dificuldade e o poder de discriminação de cada item?

Desta maneira, para responder a essas questões, foram definidos o objetivo geral e os objetivos específicos do presente estudo.

Objetivo geral: Analisar o desempenho dos estudantes participantes da Primeira Fase da OBMEP, utilizando os fundamentos da Teoria Clássica dos Testes e da Teoria de Resposta ao Item, a fim de compreender em que medida a prova pode servir de instrumento para dar *feedback* aos estudantes e docentes das escolas em termos da aprendizagem em Matemática e um retorno aos elaboradores a respeito da intencionalidade e qualidade do instrumento.

Objetivos específicos:

- Realizar o confronto da avaliação da OBMEP com a “*engenharia*” de construção de itens para avaliação de larga escala;
- Analisar as respostas dos estudantes usando fundamentos da Teoria Clássica dos Testes (TCT);
- Construir a Análise Gráfica dos Itens (AGI) aplicados;
- Analisar as respostas dos estudantes usando fundamentos da Teoria de Resposta ao Item (TRI);
- Construir a Curva Característica dos Itens (CCI) aplicados;
- Propor tratamento dos resultados a partir das respostas de todos os estudantes;
- Sugerir um método de seleção, mais formal e único, dos alunos que devem seguir para a Segunda Fase da OBMEP;
- Criar um instrumento de *feedback* para os alunos que participarem da Primeira Fase e para a escola;
- Propor um *feedback* para os organizadores da OBMEP no que diz respeito ao comportamento dos itens para os respondentes.

Para atingir os objetivos acima explicitados, dividiu-se este trabalho em duas partes. Na primeira, apresenta-se os referenciais teóricos que fundamentam esta investigação, a saber:

- **Olímpiada Brasileira de Matemática (OBM) e Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP):** apresenta-se um breve histórico das Olimpíadas de Matemática no Brasil, comentando a estrutura, os objetivos e o regulamento de cada umas delas;
- **Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI):** comenta-se a importância da análise de resultados de uma avaliação de larga escala, apresentando os fundamentos das duas teorias. Para cada uma delas, explicam-se os parâmetros dos itens observados no comportamento dos estudantes perante a avaliação, como dificuldade e discriminação do item, coeficiente ponto-bisserial, probabilidade de acerto ao acaso, nível de proficiência, AGI's e CCI's.

Na segunda parte, são apresentados os capítulos que qualificam este estudo:

- **Material e Métodos:** apresentam-se esclarecimentos acerca das escolhas metodológicas e do caminho percorrido nessa investigação;
- **Aplicação da TCT e da TRI na Primeira Fase da OBMEP de 2014:** estudo do instrumento avaliativo usado na OBMEP e do desempenho dos participantes do grupo pesquisado. São calculados os parâmetros e construídos os gráficos relativos a cada item, utilizando-se fundamentos da TCT e da TRI;

- **Considerações Finais:** apresentam-se considerações e conclusões sobre a investigação realizada.

1 Olimpíadas Brasileiras de Matemática

1.1 A Olimpíada Brasileira de Matemática (OBM)

A Olimpíada Brasileira de Matemática (OBM) é uma competição aberta entre alunos de escolas públicas e privadas de todo o Brasil. A OBM iniciou-se em 1979, idealizada e organizada pela Sociedade Brasileira de Matemática (SBM) e o Instituto de Matemática Pura e Aplicada (IMPA), e conta com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), do Instituto Nacional de Ciências e Tecnologia de Matemática (INCTMat) e da Academia Brasileira de Ciências.

Alguns dos principais objetivos da OBM são interferir na melhoria do ensino de Matemática no Brasil, descobrir jovens talentos e selecionar e preparar estudantes para participar de disputas internacionais. O processo de seleção de estudantes para a formação de equipes internacionais leva em consideração a colocação conquistada na OBM no ano anterior e resultados obtidos em provas seletivas e em listas de exercícios oferecidas durante o treinamento nos meses que antecedem as provas.

Inicialmente, a OBM era dividida em dois níveis, Júnior e Sênior, que correspondiam a alunos de até 15 anos de idade e de ensino médio, respectivamente. Em 1998, o formato da prova foi modificado e foi incluído mais um nível. De acordo com o novo modelo, o Nível I correspondia a alunos de 5^a e 6^a séries, o Nível II correspondia a alunos de 7^a e 8^a séries e o Nível III, a alunos de ensino médio. Em 2001, mais uma alteração foi realizada, incluindo-se o Nível Universitário.

Nos Níveis I, II e III, a estrutura da competição é compreendida por três fases. A Primeira Fase é uma prova de múltipla escolha com 20 a 25 questões que pode ser realizada em até 3 horas. A Segunda Fase é uma prova mista com 4 horas e 30 minutos de duração. A Terceira Fase é diferenciada de acordo com o nível: no Nível I, é aplicada uma prova discursiva, com cinco problemas a serem resolvidos em até 4 horas e 30 minutos; já nos Níveis II e III, são aplicadas duas provas discursivas realizadas em dois dias consecutivos, cada uma é composta por três problemas a serem resolvidos em um período de 4 horas e 30 minutos.

A Olimpíada Brasileira de Matemática de Nível Universitário é compreendida por duas fases, ambas com provas discursivas compostas por 6 problemas que devem ser resolvidos em, no máximo, 4 horas e 30 minutos, cada uma.

Nos Níveis I, II e III, cada questão da Primeira Fase vale 1 ponto; a Segunda Fase totaliza 60 pontos e na Terceira Fase, cada problema vale 50 pontos. No Nível Universitário, cada questão da Primeira e da Segunda Fase vale 10 e 50 pontos, respectivamente, conforme apresentado na **Tabela 1** (Fonte: Portal OBMEP).

Tabela 1: Estrutura e Pontuação da OBM.

	Primeira Fase		Segunda Fase		Terceira Fase	
	Estrutura	Pontuação	Estrutura	Pontuação	Estrutura	Pontuação
Nível I	20 a 25 itens de múltipla escolha	1 ponto cada item	Prova mista	60 pontos no total	5 problemas discursivos	50 pontos cada problema
Nível II	20 a 25 questões	1 ponto cada item	Prova mista	60 pontos no total	Duas provas discursivas compostas por 3 problemas em cada	50 pontos cada problema
Nível III	20 a 25 questões	1 ponto cada item	Prova mista	60 pontos no total	Duas provas discursivas compostas por 3 problemas em cada	50 pontos cada problema
Nível Universitário	6 problemas discursivos	10 pontos	6 problemas discursivos	50 pontos	-	-

Os alunos mais bem pontuados ao final de todas as etapas são premiados com medalhas de ouro, prata e bronze e menção honrosa.

A Comissão Nacional de Olimpíada da SBM é responsável por definir e divulgar as datas das OBM, solicitar colaborações e instruir as escolas e estabelecimentos de Ensino Superior para tornar exequível a realização da OBM em âmbito nacional, nomear bancas especializadas em elaboração e correção das provas, divulgar critérios a serem utilizados para a promoção dos alunos para Segunda Fase e patrocinar despesas dos alunos premiados.

Com o apoio do CNPq, os organizadores da OBM puderam ampliar a Comissão Nacional de Olimpíadas da SBM, composta atualmente por 21 membros, incluindo um coordenador geral e um coordenador acadêmico, além de uma Comissão Júnior. Também criaram a revista EUREKA!, que ajuda na preparação de alunos para a OBM em seus diversos níveis, com materiais e informações direcionadas a essa disputa e também às Olimpíadas Internacionais. Além disso, disponibilizaram uma página eletrônica como meio de constante comunicação entre alunos, coordenadores e diversas partes envolvidas. Nessa página, podem ser encontrados artigos, banco de questões, provas, gabaritos antigos e informações importantes para os alunos interessados em realizar a prova.

A Comissão de Olimpíadas da SBM, assim como toda a organização e logística da OBM ficam centralizadas no IMPA, no qual existe uma secretaria responsável por concentrar os trabalhos de divulgação e coordenação das atividades, além de dar o apoio necessário aos envolvidos nesse projeto.

1.2 A Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP)

A Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP) iniciou-se em 2005, apoiada financeiramente pelo Ministério da Educação (MEC) e Ministério de

Ciência, Tecnologia e Inovação (MCTI), sendo realizada pelo IMPA e pela SBM.

Segundo César Camacho, diretor geral do IMPA, “a OBMEP foi criada destinada a detectar talentos em todos os cantos do país e contribuir para a melhoria do ensino da Matemática” (OBMEP, 2014).

Segundo a consultora do Centro de Gestão e Estudos Estratégicos (CGEE), Tatiana de P. A. Maranhão (2011, p.35),

a principal razão para a existência da OBMEP são os alunos das escolas públicas, seus desempenhos, interesse e motivação pela Matemática. Este grupo de atores individuais é o foco principal dessa política porque está no cerne de problemas existentes e inter-relacionados: o baixo desempenho dos alunos em Matemática, a importância da Matemática para o desenvolvimento tecnológico do país, a baixa adesão dos profissionais a esta carreira, a necessidade de profissionais para a formação de novos alunos.

No entanto, de acordo com informações do regulamento da OBMEP, disponível no site *www.obmep.org.br*, essa avaliação exerce outros papéis importantes, entre os quais se destacam:

- Estimular o estudo de Matemática por meio de resolução de problemas que despertem o interesse e a curiosidade de estudantes e professores;
- Incentivar o aperfeiçoamento dos professores das escolas públicas contribuindo assim para a valorização dos docentes;
- Identificar jovens talentos e incentivar seu ingresso nas áreas científicas e tecnológicas;
- Promover a inclusão social por meio da difusão de conhecimento;
- Contribuir para a integração das escolas públicas com as universidades públicas, os institutos de pesquisa e as sociedades científicas.

Para Cláudio Landin (2014), coordenador geral da OBMEP e professor do IMPA,

os testes são concebidos de forma a detectar alunos que tenham potencial para Matemática. Queremos detectar alunos talentosos. As perguntas formuladas envolvem raciocínio, abstração e criatividade, não é necessário qualquer conhecimento formal sobre Matemática.

Qualquer aluno matriculado em escolas públicas municipais, estaduais ou federais do Brasil pode participar da OBMEP. O aluno não deve se preocupar com o momento da inscrição, pois não existe inscrição individual. Assim, para se inscrever, cada escola deve acessar a página eletrônica do programa e informar quantos alunos realizarão a prova. Não

há limite para a quantidade de alunos, sendo papel da escola estimulá-los a participar do programa.

Os estudantes que participam da OBMEP são divididos em três níveis, de acordo com o seu grau de escolaridade:

- Nível I - corresponde a estudantes de quinto e sexto ano do Ensino Fundamental;
- Nível II - corresponde a estudantes do sétimo e oitavo ano do Ensino Fundamental;
- Nível III - corresponde a estudantes do primeiro, segundo e terceiro ano do Ensino Médio.

Alunos matriculados na Educação de Jovens e Adultos (EJA) também podem realizar a prova, e são enquadrados nos mesmos níveis citados acima, de acordo com a série em que estão cursando.

A OBMEP é realizada em duas fases. A primeira corresponde a uma prova composta por 20 questões de múltipla escolha, diferenciada entre os níveis. Nessa fase da Olimpíada, a prova, com duração de duas horas e trinta minutos, é aplicada na própria escola inscrita, pelos próprios professores da escola, em um dia específico marcado no calendário oficial da OBMEP. A correção da prova também é realizada pelos professores da escola, de acordo com o gabarito enviado pela comissão organizadora.

No Distrito Federal, o dia da realização da Primeira Fase da OBMEP já é destacado e previsto no Calendário Escolar. Neste dia, as atividades das escolas são exclusivamente direcionadas a essa avaliação.

Cada escola, no momento da inscrição, deve designar um responsável pela OBMEP na sua escola. Esse responsável é quem recebe as provas da Primeira Fase e as orientações sobre sua aplicação e correção.

Para critério de classificação e premiação, as escolas participantes são divididas de acordo com o número de alunos inscritos na Primeira Fase da OBMEP em cada nível, conforme a **Tabela 2** (Fonte: Portal OBMEP).

Tabela 2: Classificação das escolas em grupos de acordo com o número de alunos inscritos na Primeira Fase da OBMEP em cada nível.

	Nível I	Nível II	Nível III
Grupo A	1 a 40	1 a 40	1 a 120
Grupo B	41 a 80	41 a 80	121 a 240
Grupo C	81 a 140	81 a 140	241 a 380
Grupo D	141 a 240	141 a 240	381 a 620
Grupo E	241 ou mais alunos inscritos	241 ou mais alunos inscritos	621 ou mais alunos inscritos

Após a correção das provas da Primeira Fase, o responsável pela OBMEP de cada escola deve selecionar os alunos que obtiveram as maiores notas para realizar a Segunda Fase da Olimpíada.

A quantidade de alunos selecionados para a Segunda Fase depende do grupo no qual a escola está inserida, referentes aos Níveis I e II. O regulamento da prova de 2014 contém a seguinte distribuição:

- Escolas do grupo A que inscreverem apenas um aluno na Primeira Fase, deverão selecionar um aluno para a Segunda Fase. Escolas que inscreverem de dois a quarenta alunos, selecionarão dois alunos para a Segunda Fase;
- Escolas do grupo B deverão selecionar quatro alunos para a Segunda Fase;
- Escolas do grupo C selecionarão sete alunos para a Segunda Fase;
- Escolas do grupo D selecionarão doze alunos para a Segunda Fase;
- Escolas do grupo E selecionarão cinco por cento (5%) do total de alunos inscritos na Primeira Fase.

A quantidade de alunos classificados para Segunda Fase da OBMEP no Nível III também varia de acordo com o grupo da escola:

- Escolas do grupo A que inscreverem até seis alunos na Primeira Fase, deverão selecionar todos os inscritos para a Segunda Fase. Se a escola inscreveu entre sete e cento e vinte alunos, deverão ser selecionados seis alunos para realizar a Segunda Fase;
- Escolas do grupo B deverão selecionar doze alunos para a Segunda Fase;
- Escolas do grupo C selecionarão dezenove alunos para a Segunda Fase;
- Escolas do grupo D selecionarão trinta e um alunos para a Segunda Fase;
- Escolas do grupo E, selecionarão cinco por cento do total de inscritos na Primeira Fase para realizar a Segunda Fase da OBMEP.

Caso haja empate entre os alunos com melhores pontuações, o critério de desempate deverá ser definido e divulgado pela própria escola. Além disso, alunos com nota zero não poderão ser selecionados para realizar a Segunda Fase. Escolas do grupo E deverão arredondar para cima a quantidade de alunos selecionados para a Segunda Fase, caso o número equivalente a cinco por cento não seja inteiro.

A Segunda Fase da OBMEP é uma prova discursiva composta de 6 a 8 questões, com duração total de três horas. As provas são diferenciadas de acordo com o nível e são aplicadas em locais determinados pela comissão organizadora, designados centros de aplicação.

A Coordenação Geral da OBMEP responsabiliza-se por selecionar fiscais para a aplicação da Segunda Fase, além de garantir a correção e indicação de todas as premiações, mantendo-se atualizadas todas as informações no sítio da OBMEP.

Com base exclusivamente no resultado da Segunda Fase, alunos, professores, escolas e secretarias municipais podem ser contemplados com diversos prêmios. A título de ilustração, a premiação da OBMEP em 2014 foi realizada da seguinte maneira:

- 500 medalhas de ouro para os alunos que obtiveram as maiores notas na prova da Segunda Fase, sendo 200 para o nível 1, 200 para o nível 2 e 100 para o nível 3;
- 1500 medalhas de prata para os alunos com maiores notas na prova da Segunda Fase que não foram contemplados com medalha de ouro;
- 4500 medalhas de bronze;
- 30000 Certificados de Menção Honrosa;
- Programa de Iniciação Científica Jr. (PIC) aos 6500 medalhistas;
- Programa de Iniciação Científica e de Mestrado aos medalhistas de qualquer edição da OBMEP ou OBM, com bolsa oferecido pela parceria com o CNPq e com a Capes;
- Um tablet, um diploma e uma assinatura anual da Revista do Professor de Matemática (RPM-SBM) para os 159 professores mais bem pontuados, de acordo com a pontuação de seus alunos;
- Um diploma e uma assinatura anual da Revista do Professor de Matemática (RPM-SBM) para os 870 professores mais bem pontuados que não foram premiados no item anterior;
- Kits esportivos para 75 escolas mais bem pontuadas de acordo com a pontuação de seus alunos;
- Kits de material didático para 405 escolas com maior número de pontos excluídas as escolas premiadas no item anterior;
- Troféus às 15 escolas com maior número de pontos;
- Troféus às duas secretarias municipais com maior pontuação em cada Unidade da Federação.

Para efeito de premiação, a organização da OBMEP designa como escolas seletivas aquelas que realizam processo de seleção por meio de provas ou concursos ou priorizam o acesso de algum grupo de aluno como por exemplo, filhos de militares ou filhos de funcionários públicos.

A quantidade de premiação a alunos de escolas seletivas é estabelecida e divulgada no regulamento de cada prova. Por exemplo, em 2014, no máximo 130 medalhas de ouro, 450 medalhas de prata e 650 medalhas de bronze foram destinados a esses alunos.

Com relação aos eventos de cerimônia, a Coordenação Geral da OBMEP organiza uma cerimônia de premiação nacional para os medalhistas de ouro. As medalhas de prata e bronze são entregues em cerimônias regionais enquanto os certificados de menção honrosa são entregues na própria escola.

A premiação com distribuição de materiais didáticos, bolsas de iniciação científica e reconhecimento aos educadores, escolas e secretarias de educação é um grande incentivo à participação. Segundo Cláudio Landin (2014) “[...] A Olimpíada cria um ambiente estimulante para que os alunos anualmente se esforcem para participar.”

Nesses dez anos de OBMEP, nota-se um crescimento vertiginoso na quantidade de escolas e estudantes inscritos, o que retrata reconhecimento e confirmação de que o programa tem atingido seus objetivos. De acordo com César Camacho, diretor geral do IMPA, o extraordinário nível de participação escolar indica que a OBMEP tem tido uma benéfica influência nas escolas públicas e estimulado de alguma maneira o estudo da Matemática. Essa expansão de inscritos na Primeira Fase da OBMEP no decorrer dos anos e o alcance da OBMEP nos municípios brasileiros podem ser constatados na **Tabela 3** (Fonte: Portal OBMEP).

Tabela 3: Número de escolas e estudantes inscritos e percentual de municípios brasileiros participantes da Primeira Fase da OBMEP nos anos de 2005 a 2014.

	Escolas	Estudantes	Municípios
2005	31.031	10.520.831	93,5%
2006	32.655	14.181.705	94,5%
2007	38.450	17.341.732	98,1%
2008	40.397	18.326.029	98,7%
2009	43.854	19.198.710	99,1%
2010	44.717	19.665.928	99,16%
2011	44.691	18.720.068	98,9%
2012	46.728	19.166.371	99,42%
2013	47.144	18.762.859	99,35%
2014	46.711	18.192.526	99,41%

É interessante ressaltar que a OBMEP tem conseguido atingir municípios pequenos no interior do Brasil e que alguns deles têm apresentado bons desempenhos no programa. Dores do Turvo (MG) e Cocal dos Alves (PI) foram os municípios que lideraram o *ranking* de melhores municípios nos nove primeiros anos de OBMEP.

Para ranquear os municípios pelo índice de desempenho, foi acumulado o total de pontos durante os nove primeiros anos de OBMEP e dividido pelo total de alunos selecionados para a Segunda Fase da competição. Desse modo, a cada cidade foi atribuída uma nota. De acordo com Landim (2014), os excelentes resultados quase sempre estão atrelados ao trabalho de um professor.

A diferença é que um professor no interior do país tem um *status* social que não existe nas grandes cidades. Um professor com alunos medalhistas de olimpíadas é reconhecido. Todos os habitantes dali foram um dia alunos daquele professor. Há um reconhecimento, o que torna a carreira mais atrativa.

Apesar dos bons resultados apresentados pelo *ranking* dos municípios divulgado pela imprensa, é interessante ressaltar que a medição e a atribuição de notas aos municípios não podem ser vistas como fundamentais no processo avaliativo e nos indicadores de qualidade da educação. De acordo com LIBANEO *et al* (2011, p.205), existe uma diferença significativa entre medir e avaliar.

Avaliar, como parte de um projeto educativo, é diferente de medir. Medida refere-se à aferição, expresso em notas, conceitos ou menções, do aprendizado pretendido. [...] Avaliar, por sua vez, reflete determinada concepção de educação, do papel do professor e do que é conhecimento. Só pela diferença entre medir e avaliar podem-se perceber as variadas concepções do que seja avaliar.

Para incentivar a formação continuada dos professores e a preparação de alunos para a OBMEP, o IMPA e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) estão lançando o programa “OBMEP na Escola”, que visa estimular o uso dos materiais das Olimpíadas, tais como provas e banco de questões, em atividades extraclasse a serem realizadas em escolas públicas de todo o país.

Os preparadores do programa serão professores selecionados por meio de uma prova de habilitação e que tenham realizado um projeto analisado e aprovado pelo Comitê Nacional designado pela coordenação da OBMEP. Esses professores ainda receberão um treinamento por meio de oficinas mensais para prepará-los para o desenvolvimento do projeto.

Espera-se que esta iniciativa do IMPA ofereça subsídios e contribua ainda mais com o avanço da qualidade do ensino de Matemática nas escolas públicas em todo território nacional. O regulamento desse e de outros programas podem ser encontrados na página eletrônica da OBMEP.

Cláudio Landin (2014) avaliou positivamente o impacto desses dez anos de OBMEP na Educação Brasileira, porém acredita que ainda tenha muito a melhorar.

O Brasil já avançou muito com a OBMEP no ensino da Matemática, mas ainda é preciso que maiores esforços sejam feitos por todos os agentes da educação para que o Brasil atinja o nível de excelência que outros países apresentam. Isso não apenas na Matemática.

A **Tabela 4** a seguir apresenta, de maneira simplificada, as principais informações e diferenças entre as duas principais Olimpíadas de Matemática realizadas no país, citadas neste capítulo.

Tabela 4: Informações sobre a OBM e OBMEP.

	Olimpíada Brasileira de Matemática (OBM)	Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP)
O que é	Competição aberta a todos os estudantes, a fim de estimular o ensino de Matemática no País.	Competição que promove o estudo de Matemática entre os alunos das escolas públicas brasileiras.
Quem pode participar?	Alunos do 6º ao 9º ano do Ensino Fundamental, alunos de Ensino Médio e Universitário das escolas públicas e privadas.	Alunos do 6º ao 9º ano do Ensino Fundamental e do Ensino Médio de escolas públicas.
Como participar?	Inscrições devem ser feitas pelas escolas, a partir de formulário disponível no site da competição.	Inscrições devem ser feitas pelas escolas, a partir de fichas de inscrições disponíveis no site da competição.
Níveis	Nível I: 6º e 7º anos; Nível II: 8º e 9º anos; Nível III: Ensino Médio; Nível Universitário.	Nível I: 6º e 7º anos; Nível II: 8º e 9º anos; Nível III: Ensino Médio.
Fases	Três fases de provas para os Níveis I, II e III e duas fases para o Nível Universitário.	1ª fase: Prova objetiva para todos os alunos; 2ª fase: Prova discursiva para alunos selecionados.
Tipos de questões	1ª fase: 20 a 25 itens de múltipla escolha realizadas em 3 horas; 2ª fase: Prova mista realizada em 4 horas e 30 minutos; 3ª fase: Nível I - uma prova discursiva; Nível II e III - duas provas discursivas; Nível Universitário - duas fases com provas discursivas.	1ª fase: Prova de múltipla escolha com 20 itens; 2ª fase: Prova discursiva com 6 a 8 questões em cada nível.
Pontuação Níveis	Nível I, II e III - 1 ponto para cada item da 1ª fase, 60 pontos na 2ª fase e 50 pontos por questão na 3ª fase. No Nível Universitário, cada problema da 1ª fase vale 10 pontos e cada problema da 2ª fase vale 50 pontos.	Na Primeira Fase são selecionados uma quantidade determinada de alunos com os maiores escores. Cada escola tem autonomia de definir e divulgar o critério de desempate de alunos selecionados para a Segunda Fase.
Classificação	Estudantes seguem para Segunda Fase se atingirem pontuação mínima em cada um dos níveis.	Estudantes com as maiores notas na Primeira Fase seguem para próxima fase da OBMEP.
Prêmios	Medalhas e certificados.	Medalhas e certificados.
Site	www.obm.org.br	www.obmep.org.br

Fonte: www.brasil.gov.br

2 Análise Psicométrica de Itens

2.1 Teoria Clássica dos Testes (TCT)

Segundo Pasquali (2009), a Psicometria representa a teoria e a técnica de medidas de processos mentais e procura, de modo geral, explicar o sentido que tem as respostas dadas pelo sujeito a uma série de tarefas ou itens. No século XX, muitos estudos se embasaram principalmente na Teoria Clássica dos Testes (TCT), entretanto nas últimas décadas outra vertente da Psicometria moderna tem-se destacado: a Teoria de Resposta ao Item - TRI.

Em avaliação educacional, a TCT tem como elemento central a prova como um todo e seus resultados são expressos em escore bruto, ou seja, no número total ou no percentual de itens respondidos corretamente. As propriedades psicométricas dos itens de uma prova relacionam-se aos parâmetros a seguir: índice de dificuldade, índice de discriminação e correlação bisserial (BORGATTO & ANDRADE, 2012).

Para definir o índice de dificuldade do item, calcula-se a proporção de acertos, isto é, a razão entre o número de estudantes que responderam o item corretamente e o número total de estudantes submetidos ao item. O índice varia de 0 a 1, em que o extremo inferior significa que ninguém acertou e o extremo superior que todos acertaram. Logo, quanto menor a porcentagem de acerto, maior o grau de dificuldade do item.

Para que uma avaliação educacional tenha um nível ideal de dificuldade, Pasquali (2003) recomenda uma distribuição de níveis de dificuldades de itens no teste dentro de uma curva normal, conforme a **Tabela 5** a seguir.

Tabela 5: Classificação e percentual esperado para os índices de dificuldade na TCT.

Quantitativo ideal de itens na avaliação (% esperado)	Índice de dificuldade do item	Classificação do item em relação ao índice de dificuldade
10%	Superior a 0,9	Muito fáceis
20%	De 0,7 a 0,9	Fáceis
40%	De 0,3 a 0,7	Medianos
20%	De 0,1 a 0,3	Díficeis
10%	Até 0,1	Muito díficeis

O índice de discriminação mede a capacidade do item de diferenciar os participantes com maior habilidade daqueles com menor habilidade. Para o cálculo do índice de discriminação, os participantes são divididos em três grupos: o grupo superior (27% dos participantes com maiores pontuações), o grupo inferior (27% dos participantes com menores pontuações) e os demais 46% dos participantes, que compõem o grupo intermediário. Esse parâmetro corresponde à diferença entre o percentual de acerto do primeiro grupo e do segundo grupo. Quanto maior foi essa diferença, maior será a discriminação do item. Espera-se que em uma avaliação educacional, que o poder de discriminação do item seja superior a 40, conforme mostra a **Tabela 6** a seguir.

Tabela 6: Classificação dos itens de acordo com o poder de discriminação na TCT.

Valores	Classificação
Discriminação $< 0,20$	Item deficiente, deve ser rejeitado
$0,20 \leq$ Discriminação $< 0,30$	Item marginal, sujeito a reelaboração
$0,30 \leq$ Discriminação $< 0,40$	Item bom, mas sujeito a aprimoramento
Discriminação $\geq 0,40$	Item bom

Fonte: *Rabelo, 2013*

Segundo Rabelo (2013), a correlação ponto-biserial é uma associação entre o desempenho do indivíduo no item e no teste como todo, ou seja, com seu escore bruto. Esta medida varia no intervalo $(-1, 1)$ e valores negativos e próximos a zero indicam que indivíduos com maiores notas no teste estão errando o item, ou seja, o item tem baixa discriminação em relação ao resultado do teste.

Para o cálculo da correlação ponto-biserial, utiliza-se a seguinte fórmula:

$$\rho_{pb} = \frac{\bar{S}_p - \bar{S}}{\bar{\sigma}_s} \sqrt{\frac{p}{q}},$$

na qual \bar{S}_p é a nota média no teste para os indivíduos que responderam corretamente o item, \bar{S} é o escore médio do teste para todos os respondentes, $\bar{\sigma}_s$ é o desvio padrão dos escores obtidos no teste, p é o índice de dificuldade do item e $q = (1 - p)$, ou seja, a proporção de indivíduos que erraram o item.

A correlação ponto-biserial é designada correlação biserial quando uma das variáveis assume apenas dois valores e, para efeito de cálculo, supõe-se que o desempenho no item segue uma distribuição normal. A fórmula deste coeficiente é dada por

$$r_{bis} = \frac{\rho_{pb} \sqrt{p(1-p)}}{h(p)},$$

na qual $h(p)$ é a densidade da distribuição normal no ponto em que a área da curva a esquerda desse ponto é igual ao índice de dificuldade p .

Espera-se que a opção correta do item tenha correlação positiva, ou seja, $\bar{S}_p > \bar{S}$, e que as opções erradas, chamadas de distratores, tenham correlação negativa. Assim, poderemos dizer que alunos com melhores desempenhos no teste como um todo estão acertando o item, ou seja, o item está cumprindo um de seus objetivos fundamentais.

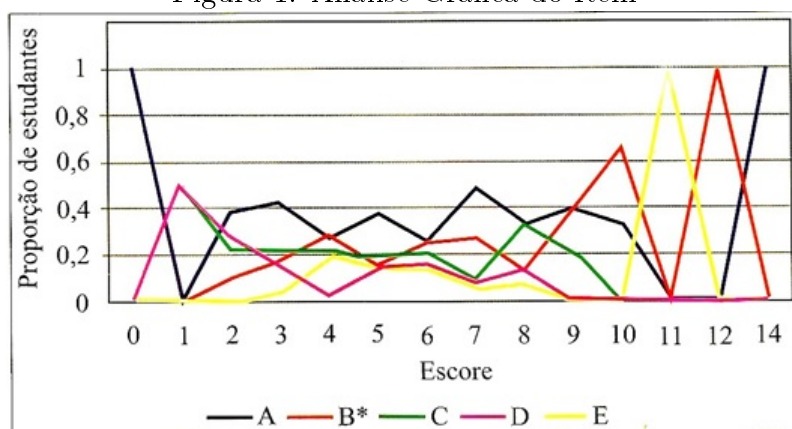
Na Teoria Clássica dos Testes, o modelo matemático que apresenta o alcance do item é denominado Análise Gráfica do Item (AGI). A AGI corresponde a um gráfico de linhas, no qual, no eixo das abscissas, ficam ordenados os escores brutos total no teste e, no eixo das ordenadas, a proporção de alunos que marcaram corretamente cada uma das alternativas do item, em cada uma das faixas de notas.

Espera-se na AGI que a linha correspondente ao gabarito do item cresça e que as linhas correspondentes aos distratores decresçam na medida em que o escore aumenta, isto é, respondentes com maior desempenho global tendem a acertar o item e indivíduos com menores desempenhos tendem a errar.

Com auxílio da AGI, é possível analisar pedagogicamente o item podendo-se detectar alguma falha na sua elaboração. Esse modelo pode revelar se os indivíduos estão sendo atraídos para alguma resposta incorreta, se existe algum distrator “meio de caminho” ou “peguinha”, além de outras conclusões acerca das habilidades e competências exigidas que possivelmente não foram alcançadas pelos respondentes. No caso de uma avaliação educacional, a AGI contribui também para que o professor entenda melhor o processo de aprendizagem do aluno.

A figura a seguir mostra um gráfico da AGI de um item de múltipla escolha de cinco opções de respostas, no qual a alternativa (B), representada pela linha vermelha, é o gabarito oficial.

Figura 1: Análise Gráfica do Item

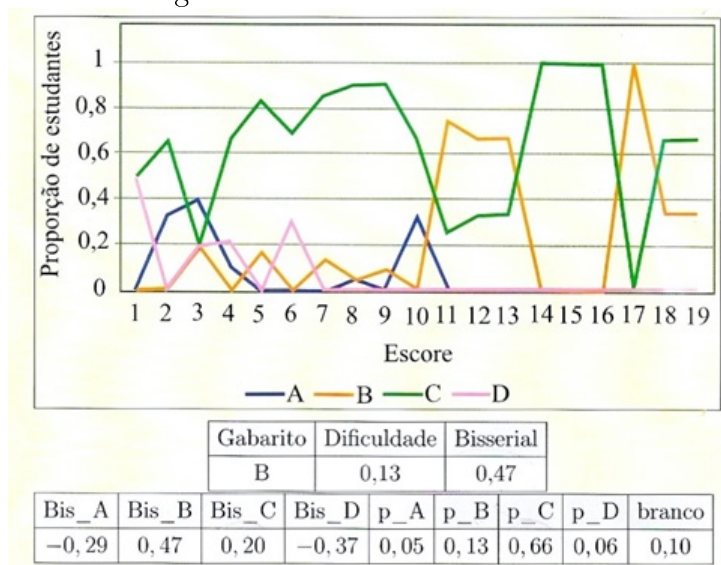


Na **Figura 1** (RABELO, 2013, p.149), podemos perceber que o item não correspondeu às expectativas de um item ideal, apresentando alguma falha em sua construção. Espera-se que indivíduos com maiores notas acertem a questão, o que, de fato não aconteceu. Ninguém com escore 14 marcou a opção correta. Nota-se também que a opção incorreta (A) foi escolhida por todos os respondentes de melhor desempenho global. É interessante observar também que todos os alunos com escore 11 optaram pela alternativa (E), portanto, sugere-se atenção a esta alternativa, pois pode conter algum “peguinha” que atraiu alunos com aptidão elevada.

Somente com a análise da AGI não é possível classificar o item em relação ao seu nível de dificuldade e discriminação, pois precisaríamos saber quantos alunos existem em cada faixa de nota para saber o total de alunos que acertaram o item e o total de alunos que realizaram o teste.

Para se realizar uma análise crítica mais detalhada de um item usando a TCT, são necessárias mais informações além da AGI, como o nível de dificuldade do item, a discriminação, a proporção de respostas em cada opção e o coeficiente bisserial geral e por alternativa. Na **Figura 2** (RABELO, 2013, p.153) a seguir, essas informações são evidenciadas nas tabelas, o que permite, assim, compreender melhor o resultado do desempenho dos estudantes.

Figura 2: Análise Gráfica do Item



A **Figura 2** mostra o gráfico de uma AGI de um item de múltipla escolha com 4 opções. Pelas informações da tabela, conclui-se que é um item difícil pois apenas 13% dos estudantes o responderam corretamente. O coeficiente bisserial 0,47 mostra que o item tem um bom poder de discriminação, ou seja, o percentual de estudantes com bom desempenho que acertaram o item foi maior que o percentual de estudantes com baixo desempenho. Isso pode ser visto na AGI, pois o traço laranja correspondente ao gabarito (B) cresce à medida que o escore aumenta.

Apesar do gabarito ser a opção (B), a grande maioria dos estudantes (66%) marcaram a opção (C), entre os alunos que optaram por essa resposta estão estudantes com escores 18 e 19. Percebe-se também que coeficiente bisserial desta opção incorreta é positivo (0,20), o que indica que esta opção atraiu muitos estudantes de bom desempenho.

Assim como o esperado, os coeficientes bisseriais dos distratores (A) e (D) foram negativos, iguais -0,29 e -0,37, respectivamente, ou seja, alunos de baixo desempenho no teste escolhem essa resposta mais do que alunos de alto desempenho. Em relação à proporção de marcação, pode-se questionar a plausibilidade dessas opções, pois apenas 5% e 6% os estudantes escolheram a alternativa (A) e (D), respectivamente.

A análise mostra que o item deve ser revisto, excluído ou reelaborado para que atinja sua finalidade de avaliar um determinado conhecimento do estudante.

Segundo Urbina (2007), apesar de ainda ser amplamente usado, os Princípios da Teoria Clássica do Teste têm sido continuamente refinados por possuírem algumas limitações, o que tem motivado o interesse de especialistas em avaliação por outras vertentes da Psicometria, como a Teoria de Resposta ao Item - TRI.

Andrade, Tavares e Valle (2000) citam como uma das principais limitações da TCT o fato de ser inviável a comparação entre indivíduos que não foram submetidos a uma mesma prova. Para Rabelo (2013, p.32), as conclusões feitas por meio da TCT se restringem a determinado momento, a determinado teste e ao grupo que respondeu à prova, o que inviabiliza e dificulta a comparação entre séries diferentes e o acompanhamento pedagógico ao longo dos anos. Além disso, com a TCT, a dificuldade do item varia de acordo com

os respondentes, ou seja, um item torna-se mais fácil ou mais difícil se for respondido por uma amostra composta por indivíduos mais ou menos inteligentes.

2.2 Teoria de Resposta ao Item (TRI)

Desde antes dos anos 30, estudiosos já haviam detectado algumas das limitações já citadas da Teoria Clássica. Uma das inquietações era o fato de os resultados obtidos pela TCT dependerem do conjunto de itens de determinado teste. Segundo Thurstone (1928/1959, p.547), *apud* Pasquali e Primi (2003):

Um instrumento de medida, na sua função de medir, não pode ser seriamente afetado pelo objeto de medida. Na extensão em que sua função de medir for assim afetada, a validade do instrumento é prejudicada ou limitada. Se um metro mede diferentemente pelo fato de estar medindo um tapete, uma pintura ou um pedaço de papel, então nesta mesma extensão a confiança neste metro como instrumento de medida é prejudicada. Dentro dos limites de objetos para os quais o instrumento de medida foi produzido, sua função deve ser independente da medida do objeto.

Pasquali (2003) cita trabalhos de Richardson (1936), Lawley (1943, 1944), Tucker (1946) e Lazarsfeld (1950) como os precursores da Teoria de Resposta ao Item (TRI), porém destaca Frederic Lord (1952, 1953) como responsável direto da introdução dessa teoria. Ele criou um novo modelo teórico e estabeleceu métodos para estimar os parâmetros dos itens dentro dessa nova teoria.

Apesar de estudos antigos, a expansão da TRI só ocorreu após os anos 80, devido a limitações computacionais existentes na época. Nesta década, com o avanço tecnológico computacional e o desenvolvimento de softwares apropriados como o *BICAL*, o *LOGISTIC*, o *BILOG* e o *BILOG-MG*, o problema entre a dependência do instrumento e o objeto de medida pôde ser solucionado, e a TRI passou a ser difundida e predominante na área de testes.

No Brasil, o uso da TRI em avaliações educacionais já tem exatamente duas décadas. Em 1995 o Sistema Nacional de Avaliação da Educação Básica - SAEB teve sua metodologia reestruturada e incorporou as técnicas da TRI. Com o passar dos anos e com as experiências adquiridas com o uso da TRI no SAEB, outras avaliações em larga escala adotaram esse modelo: o Exame Nacional para Certificação de Competências de Jovens e Adultos (ENCCEJA), a Prova Brasil e, em 2009, o ENEM.

Algumas provas no âmbito internacional também utilizam a TRI. O TOEFL - exame de proficiência em língua inglesa e o SAT (Scholastic Aptitude Test) - avaliação que serve como critério para admissão de estudantes em algumas Universidades dos Estados Unidos, são dois grandes exemplos de avaliações que utilizam essa teoria.

A TRI tem como elemento central os itens e não a prova como um todo. É um conjunto de modelos matemáticos que supõe que o desempenho de um indivíduo é definido

pela sua proficiência ou habilidade e pelas características do item, ou seja, procura definir a probabilidade de um indivíduo com determinada habilidade acertar um item com determinadas características.

Dessa maneira, a TRI permite que alunos que acertam a mesma quantidade de itens, tirem notas diferentes em uma determinada prova. Isso acontece, pois, os itens têm características específicas diferentes que são levadas em consideração para o cálculo final da nota do aluno. Além disso, alunos com o mesmo escore podem ter proficiências diferentes, pois a TRI levará em consideração a coerência das respostas dadas aos itens, e quanto mais coerente, maior a proficiência. Algumas características consideradas pela TRI são: a discriminação (a), a dificuldade (b) e o acerto ao acaso (c).

Para medir o nível de proficiência (θ) de um indivíduo, a TRI estabelece uma escala padrão de conhecimento e aptidões, denominado traço latente. Traço latente são características dos indivíduos que não podem ser aferidas diretamente e que, para medi-las, utilizam-se variáveis secundárias a elas relacionadas. Para exemplificar, caso seja necessário estimar a idade de uma pessoa, é possível fazer isso por meio de um questionário. Algumas perguntas poderiam ser feitas e daria uma boa precisão da idade da pessoa, por exemplo: “Você viu o Brasil ser tetra campeão mundial da Copa do Mundo de Futebol?”, “Você utilizou alguma moeda brasileira antes do Real?”, “Você viu os Beatles tocar?”. Análogo a essas questões que se relacionam à idade de uma pessoa, os itens de uma prova estão relacionados à proficiência dos estudantes.

Diferentemente da TCT, o instrumento de medida independe do objeto que está sendo medido, isto é, a proficiência do aluno não depende dos itens que ele respondeu. O item mede determinado conhecimento independente de quem o responda.

Fazendo uma associação entre a TRI e o exemplo dado sobre a idade do indivíduo, podemos ressaltar que não existe um único conjunto de questões capazes de realizar essa medida e que pessoas de mesma idade respondem as questões de maneira igual. Analogamente, na TRI, não existe um único conjunto de itens capazes de medir a proficiência de um aluno e itens com mesma dificuldade são respondidos de maneira igual por pessoas com mesma proficiência.

Por meio da interpretação da escala de proficiência, é possível atribuir significado prático aos valores encontrados. Para essa interpretação, deve-se primeiramente identificar os “itens âncoras”, ou seja, verificar o nível da escala no qual 65% ou mais dos avaliados situados neste nível acertam o item enquanto, no nível imediatamente inferior, somente menos da metade dos alunos consegue acertar. Além disso, a diferença entre os percentuais dos que acertaram deve ser maior que 30%. No entanto, neste trabalho, na análise psicométrica e pedagógica dos itens, a identificação dos itens âncoras se restringiu apenas ao primeiro critério citado (65% ou mais dos avaliados situados no nível acertam o item).

Após a identificação desses itens âncoras, é feita uma interpretação pedagógica dos resultados, estabelecendo as habilidades que os alunos são capazes de demonstrar nos diversos níveis que compõem a escala. Esse foi um dos grandes ganhos da TRI em relação à TCT, pois, além de inserir indivíduos na escala de proficiência, também localiza os itens dos testes.

Além da interpretação dos níveis da escala, há também a possibilidade de se fazer a equalização dos parâmetros, uma transformação linear que torna comparáveis as profici-

ências de indivíduos que se submeteram a testes diferentes, desde que alguns pressupostos sejam observados, como, por exemplo, os testes tenham um conjunto de itens comuns, denominados itens de ligação. Essa foi a grande inovação da TRI, pois permite comparar os resultados de provas aplicadas a alunos diferentes e em anos diferentes, não sendo necessário que a prova seja exatamente a mesma.

Segundo Karino (2012), essa foi uma das grandes justificativas para a implementação da Teoria de Resposta ao Item no Exame Nacional do Ensino Médio, em 2009.

A decisão de implementar no Exame Nacional do Ensino Médio (ENEM) a Teoria de Resposta ao Item (TRI) teve duas finalidades principais: (1) permitir a comparabilidade dos resultados entre os anos e (2) permitir a aplicação do Exame várias vezes ao ano.

O modelo de TRI que iremos utilizar nesse trabalho na análise da Primeira Fase da OBMEP de 2014 é o modelo logístico de três parâmetros (*Modelo 3LP*), o mesmo utilizado para a análise do desempenho dos estudantes no Enem e no SAEB.

O modelo logístico de três parâmetros tem suas características particulares que diferenciam de outros modelos de TRI. É o modelo aplicado em testes com itens dicotômicos ou itens de múltipla escolha dicotomizados, ou seja, corrigidos como certo ou errado, unidimensionais (nos quais apenas um traço latente dominante está sendo medido) e com independência local ou incondicional, ou seja, “para uma dada habilidade, as respostas aos diferentes itens da prova são independentes” (ANDRADE, TAVARES e VALLE, 2000).

Neste modelo, estima-se o traço latente do respondente a partir de uma relação, onde a probabilidade do indivíduo responder corretamente um item é dada em função da sua habilidade, e de três características do item já citadas anteriormente: dificuldade e discriminação do item e probabilidade de acerto ao acaso.

O parâmetro dificuldade do item indica o nível mínimo de proficiência para que a probabilidade de acerto ao item seja alta. Desta maneira, cada item estabelece uma proficiência mínima. A palavra alta pode soar um pouco relativa, e realmente é: a probabilidade de acerto ao item ser alta dependerá da probabilidade de acerto ao acaso, que varia de prova para prova.

O parâmetro discriminação é o poder do item em diferenciar indivíduos com diferentes habilidades, enquanto a probabilidade de acerto ao acaso, também conhecida como “chute” é a probabilidade de um aluno com baixa aptidão acertar um item.

Conforme Rabelo (2013, p.130), para o cálculo da probabilidade de um indivíduo j responder corretamente o item i , utiliza-se a seguinte fórmula matemática:

$$P(X_{ji} = 1|\theta_j) = c_i + \frac{(1 - c_i)}{1 + e^{-Da_i(\theta_j - b_i)}}, \text{ com } i, j \in \mathbb{N},$$

em que

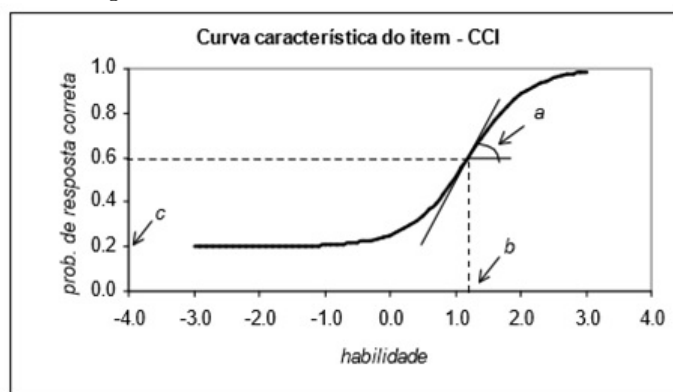
- X_{ji} é a resposta do indivíduo j ao item i . É uma variável binária cujo valor é 1, se o indivíduo acertar o item, ou 0, caso a resposta esteja incorreta;

- a_i é a discriminação do item, que é sempre positiva;
- b_i é a dificuldade do item, medida na mesma escala da habilidade;
- c_i é a chance de acerto ao acaso do item, que varia entre 0 e 1;
- θ_j é o traço latente ou habilidade do indivíduo j ;
- D é um fator de escala, que é igual a 1 na métrica logística e igual a 1,7 na métrica normal.

Análogo à AGI utilizada na Teoria Clássica do Item, na TRI também existe um gráfico que descreve resultados e fornece informações importantes sobre a qualidade do item: A Curva Característica do Item (CCI). A CCI representa a probabilidade de uma resposta estar correta em cada grupo de indivíduo com habilidades diferentes.

O gráfico da CCI é uma sigmoide, ou seja, uma curva cujo formato é semelhante à letra “S”, com duas assíntotas horizontais, conforme ilustra a figura a seguir (Fonte: ANDRADE, TAVARES e VALLE, 2000, p.11, SINAPE).

Figura 3: Curva Característica do Item



Infere-se com o formato da curva que quanto maior a habilidade do aluno, maior a probabilidade de responder o item corretamente, porém, essa relação não é linear.

O formato da curva pode ser mais ou menos inclinado e deslocado, os parâmetros dificuldade do item (b), discriminação do item (a) e acerto ao acaso (c) que determinam o formato da CCI.

O parâmetro acerto ao acaso (c) é uma probabilidade, portanto, seu valor pertence ao intervalo $[0, 1]$. Em provas em que o “chute” não é permitido, esse valor é zero. Na TRI, esse parâmetro é a probabilidade de um aluno com baixa habilidade responder corretamente o item e corresponde ao ponto em que a assíntota horizontal inferior intercepta o eixo das probabilidades. Na **Figura 3**, o valor de c é igual a 0,2, o que faz acreditar que o exemplo representa a CCI de um item com 5 opções de escolha.

É interessante observar que nem sempre o valor de c é exatamente o esperado, isso pode indicar uma falha na formulação e construção do item, com distratores não plausíveis, “peguinhas”, opção com estrutura ou extensão discrepante em relação aos outros itens, entre outras. Nesses casos, é interessante analisar o item e se possível, reelaborá-lo para que ele cumpra sua função avaliativa.

O parâmetro dificuldade do item (b) é o valor da habilidade mínima necessária para que a probabilidade de acerto daquele item seja igual a $\frac{(1+c)}{2}$. Quanto menor o valor de b , mais fácil é o item. Para determinar esse parâmetro, sugere-se que seja traçada uma reta horizontal na probabilidade $\frac{(1+c)}{2}$ e na intersecção com a CCI ver qual a habilidade correspondente.

Note-se que, quando não é permitido “chutar”, ou seja, quando $c = 0$, tem-se que $\frac{(1+c)}{2} = \frac{(1+0)}{2} = \frac{1}{2}$. Isso indica que, em provas em que o “chute” é proibido, a dificuldade do item indica a habilidade necessária para que os alunos tenham 50% de chance de acertar o item.

Teoricamente, a escala que representa a dificuldade do item pertence ao intervalo $(-\infty, +\infty)$. Porém, na prática seus valores variam entre $(-3, +3)$. Para Rabelo (2013), em uma avaliação educacional, sugere-se que a distribuição e a classificação de itens seguindo a sua dificuldade siga a seguinte tabela:

Tabela 7: Distribuição esperada e classificação do item em relação ao seu nível de dificuldade.

Distribuição esperada	Classificação	Dificuldade do item
10%	Muito fáceis	Até -1,28
20%	Fáceis	De -1,27 a -0,52
40%	Medianos	De -0,51 a 0,51
20%	Díficeis	De 0,52 a 1,27
10%	Muito difíceis	1,28 ou mais

Fonte: Rabelo, 2013, p. 143

O parâmetro discriminação do item (a) é a capacidade de diferenciar alunos que têm habilidades distintas. De fato, um item pode ser considerado muito fácil por um aluno com habilidade altíssima, porém pode ser considerado muito difícil por outro respondendo com baixa proficiência.

Na CCI, a discriminação é a derivada da curva no ponto de inflexão da curva, ou seja, no ponto em que a curva muda de concavidade. Geralmente, a discriminação de um item tem valor entre 0 e 2.

Nota-se que o parâmetro a influencia diretamente no formato da CCI. Se $a = 0$, a CCI deixaria de ser uma sigmoide e seria uma reta horizontal, indicando que alunos com habilidades diferentes tem a mesma probabilidade de acertar o item ou seja, o item não discrimina. Se $a < 0$, o item não cumpriria um de seus propósitos fundamentais pois alunos com habilidades baixas teriam maior probabilidade de acertar o item se comparados com alunos com habilidades maiores, não fazendo sentido em um processo avaliativo. Portanto, espera-se de um bom item que o parâmetro discriminação seja sempre positivo e distante do zero, conforme mostra a **Tabela 8**.

Tabela 8: Classificação do item quanto ao seu poder de discriminação.

Discriminação do item	Valores do parâmetro a
Nenhuma	$a = 0$
Muito baixa	$0 < a \leq 0,35$
Baixa	$0,35 < a \leq 0,65$
Moderada	$0,65 < a \leq 1,35$
Alta	$1,35 < a \leq 1,70$
Muito alta	$a \geq 1,70$

Fonte: *Rabelo, 2013, p.138*

A influência do parâmetro de discriminação sobre a inclinação da curva pode ser facilmente observado nas figuras a seguir. As figuras representam as CCI's de itens considerados difíceis, nos quais os parâmetros dificuldade do item e probabilidade de acerto ao acaso foram mantidos constantes e iguais a 1 e 0,2, respectivamente. As mudanças observadas no formato das CCI's foram obtidas variando o valor do parâmetro discriminação.

Figura 4: CCI de um item com $a = 0$

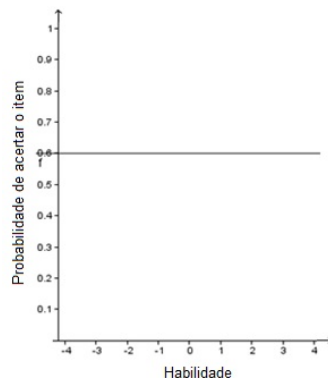


Figura 5: CCI de um item com $a = 0,5$

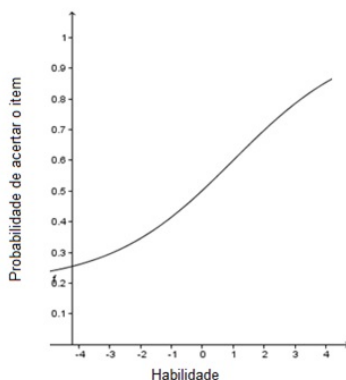


Figura 6: CCI de um item com $a = 1$

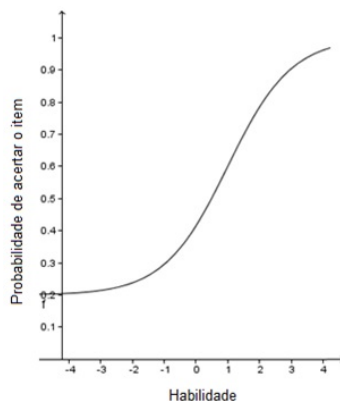
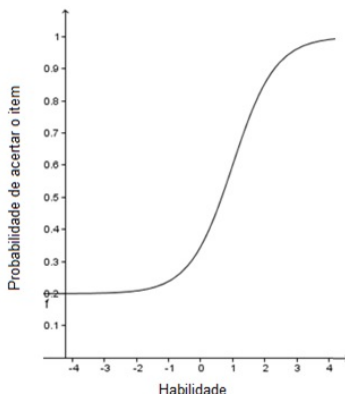


Figura 7: CCI de um item com $a = 1,5$



Infere-se das figuras que quanto mais alto o valor da discriminação, mais íngreme é a curva no ponto de inflexão, portanto, mais discriminativo é o item, o que o torna mais adequado para uma avaliação educacional.

Na TRI, a habilidade de um aluno pode assumir qualquer valor real, porém, é necessário a definição de uma escala, com o estabelecimento de uma origem e uma unidade de medida, que representarão, respectivamente, a média e o desvio padrão das habilidades dos respondentes.

Na **Figura 3**, o gráfico representa um item cuja média foi 0 e o desvio-padrão 1, ou seja, utilizou-se uma escala (0, 1). Os valores dos parâmetros dificuldade e discriminação também foram definidos nas **Tabelas 7 e 8**, utilizando-se essa mesma escala.

Independentemente da escala utilizada, o importante é interpretar o valor e as relações existentes entre os pontos. Assim como no Enem, neste trabalho utilizaremos a escala (500, 100), ou seja, uma escala com média igual a 500 e desvio-padrão igual a 100. Já a Avaliação Nacional da Educação Básica (ANEB) e a Prova Brasil utilizam a escala com média 250 e desvio-padrão 50, ou seja, escala (250, 50).

Para exemplificar, suponha que, na **Figura 3** mostrada anteriormente, os valores dos parâmetros a e b sejam 0,6 e 1,2, respectivamente. Na escala (500, 100), esses valores corresponderiam a 0,006 e 620.

Para transformação da escala $(0, 1)$ para uma escala qualquer (μ, σ) , utilizam-se as seguintes fórmulas:

$$\theta^* = \sigma\theta + \mu;$$

$$a^* = \frac{a}{\sigma};$$

$$b^* = \sigma b + \mu,$$

em que, θ^* , a^* e b^* são os novos valores da habilidade (θ), da discriminação (a) e dificuldade (b) do item na escala (μ, σ) .

As **Tabelas 9 e 10** representam a classificação do item em relação aos parâmetros dificuldade e discriminação na escala $(500, 100)$. Os valores estão relacionados com as **Tabelas 7 e 8**.

Tabela 9: Classificação do item na escala $(500, 100)$ em relação a sua dificuldade.

Classificação	Dificuldade do item
Muito fáceis	Até 372
Fáceis	De 373 a 448
Medianos	De 449 a 551
Díficeis	De 552 a 627
Muito díficeis	628 ou mais

Tabela 10: Classificação do item na escala $(500, 100)$ quanto ao seu poder de discriminação.

Discriminação do item	Valores do parâmetro a
Nenhuma	$a = 0$
Muito baixa	$0 < a \leq 0,0035$
Baixa	$0,0035 < a \leq 0,0065$
Moderada	$0,0065 < a \leq 0,0135$
Alta	$0,0135 < a \leq 0,0170$
Muito alta	$a > 0,0170$

É interessante ressaltar que independentemente da escala utilizada, a probabilidade de um indivíduo responder corretamente a determinado item é sempre a mesma.

A expansão do uso da TRI no Brasil ficou mais evidente com a reestruturação do ENEM em 2009. O uso dessa teoria promoveu melhorias no SAEB, permitindo uma melhor interpretação e análise de desempenho dos estudantes ao longo dos anos, além de melhorias no Sistema Nacional de Avaliação da Educação Superior (SINAES).

O novo formato do ENEM permite a avaliação do desempenho de estudantes que ingressam nas Instituições de Educação Superior. Desde 2012, as notas desta avaliação são também utilizadas para o cálculo do Indicador de Diferença entre os Desempenhos Observado e Esperado (IDD), um dos indicadores associados ao SINAES.

Essa funcionalidade do novo exame passou a vigorar a partir de 2012, quando as notas dos estudantes do ENEM passaram a ser usadas como preditoras de seus desempenhos como concluintes da educação superior, na composição de um índice associado a esse nível de ensino, denominado Indicador de Diferença entre os Desempenhos Observados e Esperado (IDD). (RABELO, 2013, p.59)

Outros indicadores associados ao SINAES que utilizam indiretamente a TRI por meio da nota do ENEM são: o Conceito Preliminar de Curso (CPC) e o Índice Geral de Cursos Avaliados da Instituição (IGC).

Para o cálculo da nota do CPC é realizado uma média ponderada entre alguns elementos, entre eles a Nota IDD, que tem o peso de 35%. Para o cálculo do IGC é utilizada a média dos últimos CPC's disponíveis dos cursos avaliados da instituição no ano do cálculo e nos dois anos anteriores, levando-se em consideração a quantidade de matrículas em cada um dos cursos computados.

Portanto, é possível perceber a grande importância do uso da TRI na análise de desempenho nas avaliações educacionais de larga escala, permitindo extrair conclusões e fornecendo insumos e argumentos para formular e reformular ações educacionais e novas políticas públicas voltadas à educação.

3 Material e Métodos

Este trabalho busca analisar o desempenho dos estudantes e investigar a qualidade da Primeira Fase do Nível II da OBMEP 2014, utilizando a Teoria Clássica dos Testes e a Teoria de Resposta ao Item, com finalidade de maior valorização dessa avaliação.

Embora existam várias provas de Matemática de larga escala que pudessem ser analisadas, a Primeira Fase da OBMEP foi escolhida em razão de esta autora ser professora de Matemática em uma escola da rede pública do Distrito Federal, ter sido coordenadora da OBMEP na escola e ter notado que poderia contribuir para que essa avaliação atingisse em sua totalidade os objetivos por ela propostos. Pelo fato de a autora ter mais experiência lecionando nas séries finais do Ensino Fundamental, o instrumento de estudo deste trabalho restringiu-se ao Nível II da Primeira Fase da OBMEP, prova realizada por estudantes de oitavo e nonos anos.

Para entender um pouco melhor sobre a Olimpíada Brasileira de Matemática e esclarecer possíveis dúvidas, a autora dessa pesquisa reuniu-se pessoalmente com o Coordenador da OBMEP no DF e manteve contato via e-mail com dois coordenadores nacionais.

O levantamento de referencial teórico bibliográfico foi um passo muito importante para a elaboração deste trabalho, iniciado com uma investigação sobre o regulamento, a relevância e fatos históricos relacionados à OBM e à OBMEP e um estudo sobre a TCT e a TRI.

3.1 Coleta de Dados

Com a finalidade de investigar e analisar as respostas dos estudantes, foi escolhida uma escola em uma Região Administrativa do Distrito Federal, na qual 534 estudantes se submeteram à avaliação. Portanto, para realização deste trabalho, utilizou-se uma amostragem não probabilística por conveniência, definida pela facilidade de contato entre a autora do estudo e os diretores e coordenadores da escola.

Foi solicitado junto ao professor coordenador da OBMEP na escola e ao diretor da instituição de ensino, a cessão dos gabaritos dos estudantes. Para isso, foi redigido um Termo de Consentimento Livre e Esclarecido (**Apêndice A**), que apresentava os objetivos do estudo, explicava a necessidade do uso das respostas dos estudantes e evidenciava o sigilo do nome da escola e dos alunos participantes.

Os gabaritos originais dos estudantes que não foram selecionados para a Segunda Fase da OBMEP foram cedidos para o estudo, enquanto os gabaritos dos estudantes selecionados para etapa seguinte da avaliação foram fotografados e escaneados antes de serem enviados pelos correios aos organizadores da prova para dar prosseguimento à avaliação.

Apesar de terem sido inscritos 729 estudantes no Nível II, aproximadamente 25% dos estudantes não compareceram no dia de aplicação do exame. Desse modo, a prova foi realizada por apenas 534 estudantes. Entre os estudantes inscritos na Primeira Fase, aproximadamente 5%, ou seja, 37 alunos com maior desempenho foram selecionados para a segunda etapa. Para a seleção, o critério de desempate utilizado pela escola foi o valor da nota em Matemática no bimestre anterior.

A prova é constituída por 20 itens e o melhor desempenho registrado no colégio foi de dois estudantes que acertaram 50% dos itens, obtendo escore bruto de 10 pontos. Outros estudantes que passaram para segunda etapa conquistaram notas superiores ou iguais a 7, conforme será visto no **Capítulo 4**.

Todos os gabaritos dos estudantes foram numerados de 1 a 534 com a finalidade de verificar alguma divergência na digitação ou nos escores quando corrigidos manualmente pelos próprios professores do colégio. Os gabaritos dos estudantes selecionados para a Segunda Fase foram numerados de 498 a 534. Após a numeração, os gabaritos foram digitados e armazenados em uma planilha eletrônica do programa *Excel*, em formato de tabela contendo algumas informações tais como: número do estudante, opção de marcação em cada uma das alternativas e escore geral. Nos casos de dupla marcação ou itens sem resposta, foi atribuído um ponto-final (.), para que não gerasse dúvidas no processamento das macros e utilização de fórmulas no programa.

Com os dados tabelados e com o apoio do Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos (CEBRASPE), cujo requerimento de apoio e aceite da instituição estão contidos no **Apêndice B**, foram produzidas estatísticas do tipo descritivas, no caso da TCT, e do tipo descritivo-probabilísticas, no caso da TRI, utilizadas posteriormente para a análise de cada um dos itens e da prova como um todo, gerando valores para os parâmetros de dificuldade, discriminação, acerto ao acaso, ponto-bisserial e proficiência. Nessa pesquisa, utilizou-se uma abordagem quantitativa e qualitativa, pois, além de determinar valores para parâmetros via TCT e TRI, buscou-se entender e dar significado pedagógico aos resultados, além de se apresentar uma sugestão de retorno aos elaboradores da prova com propostas de valorização do instrumento.

Na análise quantitativa, o que serve de informação é a frequência com que surgem certas características do conteúdo. Na análise qualitativa é a presença ou a ausência de uma dada característica de conteúdo ou de um conjunto de características num determinado fragmento de mensagem que é tomado em consideração (BARDIN, 1977, p.21).

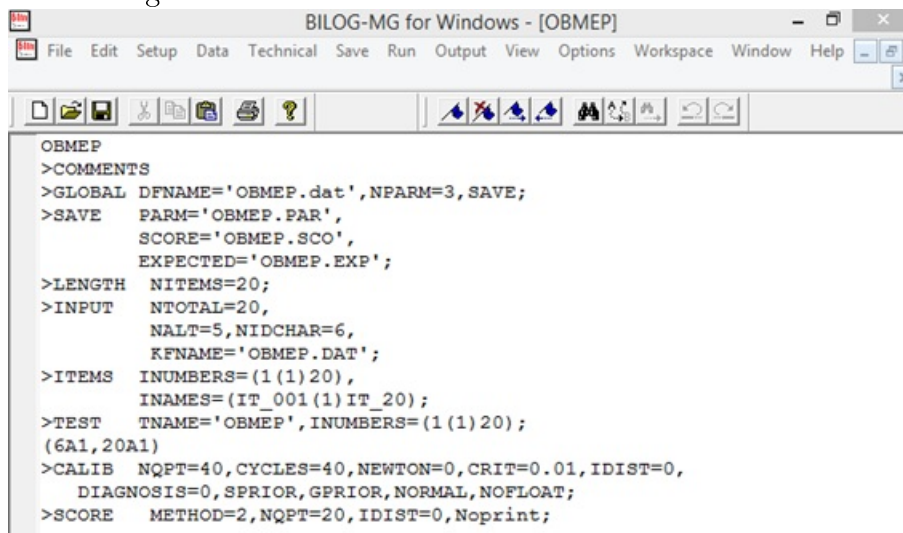
3.2 Programas Computacionais Utilizados

Primeiramente, foi realizado o estudo dos dados via TCT utilizando o programa *Excel*. Com os gabaritos inseridos na planilha eletrônica, os resultados dos itens foram transformados em valores numéricos dicotômicos. Atribuiu-se 1 para cada resposta correta e 0 para as respostas incorretas, permitindo, assim, uma nova correção dos gabaritos, dessa vez, computacionalmente e isento de erros. Com a utilização de uma macro, foi possível obter a dificuldade e a discriminação de cada item, o percentual de marcação de cada opção incorreta, o coeficiente bisserial do item e de cada um dos distratores. Além disso, utilizando-se macros cedidas pelo CEBRASPE, foram construídos os gráficos das AGI's, que revelaram o comportamento de cada item para aquela amostra de estudantes, apresentando o percentual de marcações de cada opção de acordo com o escore do aluno. Apesar da prova ter uma pontuação máxima de 20 pontos, o maior escore alcançado pelos estudantes da amostra foi 10 pontos, portanto, as AGI's construídas só apresentam as faixas de notas entre 0 e 10. Para facilitar o manuseio e a interpretação dos gráficos, as AGI's foram salvas em um arquivo em formato PDF.

Para a análise via TRI foram utilizados os softwares *SPSS (Statistical Package for Social Sciences)* para *Windows* e o *BILOG-MG*. O *SPSS* foi utilizado para organizar o banco de dados, que, posteriormente, foi salvo com a extensão *.dat* e utilizado pelo *BILOG-MG* para a geração das estatísticas e parâmetros da TRI.

Após a organização do banco de dados, iniciou-se a programação. A figura a seguir apresenta a estrutura do arquivo de comando do *BILOG-MG*.

Figura 8: Tela de entrada do software *BILOG-MG*



```

OBMEP
>COMMENTS
>GLOBAL  DFNAME='OBMEP.dat',NPARM=3,SAVE;
>SAVE    PARM='OBMEP.PAR',
        SCORE='OBMEP.SCO',
        EXPECTED='OBMEP.EXP';
>LENGTH NITEMS=20;
>INPUT  NTOTAL=20,
        NALT=5,NIDCHAR=6,
        KFNAME='OBMEP.DAT';
>ITEMS  INUMBERS=(1(1)20),
        INAMES=(IT_001(1)IT_20);
>TEST   TNAME='OBMEP',INUMBERS=(1(1)20);
(6A1,20A1)
>CALIB  NQPT=40,CYCLES=40,NEWTON=0,CRIT=0.01,IDIST=0,
        DIAGNOSIS=0,SPRIOR,GPRIOR,NORMAL,NOFLOAT;
>SCORE  METHOD=2,NQPT=20,IDIST=0,Noprint;

```

Temos que

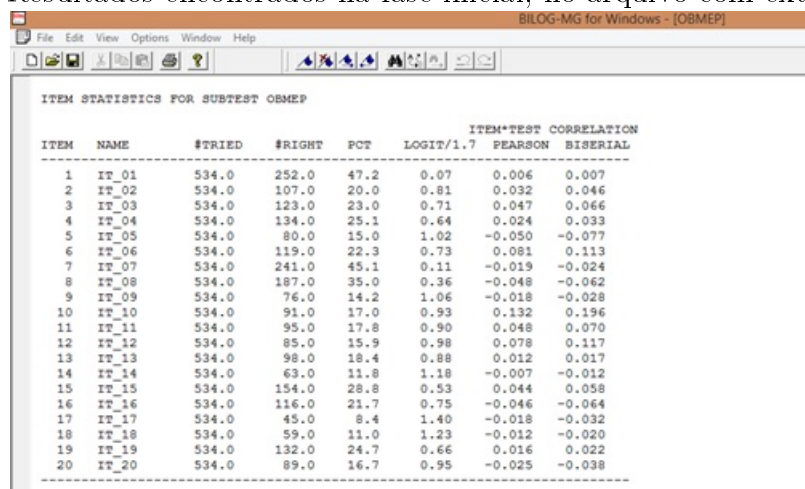
- $NPARM = 3$ é o número de parâmetros utilizado no teste;
- *SAVE* é o local onde foram salvos os arquivos;
- $NITEMS = 20$ é o número de itens da prova 20 que será utilizado para o processamento da TRI no presente estudo.
- $NTOTAL = 20$ é o número total de itens da prova;
- $NALT = 5$ é o número total de alternativas de cada item;
- $NIDCHAR = 6$ é a quantidade de caracteres de identificação de cada indivíduo no banco de dados;
- *KFNAME* é o local onde foi salvo o gabarito;
- $NQPT = 40$ é o número de pontos de quadratura;
- $CYCLES = 40$ é o número de ciclos para convergência;
- $CRIT = 0,01$ é o critério de convergência.

O *BILOG-MG* desenvolve-se através de três fases, apresentando 3 arquivos de saída, o *.PH1*, *.PH2* e *.PH3*. A fase inicial, com extensão *.PH1*, apresenta os resultados das estatísticas via TCT. Os dados gerados na primeira fase são utilizados para a obtenção

dos parâmetros da TRI, na próxima fase do programa. A fase 2, com extensão .PH2, contém detalhes sobre a estimativa, e se o algoritmo convergir, apresenta as estimativas finais dos parâmetros dos itens e dos erros. É nessa fase que os dados são calibrados e obtêm-se os valores dos parâmetros utilizados pela TRI, a discriminação do item (a), a dificuldade (b) e o parâmetro de acerto ao acaso (c). Na fase 3, com extensão .PH3, são geradas as estimativas das proficiências dos estudantes.

Nesse primeiro momento, aparecia uma das maiores dificuldades deste trabalho: 9 itens apresentaram discriminação e correlação biserial negativos e impossibilitaram a continuidade do estudo, pois não possibilitaram convergência dos dados para dar prosseguimento e determinar os parâmetros e os resultados pela TRI.

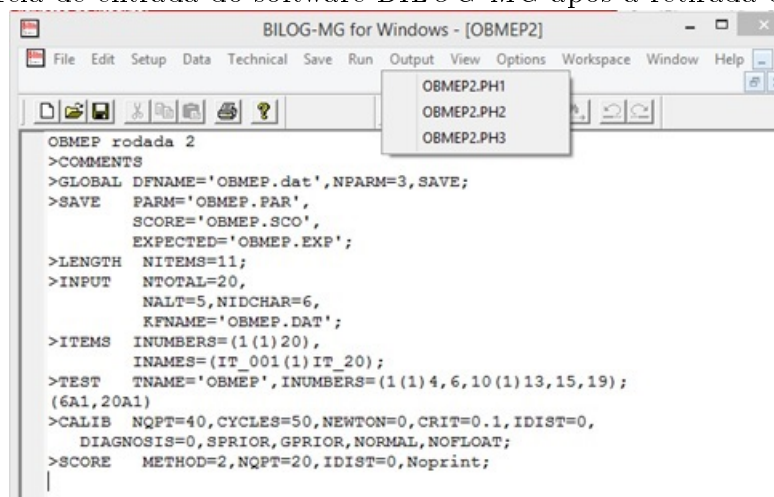
Figura 9: Resultados encontrados na fase inicial, no arquivo com extensão .PH1



ITEM	NAME	#TRIED	#RIGHT	PCT	LOGIT/1.7	ITEM*TEST CORRELATION	
						PEARSON	BISERIAL
1	IT_01	534.0	252.0	47.2	0.07	0.006	0.007
2	IT_02	534.0	107.0	20.0	0.81	0.032	0.046
3	IT_03	534.0	123.0	23.0	0.71	0.047	0.066
4	IT_04	534.0	134.0	25.1	0.64	0.024	0.033
5	IT_05	534.0	80.0	15.0	1.02	-0.050	-0.077
6	IT_06	534.0	119.0	22.3	0.73	0.081	0.113
7	IT_07	534.0	241.0	45.1	0.11	-0.019	-0.024
8	IT_08	534.0	187.0	35.0	0.36	-0.048	-0.062
9	IT_09	534.0	76.0	14.2	1.06	-0.018	-0.028
10	IT_10	534.0	91.0	17.0	0.93	0.132	0.156
11	IT_11	534.0	95.0	17.8	0.90	0.048	0.070
12	IT_12	534.0	85.0	15.9	0.98	0.078	0.117
13	IT_13	534.0	98.0	18.4	0.88	0.012	0.017
14	IT_14	534.0	63.0	11.8	1.18	-0.007	-0.012
15	IT_15	534.0	154.0	28.8	0.53	0.044	0.058
16	IT_16	534.0	116.0	21.7	0.75	-0.046	-0.064
17	IT_17	534.0	45.0	8.4	1.40	-0.018	-0.032
18	IT_18	534.0	59.0	11.0	1.23	-0.012	-0.020
19	IT_19	534.0	132.0	24.7	0.66	0.016	0.022
20	IT_20	534.0	89.0	16.7	0.95	-0.025	-0.038

Para dar continuidade aos estudos, foi necessário remover esses itens da avaliação e ser mais flexível e tolerante quanto ao critério de convergência. Portanto, os dados gerados pela TRI levaram em consideração apenas 11 itens da avaliação.

Figura 10: Tela de entrada do software *BILOG-MG* após a retirada de alguns itens



```

BILOG-MG for Windows - [OBMEP2]
File Edit Setup Data Technical Save Run Output View Options Workspace Window Help
OBMEP2.PH1
OBMEP2.PH2
OBMEP2.PH3
OBMEP rodada 2
>COMMENTS
>GLOBAL DFNAME='OBMEP.dat', NPARAM=3, SAVE;
>SAVE PARM='OBMEP.PAR',
SCORE='OBMEP.SCO',
EXPECTED='OBMEP.EXP';
>LENGTH NITEMS=11;
>INPUT NTOTAL=20,
NALT=5, NIDCHAR=6,
KFNAME='OBMEP.DAT';
>ITEMS INUMBERS=(1(1)20),
INAMES=(IT_001(1)IT_20);
>TEST TNAME='OBMEP', INUMBERS=(1(1)4,6,10(1)13,15,19);
(6A1,20A1)
>CALIB NQPT=40, CYCLES=50, NEWTON=0, CRIT=0.1, IDIST=0,
DIAGNOSIS=0, SPRIOR, GPRIOR, NORMAL, NOFLOAT;
>SCORE METHOD=2, NQPT=20, IDIST=0, Noprint;
  
```

Nota-se que houve alterações nos valores nas entradas NITEMS = 11, no INUMBERS, onde foram apresentados só os itens que seriam utilizados no estudo, no CRIT = 0,1, definindo-se o novo critério de convergência, e no CYCLES, no qual foi ampliada a quantidade de ciclos para a convergência dos dados.

Os valores foram apresentados no *BILOG-MG* na escala (0, 1) e posteriormente foram convertidos, no *Excel*, para escala (500, 100) para que pudessem ser comparados com outras provas que utilizam essa escala, como o ENEM. Em seguida, foram geradas as CCI's, feitas com macros do *Excel* criadas pelo CEBRASPE, e o arquivo foi salvo em formato PDF para facilitar o manuseio e a interpretação dos gráficos e as análises dos resultados.

4 Aplicação da TCT e da TRI na Primeira Fase da OBMEP de 2014

4.1 Visão Geral da Prova

A partir do desempenho geral dos estudantes da escola pesquisada, a análise do comportamento dos itens que compõem a prova da Primeira Fase da OBMEP de 2014, segundo os parâmetros originados pela Teoria Clássica dos Testes (TCT) e pela Teoria de Resposta ao Item (TRI), pode ser resumida na **Tabela 11** a seguir. Observa-se que os parâmetros escolhidos para a TCT foram os índices de dificuldade e de discriminação e o coeficiente bisserial. Para a TRI, foi utilizado o modelo logístico de três parâmetros, obtendo-se, para cada item, os parâmetros de discriminação (a), dificuldade (b) e de acerto ao acaso (c).

Tabela 11: Visão geral da prova da Primeira Fase da OBMEP de 2014.

Itens	TCT			TRI		
	Dificuldade	Discriminação	Bisserial	Discriminação	Dificuldade	Acerto ao acaso
1	0,472	0,326	0,344	0,001775	841,6984057	0,26276
2	0,200	0,245	0,354	0,003698	1298,454384	0,18796
3	0,230	0,254	0,379	0,004451	1300,400797	0,22282
4	0,251	0,270	0,351	0,003842	1108,928255	0,21693
5	0,150	0,113	0,220	-	-	-
6	0,223	0,325	0,422	0,003981	970,659876	0,15527
7	0,452	0,280	0,315	-	-	-
8	0,350	0,251	0,273	-	-	-
9	0,142	0,119	0,266	-	-	-
10	0,170	0,326	0,489	0,004258	973,1023029	0,10565
11	0,178	0,218	0,372	0,005395	1055,728521	0,15347
12	0,159	0,242	0,412	0,005445	939,9269265	0,10197
13	0,182	0,214	0,330	0,003633	1196,581045	0,15988
14	0,118	0,115	0,273	-	-	-
15	0,290	0,294	0,373	0,004236	879,6523472	0,19127
16	0,215	0,196	0,256	-	-	-
17	0,084	0,084	0,238	-	-	-
18	0,110	0,132	0,262	-	-	-
19	0,247	0,269	0,341	0,003834	1164,685562	0,22201
20	0,167	0,130	0,264	-	-	-

Pela TCT, infere-se dos resultados apresentadas na tabela que a prova não apresentou itens de todas as variações de graus de dificuldade, fato que sugere inadequação do instrumento no processo avaliativo. Segundo Vianna (1989), “para fins de avaliação de larga escala, os testes devem ser compostos de itens que alcancem todo o *continuum* da escala, ou seja, devem ter uma amplitude que inclua itens fáceis, medianos e difíceis.”

No presente estudo, apenas os itens 1, 7 e 8 revelaram-se com dificuldade mediana, a grande maioria dos itens pode ser classificada como difícil, com exceção do item 17, que

teve nível de dificuldade extremamente alto, com apenas 8,4% de acerto, conforme está apresentado na **Tabela 12** a seguir.

Tabela 12: Distribuição dos itens em relação ao parâmetro dificuldade, segundo a TCT.

Classificação	Valores	Itens	Percentual de itens na prova
Muito fácil	0,9 ou mais	Nenhum	0%
Fácil	Maior ou igual a 0,7 e menor que 0,9	Nenhum	0%
Moderado	Entre 0,3 e 0,7	1, 7 e 8	15%
Difícil	Entre 0,1 e 0,3	2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19 e 20	80%
Muito difícil	Até 0,1	17	5%

Em relação à discriminação, a prova apresentou sete itens deficientes, com discriminação muito baixa e que deveriam ser rejeitados, dez itens marginais, sujeitos a reelaboração e três itens bons, mas sujeitos a aprimoramento. Os itens que menos discriminaram foram, em ordem crescente, os itens 17, 5, 14, 9, 20, 18 e 16. Já os itens que tiveram o maior poder de discriminação foram os itens 1, 6 e 10. A **Tabela 13** resume essa situação:

Tabela 13: Distribuição dos itens em relação a discriminação, pela TCT.

Classificação	Valores	Itens	Percentual de itens na prova
Item deficiente	Até 0,2	5, 9, 14, 16, 17, 18 e 20	35%
Item marginal	$0,2 \leq \text{Disc} < 0,3$	2, 3, 4, 7, 8, 11, 12, 13, 15 e 19	50%
Item bom, sujeito a aprimoramento	$0,3 \leq \text{Disc} < 0,4$	1, 6 e 10	15%
Item bom	$\text{Disc} \geq 0,4$	Nenhum	0%

Os escores totais dos alunos variaram apenas entre 0 e 10, em uma escala de 0 a 20 pontos, com média igual a 4,39 e desvio padrão igual a 1,85. A nota mais frequente foi 4,00, com total de 119 estudantes. Na **Figura 11**, apresenta-se uma proposta de boletim de desempenho dos estudantes da escola utilizando-se a TCT. O boletim contempla a distribuição da quantidade dos estudantes em relação aos escores, o que possibilita a cada participante se localizar e comparar seu desempenho com os demais colegas da escola.

Ressalta-se que, em uma avaliação educacional, espera-se uma distribuição balanceada de acertos, com o gráfico equilibrado e centralizado. Nota-se que a curva ilustrada na **Figura 11** distancia-se bastante do padrão de curva normal esperado em avaliações de desempenho dessa natureza.

Figura 11: Modelo de Boletim de desempenho pela TCT



A tentativa de entender o comportamento dos itens pela TRI foi carregada de dificuldades. O processo de obtenção dos parâmetros dos itens e das proficiências dos estudantes envolve a convergência de sequências numéricas, a partir da fixação de um valor de tolerância para as aproximações. Para esta prova, foi necessário retirar 9 dos 20 itens para que houvesse convergência, pois estes itens não se comportaram bem do ponto de vista dos pressupostos da teoria. Sempre que isso ocorre, os itens “problemáticos” precisam ser retirados da análise. Isso por si só já indica que parte substancial da prova tem problemas do ponto de vista da teoria da medida ou é inadequada aos estudantes dessa faixa etária, no grupo estudado.

Considerando apenas os onze itens que possibilitaram a convergência do método, a prova, em sua totalidade, se revelou muito difícil, já que todos os itens apresentaram o nível de dificuldade superior a 628, em uma escala com média 500 e desvio-padrão 100. Em relação ao poder de discriminação, os itens apresentaram baixo poder de discriminação, sendo que o item 1 praticamente não discriminou, sugerindo sua eliminação no processo de análise.

Em relação à escala de proficiência dos alunos, os valores do parâmetro habilidade variaram de 453,35 e 748,69. A maioria obteve proficiência inferior a 500, sendo 31,65% dos estudantes com habilidades entre 470 e 490. Apenas dois alunos alcançaram habilidades superiores a 700. O desempenho dos alunos pode ser observado e comparado usando-se o modelo de boletim de desempenho sugerido na **Figura 12**, no qual os alunos podem se localizar na distribuição e comparar seu desenvolvimento com os demais estudantes da escola.

Figura 12: Modelo de Boletim de desempenho pela TRI



A análise geral da prova revela uma preocupação com a qualidade psicométrica do instrumento, sugerindo uma inadequação para a proposta da OBMEP de servir como mecanismo de seleção de estudantes para a Segunda Fase. Considerando os parâmetros da TRI, dos 20 itens que compõem a prova, apenas 11 deveriam ser considerados para o cálculo da proficiência dos estudantes. A distribuição da quantidade de estudantes segundo os escores totais (**Figura 11**) mostra que não há indivíduos em todas as faixas de notas possíveis (pontuação de 0 a 20). Esse resultado explica, de certo modo, o fato de não ter sido possível a “calibração” de todos os itens pela TRI.

Para entender melhor esse comportamento, faremos a análise psicométrica e pedagógica de cada item na seção a seguir. Serão utilizadas as siglas p_A, p_B, p_C, p_D e p_E para representar a proporção de marcações em cada uma das respectivas alternativas. A sigla p_ indica a proporção de marcações duplas ou respostas em branco. As siglas Bis_A, Bis_B, Bis_C, Bis_D e Bis_E representam a correlação bisserial de cada uma das alternativas. A sigla Bis_ indica a correlação bisserial das marcações duplas e respostas em branco.

Ressalta-se novamente que, embora o valor total da prova seja de 20 pontos, o escore máximo alcançado pelos estudantes da amostra foi 10 pontos. Portanto, as AGI's dos itens foram construídas apresentando apenas as faixas de notas de 0 a 10.

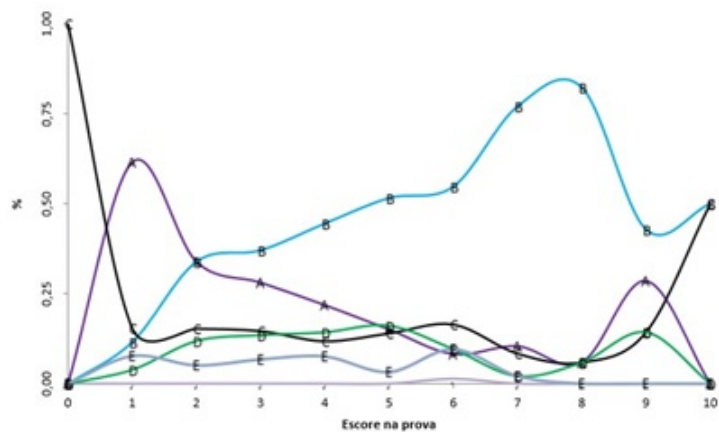
4.2 Análise Psicométrica e Pedagógica dos Itens

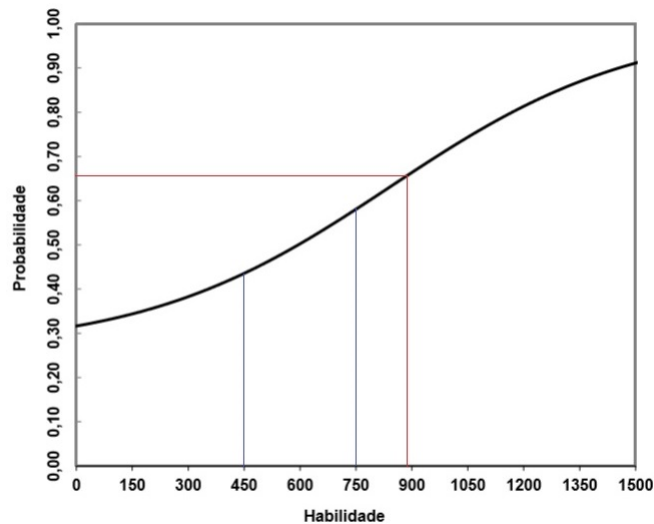
Questão 1

Paula numerou todas as casas do tabuleiro quadrado abaixo, da esquerda para a direita e de cima para baixo, começando com o número 1. A casa central recebeu o número 5. Se ela fizer o mesmo com outro tabuleiro quadrado com 49 casas, qual número será escrito em sua casa central?

1	2	3
4	5	6
7	8	9

- a. 23
- b. 25
- c. 27
- d. 29
- e. 31





TCT	Gabarito	Dificuldade	Discriminação	Bisserial
	B	0,472	0,326	0,344

p_A	p_B	p_C	p_D	p_E	p_.
0,215	0,472	0,137	0,116	0,058	0,002

Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
-0,356	0,344	-0,053	-0,064	-0,088	0,273

TRI	a	b	c
	0,001775	841,6984057	0,26276

Segundo os resultados apresentados pela TCT, o item apresentou grau de dificuldade mediano, já que 47,2% dos estudantes optaram pela alternativa correta, a opção (B). O item apresentou coeficiente de discriminação igual a 0,326, o que o classificaria como um bom item, mas sujeito a aprimoramento. Todos os distratores apresentaram bisserial negativo, conforme esperado. No entanto, a AGI mostra que 50% dos melhores alunos no teste como um todo acabaram optando pela resposta correta e os demais pela opção (C), o que deveria ser investigado, afinal, uma alternativa incorreta não deveria atrair alunos com bom desempenho.

Observando-se o percentual de marcação em cada um dos distratores, verifica-se que pouquíssimos alunos optaram pela alternativa (E). Neste caso, é interessante rever e re-analisar a plausibilidade dessa opção. Será que existe algum pensamento coerente que levaria os estudantes a escolherem essa resposta?

Segundo os resultados apresentados pela TRI o item é classificado como muito difícil e de discriminação muito baixa, portanto, um item que deveria ser descartado ou reelaborado. Na CCI, quase não é possível observar o formato de “S” da curva, no caso de um bom item essa inclinação ficaria evidente. Nota-se no intervalo (450, 750), que compreende a proficiência dos alunos investigados, o formato aproxima-se de uma reta.

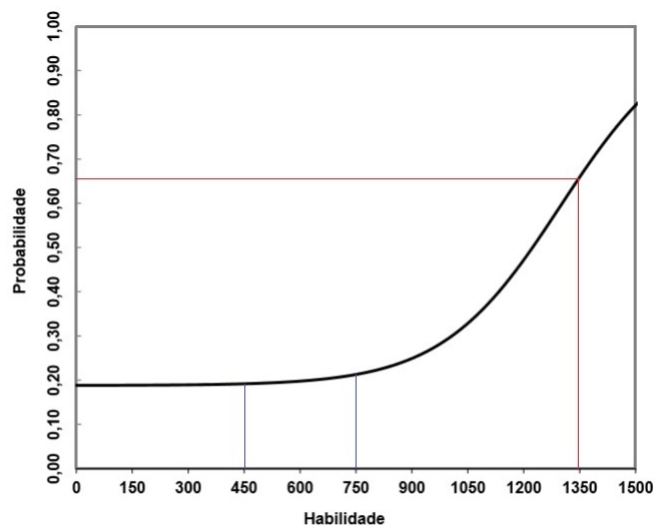
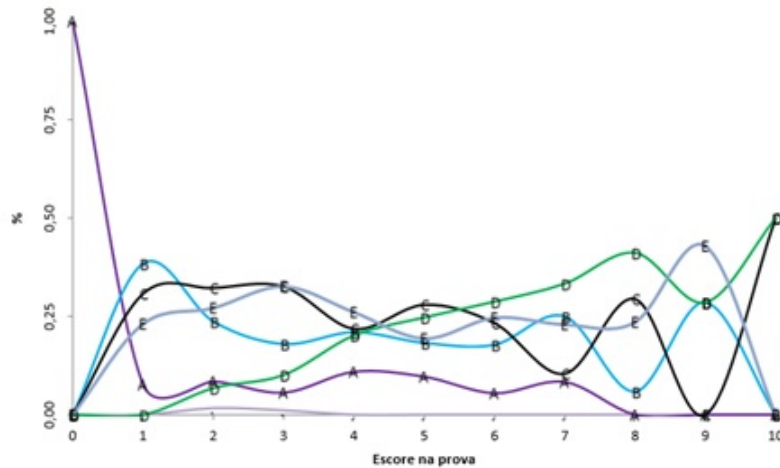
É interessante ressaltar que, de acordo com as condições citadas no capítulo anterior, observa-se na CCI que este item seria considerado item âncora de um nível próximo a 800, o que indica que somente alunos com proficiência igual ou superior a este valor

desenvolveram a habilidade ou possuem os conhecimentos avaliados no item. No universo de estudantes aqui pesquisado, nenhum deles se encontra nessa faixa de proficiência.

Questão 2

Ana Maria apertou as teclas $19 \times 106 =$ de sua calculadora e o resultado 2014 apareceu no visor. Em seguida, ela limpou o visor e fez aparecer novamente 2014 com uma multiplicação de dois números naturais, mas, desta vez, apertando seis teclas em vez de sete. Nesta segunda multiplicação, qual foi o maior algarismo cuja tecla ela apertou?

- a. 5
- b. 6
- c. 7
- d. 8
- e. 9



TCT	Gabarito	Dificuldade	Discriminação	Bisserial
	D	0,200	0,245	0,354

p_A	p_B	p_C	p_D	p_E	p_.
0,081	0,206	0,255	0,200	0,255	0,004

Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
-0,087	-0,085	-0,136	0,354	-0,045	-0,343

TRI	a	b	c
	0,003698	1298,454384	0,18796

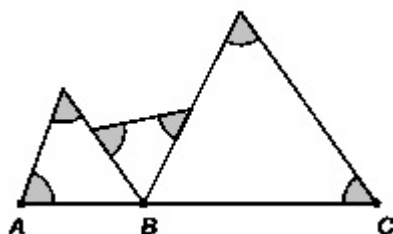
Segundo a TCT, o item apresentou elevado grau de dificuldade, com apenas 20% de acerto. O item apresentou coeficiente de discriminação igual a 0,245, o que o classificaria como um item marginal, sujeito a reelaboração. Todos os distratores apresentaram bisseriais negativos, conforme o esperado. É interessante observar que o gabarito foi uma das alternativas menos escolhida pelos estudantes que fizeram a prova, enquanto 25,5% escolheram a alternativa (C), a mesma quantidade escolheu a alternativa (E) e 20,6% optaram pela letra (B), embora a alternativa correta fosse a letra (D).

Na AGI, espera-se que as linhas referentes aos distratores decresçam à medida que a habilidade dos estudantes aumenta, o que não pôde ser observado no gráfico, sugerindo que algumas opções incorretas (alternativas (D), (C) e (E)) foram escolhidas por um percentual elevado de estudantes com bom desempenho no teste de modo geral.

Segundo os resultados apresentados pela TRI, o item é classificado como muito difícil de discriminação baixa, como se pode notar observando-se a pouca inclinação da CCI. Nota-se na CCI que os estudantes da amostra com maior e menor proficiência (748 e 453, respectivamente) tiveram praticamente a mesma probabilidade de acertar o item, em torno de 20%. No gráfico, fica evidente que este item seria um item âncora de um nível muito alto de habilidade, próximo a 1350, no qual não se encontra nenhum estudante da amostra. Infere-se, então, que o item não cumpriu sua finalidade de avaliar os estudantes, indicando que ele deveria ser descartado ou reelaborado.

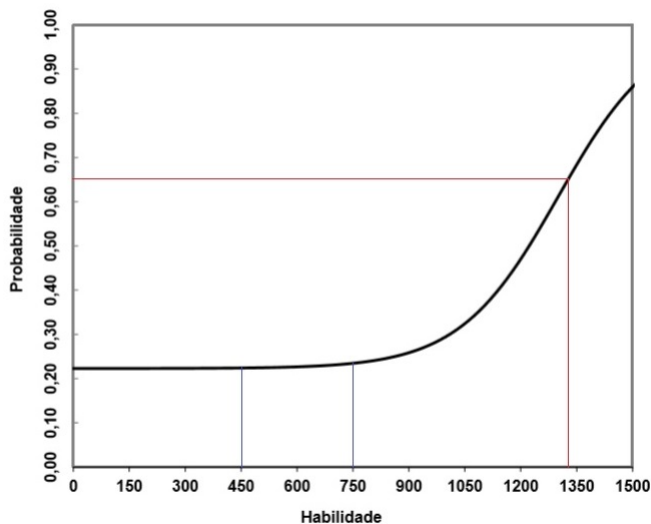
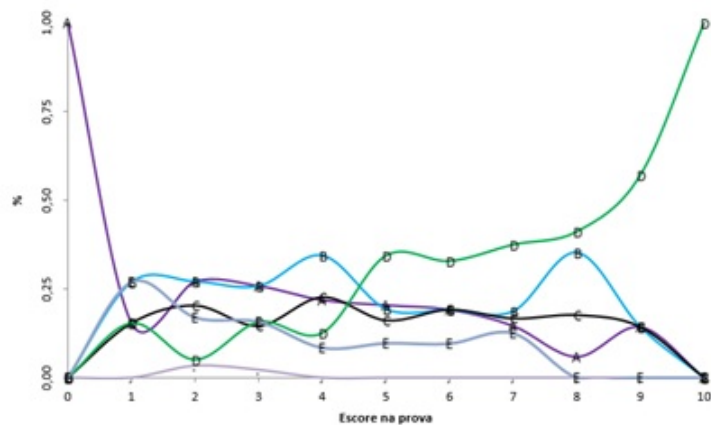
Questão 3

Na figura, os pontos A, B e C estão alinhados. Qual é a soma dos ângulos marcados em cinza?



- a. 120°

- b. 180°
- c. 270°
- d. 360°
- e. 540°



TCT	Gabarito	Dificuldade	Discriminação	Bisserial
	D	0,230	0,254	0,379

p_A	p_B	p_C	p_D	p_E	p_.
0,210	0,253	0,182	0,230	0,118	0,007

Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
-0,133	-0,085	-0,011	0,379	-0,198	-0,370

TRI	a	b	c
	0,004451	1300,400797	0,22282

Pela TCT, o item apresentou elevado grau de dificuldade, com apenas 23% de acerto. O gabarito, letra (D), não foi a alternativa mais escolhida pelos estudantes, já que 25,3% optaram pela alternativa (B), que é uma opção atraente, tendo em vista que na figura

tem apenas triângulos e a informação de que a soma dos ângulos internos de um triângulo é igual a 180° é básica e fundamental no estudo de Geometria. Desta maneira, infere-se que em alguns itens da OBMEP é necessário conhecimento formal da Matemática, e não apenas raciocínio, abstração e criatividade, contrariando a citação de Cláudio Landim (Seção 1.2, Capítulo 1).

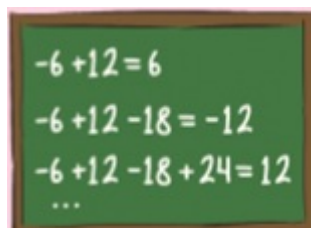
Em relação à discriminação, o valor 0,254 sugere que o item seja considerado um item marginal e sujeito à reelaboração. É interessante notar na AGI que a linha correspondente ao gabarito (D) cresceu com o aumento da habilidade, e que as outras curvas tenderam a diminuir, conforme o esperado. Isso também é característico de um item no qual o bisserial de cada distrator é negativo, sendo positivo apenas o correspondente à alternativa correta.

Pela TRI, o item apresentou um baixo poder de discriminação, 0,004451, e um altíssimo grau de dificuldade, no qual para se ter uma probabilidade alta de acerto ao item o aluno deveria ter proficiência superior a 1300. Nota-se no intervalo (450, 750) destacado na CCI que o formato da curva se aproxima de uma reta horizontal, evidenciando o baixo poder do item em diferenciar alunos com habilidades distintas.

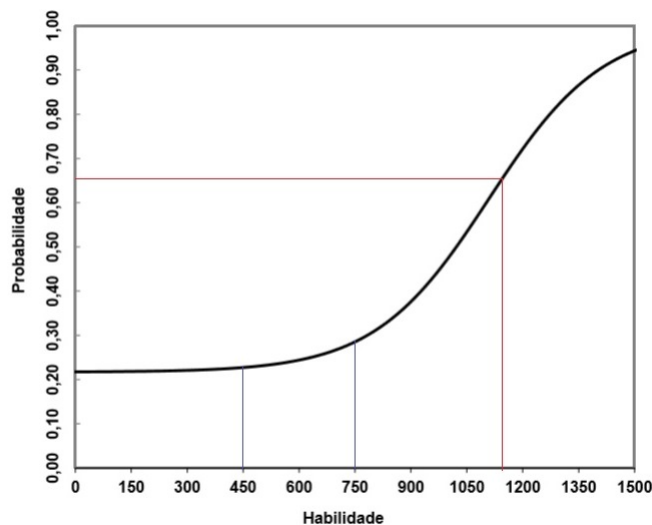
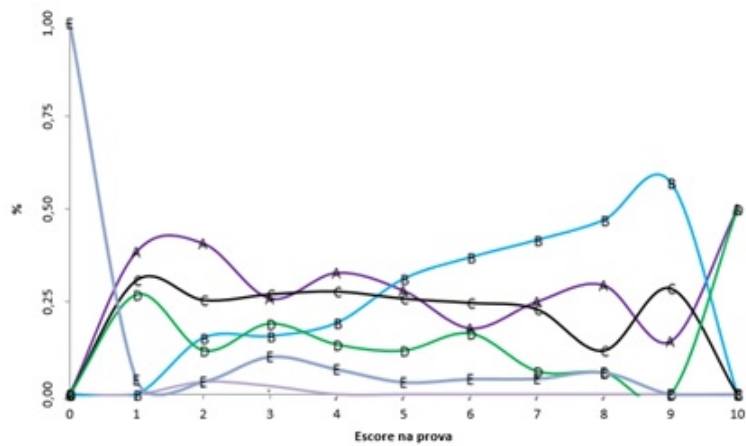
O item caracterizou-se como item âncora de um nível próximo a 1350, o que indica que nenhum aluno desta amostra apresentou as habilidades avaliadas, tendo em vista que 748,69 foi o maior nível de proficiência atingido pelos estudantes.

Questão 4

A sequência -6, 12, -18, 24, -30, 36, ... é obtida a partir dos múltiplos positivos de 6, multiplicando-se os termos nas posições ímpares por -1. Observe na figura que a soma dos dois primeiros termos da sequência é igual a 6 e a soma dos três primeiros termos é igual a -12. Quantos termos consecutivos dessa sequência devemos somar, a partir do primeiro, para obter 180 como resultado?


$$\begin{array}{l} -6 + 12 = 6 \\ -6 + 12 - 18 = -12 \\ -6 + 12 - 18 + 24 = 12 \\ \dots \end{array}$$

- a. 30
- b. 60
- c. 90
- d. 120
- e. 180



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	B	0,251	0,270	0,351		
	<u>p_A</u>	<u>p_B</u>	<u>p_C</u>	<u>p_D</u>	<u>p_E</u>	<u>p_.</u>
	0,288	0,251	0,257	0,140	0,056	0,007
	<u>Bis_A</u>	<u>Bis_B</u>	<u>Bis_C</u>	<u>Bis_D</u>	<u>Bis_E</u>	<u>Bis_.</u>
	-0,125	0,351	-0,065	-0,124	-0,114	-0,370
TRI	<u>a</u>	<u>b</u>	<u>c</u>			
	0,003842	1108,928255	0,21693			

Com 25,1% de acerto e poder de discriminação igual a 0,27, o item revela-se difícil e com discriminação marginal. Em relação à proporção de marcação, destaca-se a alternativa (E), já que apenas 5,6% dos estudantes escolheu essa opção, indicando ser uma opção pouco plausível. Já a alternativa incorreta (A), foi escolhida por 28,8% dos estudantes, superando até mesmo a quantidade de estudantes que acertaram o item.

Observando-se a AGI, percebe-se que estudantes com maiores habilidades não acertaram o item, escolhendo as alternativas (A) e (D) como respostas. Em relação aos bisseriais, todos os distratores apresentaram valores negativos, conforme o esperado.

Pela TRI, o item apresentou dificuldade igual a 1108,928255, o que é característico de um item muito difícil. Observando a CCI, nota-se uma leve inclinação na curva, o

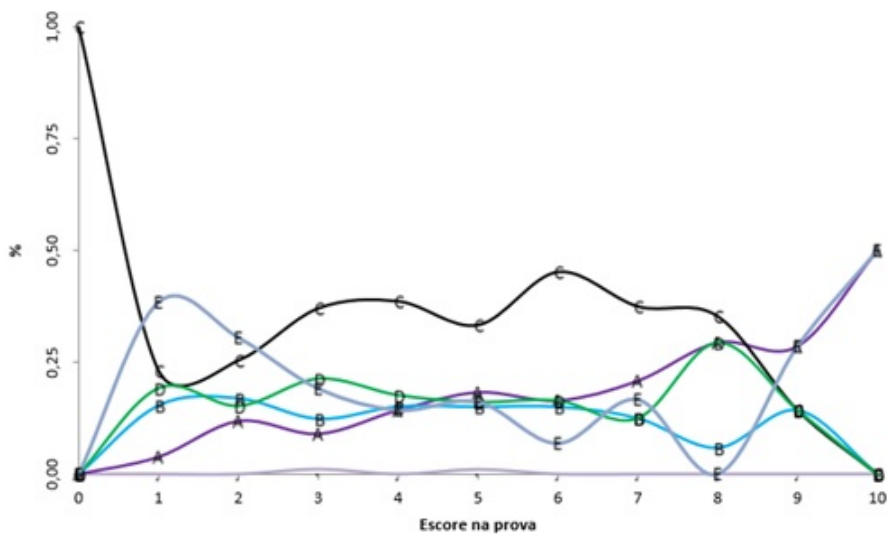
que caracteriza o baixo poder de discriminação apresentado pelo item. Essa inclinação é notada à direita do gráfico, não compreendendo o intervalo em que os alunos avaliados nesta amostra se encontram. Além disso, percebe-se na CCI que se trata de um item âncora de um nível de habilidade próximo a 1100. Na amostra pesquisada, ninguém alcançou esse nível de proficiência, sugerindo que, entre os estudantes avaliados, não houve domínio do conhecimento necessário para o entendimento do item.

Questão 5

Os irmãos Luiz e Lúcio compraram um terreno cercado por um muro de 340 metros. Eles construíram um muro interno para dividir o terreno em duas partes. A parte de Luiz ficou cercada por um muro de 260 metros e a de Lúcio, por um muro de 240 metros. Qual é o comprimento do muro interno?



- a. 80 m
- b. 100 m
- c. 160 m
- d. 180 m
- e. 200 m



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	A	0,150	0,113	0,220		
	<u>p_A</u>	<u>p_B</u>	<u>p_C</u>	<u>p_D</u>	<u>p_E</u>	<u>p_.</u>
	0,150	0,142	0,356	0,174	0,174	0,004
	<u>Bis_A</u>	<u>Bis_B</u>	<u>Bis_C</u>	<u>Bis_D</u>	<u>Bis_E</u>	<u>Bis_.</u>
	0,220	-0,044	0,056	-0,025	-0,214	-0,071

Pela TCT, os valores observados dos parâmetros dificuldade e discriminação sugerem que o item seja considerado difícil e deficiente. Apenas 15% dos estudantes escolheram a opção correta. Observa-se pelo percentual de marcações que a alternativa incorreta (C) foi escolhida por 35,6% dos estudantes e que quase todos os outros distratores tiveram um número maior de marcações do que o próprio gabarito. Analisando-se o coeficiente bisserial, a opção incorreta (C) teve valor positivo, revelando-se completamente inadequada, pois sugere que alunos com habilidade elevada optaram por essa alternativa mais do que os alunos com baixa habilidade. Isso é reflexo do pouco poder de discriminação no item, 0,113, fato também evidenciado na AGI quando curvas de alternativas incorretas não decrescem com o aumento da habilidade.

O item foi descartado para o processamento da TRI, pois apresentou discriminação e bisserial negativos na primeira fase da rodada da teoria clássica pelo software *BLOG-MG*, sendo -0,005 e -0,077, respectivamente, prejudicando e impossibilitando a convergência dos dados.

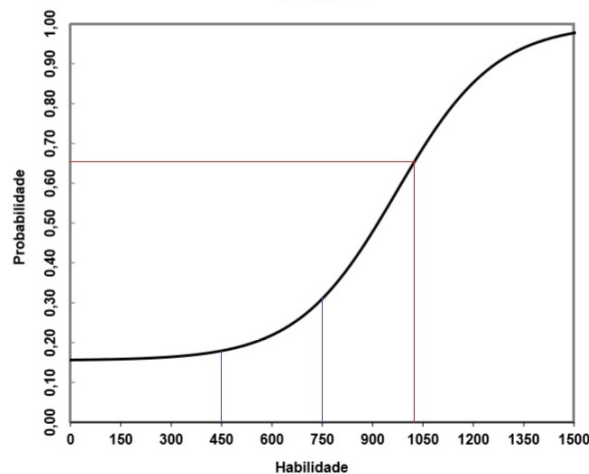
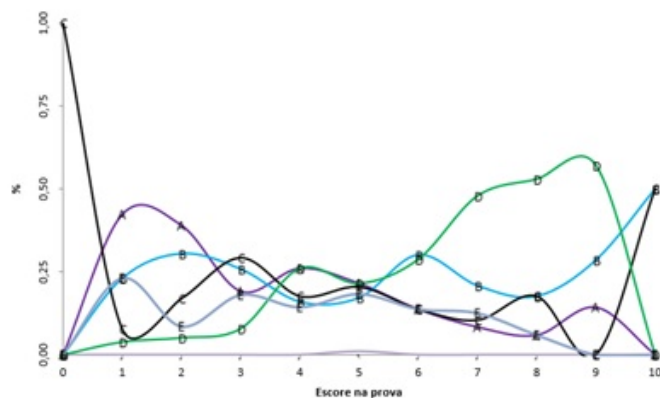
Questão 6

Cinco meninas não estão totalmente de acordo sobre a data da prova de Matemática.

- Andrea diz que será em agosto, dia 16, segunda-feira;
- Daniela diz que será em agosto, dia 16, terça-feira;
- Fernanda diz que será em setembro, dia 17, terça-feira;
- Patrícia diz que será em agosto, dia 17, segunda-feira;
- Tatiane diz que será em setembro, dia 17, segunda-feira.

Somente uma está certa, e as outras acertaram pelo menos uma das informações: o mês, o dia do mês ou o dia da semana. Quem está certa?

- a. Andrea
- b. Daniela
- c. Fernanda
- d. Patrícia
- e. Tatiane



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	D	0,223	0,325	0,422		
	<u>p_A</u>	<u>p_B</u>	<u>p_C</u>	<u>p_D</u>	<u>p_E</u>	<u>p_.</u>
	0,221	0,225	0,184	0,223	0,146	0,002
	<u>Bis_A</u>	<u>Bis_B</u>	<u>Bis_C</u>	<u>Bis_D</u>	<u>Bis_E</u>	<u>Bis_.</u>
	-0,277	-0,020	-0,085	0,422	-0,069	0,103
TRI	<u>a</u>	<u>b</u>	<u>c</u>			
	0,003981	970,659876	0,15527			

Item com grau de dificuldade elevado e bom poder de discriminação. A proporção de marcação foi bem distribuída entre as cinco alternativas. A opção correta, alternativa (D), obteve 22,3% das marcações, número muito próximo à quantidade de marcações das alternativas (A) e (B). O poder de discriminação do item, 0,325, sugere que o item deva ser aprimorado.

Observa-se na AGI que nenhum estudante com a habilidade máxima obtida respondeu corretamente o item, optando pelas alternativas (C) e (B). Com exceção do decréscimo no nível de habilidade mais extremo da AGI, a linha correspondente ao gabarito se comportou como o esperado, crescendo junto com o aumento do nível de habilidade.

Nota-se na análise via TCT que o coeficiente bisserial de duplas marcações ou que deixaram o item em branco foi positivo. Neste caso, os alunos de bom desempenho na prova de modo geral erraram na hora de marcar a resposta final, se esqueceram de responder o item por falta de atenção ou simplesmente não sabiam como responder o item. Sugere-se

assim uma investigação mais detalhada do item. O gabarito também apresentou bisse-rial positivo, enquanto as alternativas incorretas tiveram bisseriais negativos conforme esperado em uma avaliação educacional.

Pela TRI, o item apresentou nível de dificuldade superior a 970, indicando que apenas estudantes com proficiência igual ou superior a esse valor tem probabilidade alta de acertar o item. Embora seja um item muito difícil, o parâmetro acerto ao acaso igual a 0,15527 ficou abaixo do máximo permitido, revelando que não houve atratividade de alunos com baixo desempenho para a resposta correta.

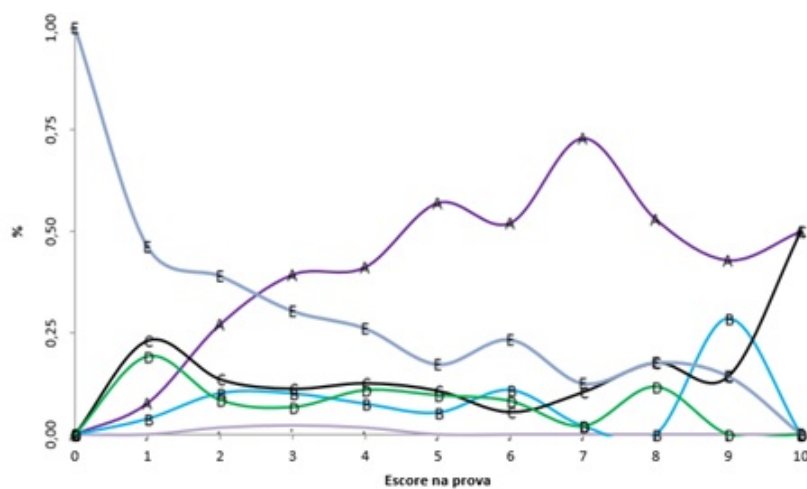
Como pode ser visto na CCI, o item é classificado como item âncora de um nível próximo de 1000, portanto, como não existem alunos da amostra com essa habilidade, infere-se que a habilidade avaliada no item não foi compreendida pelos estudantes.

Em relação ao poder de discriminação no item, 0,003981, observa-se na CCI uma leve inclinação na sigmóide, característico de um item com baixa discriminação.

Questão 7

Rodrigo comprou três cadernos iguais em uma promoção na qual o segundo e o terceiro cadernos eram vendidos, respectivamente, com 20% e 40% de desconto sobre o preço do primeiro. No dia seguinte, terminada a promoção, Gustavo comprou três cadernos iguais aos de Rodrigo, todos sem desconto. Percentualmente, quanto Rodrigo pagou a menos que Gustavo?

- a. 20%
- b. 22%
- c. 25%
- d. 28%
- e. 30%



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	A	0,452	0,280	0,315		
	<u>p_A</u>	<u>p_B</u>	<u>p_C</u>	<u>p_D</u>	<u>p_E</u>	<u>p_.</u>
	0,452	0,077	0,118	0,088	0,257	0,008
	<u>Bis_A</u>	<u>Bis_B</u>	<u>Bis_C</u>	<u>Bis_D</u>	<u>Bis_E</u>	<u>Bis_.</u>
	0,315	-0,043	-0,066	-0,111	-0,258	-0,224

Pela TCT, o item apresentou dificuldade moderada, com 45,2% de acerto, e poder de discriminação igual a 0,280, sendo classificado como um item marginal, sujeito à reelaboração. Pela proporção de marcação, infere-se que a maior parte dos estudantes optou apenas por duas alternativas - (A) e (E), os outros distratores quase não foram escolhidos, fato que indica ser necessário uma análise de plausibilidade dos mesmos.

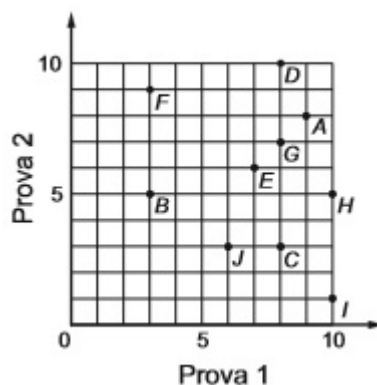
Uma possível justificativa para a grande quantidade de marcações da alternativa incorreta (E) talvez seja o fato de a opção apresentar como resposta a média aritmética entre os valores indicados no enunciado.

Em relação ao bisserial das alternativas, apenas a alternativa correta (A) obteve valor positivo, o que de fato é esperado em avaliações educacionais, pois indica que estudantes com as habilidades mais altas optaram por essa alternativa mais do que estudantes com baixa proficiência.

Por ter apresentado coeficiente bisserial e poder de discriminação negativos (-0,024 e -0,019, respectivamente), no primeiro processamento do *BILOG-MG* e, assim, dificultar a convergência dos dados, o item foi descartado, não sendo possível analisá-lo pela TRI.

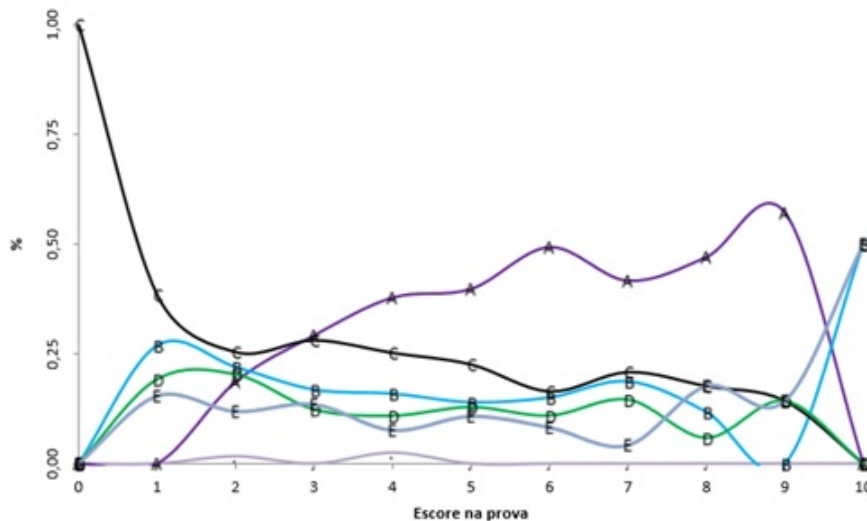
Questão 8

O professor Michel aplicou duas provas a seus dez alunos e divulgou as notas por meio do gráfico mostrado abaixo. Por exemplo, o aluno A obteve notas 9 e 8 nas provas 1 e 2, respectivamente; já o aluno B obteve notas 3 e 5. Para um aluno ser aprovado, a média aritmética de suas notas deve ser igual a 6 ou maior do que 6. Quantos alunos foram aprovados?



- a. 6
- b. 7

- c. 8
- d. 9
- e. 10



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	A	0,350	0,251	0,273		
	<u>p_A</u>	<u>p_B</u>	<u>p_C</u>	<u>p_D</u>	<u>p_E</u>	<u>p_.</u>
	0,350	0,169	0,240	0,131	0,103	0,007
	<u>Bis_A</u>	<u>Bis_B</u>	<u>Bis_C</u>	<u>Bis_D</u>	<u>Bis_E</u>	<u>Bis_.</u>
	0,273	-0,085	-0,153	-0,092	-0,048	-0,174

Segundo os resultados apresentados pela TCT, o item se revelou com grau de dificuldade mediano, já que 35% dos estudantes o responderam corretamente. Em relação à discriminação, o valor 0,251 corresponde a um item marginal, sujeito à reelaboração.

Em relação ao percentual de marcação, nota-se um equilíbrio entre os valores apresentados nos quatro distratores, porém a alternativa incorreta (C) foi a mais escolhida entre os estudantes que erraram o item, com 24% de marcações.

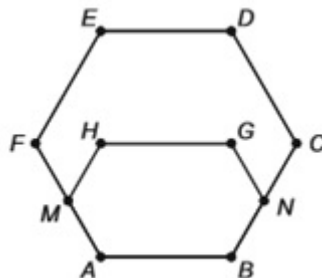
É interessante extrair da AGI que alunos com melhores desempenhos no teste erraram o item, e escolheram as alternativas (B) e (E) como opções de resposta.

Embora o item não tenha apresentado discriminação boa, os valores do parâmetro bisserial do item se comportaram conforme o esperado, ou seja, os distratores apresentaram valor negativo e o gabarito valor positivo.

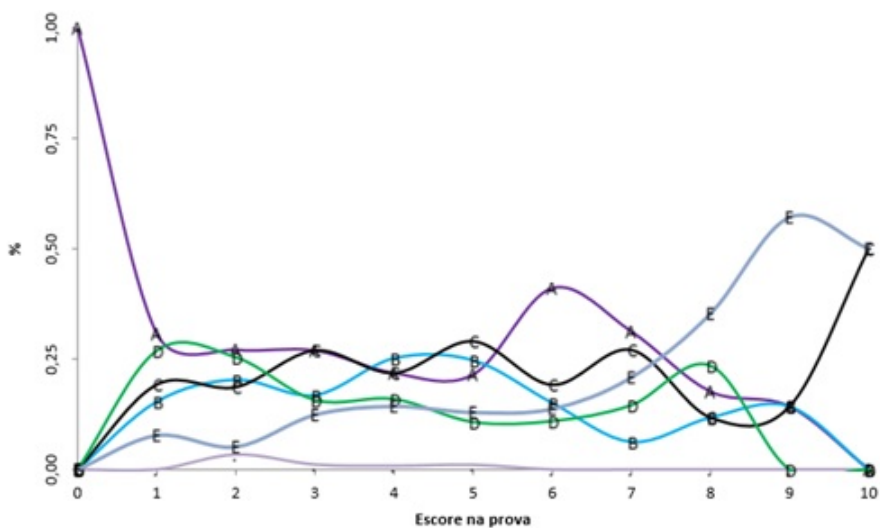
O item foi descartado para o processamento da TRI, pois apresentou poder discriminação e coeficiente bisserial negativos na primeira fase da rodada da teoria clássica pelo software *BILOG-MG*, ou seja, -0,048 e -0,062, respectivamente, prejudicando e impossibilitando a convergência dos dados.

Questão 9

O polígono ABCDEF é um hexágono regular. Os pontos M e N são pontos médios dos lados AF e BC, respectivamente. O hexágono ABNGHM é simétrico em relação à reta que passa por M e N. Qual é a razão entre as áreas dos hexágonos ABNGHM e ABCDEF?



- a. $\frac{3}{10}$
- b. $\frac{4}{11}$
- c. $\frac{3}{7}$
- d. $\frac{7}{15}$
- e. $\frac{5}{12}$



TCT	Gabarito	Dificuldade	Discriminação	Bisserial
	E	0,142	0,119	0,266

p_A	p_B	p_C	p_D	p_E	p_.
0,270	0,189	0,232	0,157	0,142	0,009

Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
0,005	-0,088	0,011	-0,147	0,266	-0,240

Pela TCT, o item se revelou difícil para os respondentes, com apenas 14,2% de acerto. Observa-se que o gabarito, opção (E), foi a alternativa menos escolhida pelos estudantes. Os distratores (A) e (C) foram as opções de resposta mais escolhidas, com 27% e 23,2% de marcações, respectivamente.

As proporções de marcação apresentadas sugerem um problema sério que pode comprometer a utilização do item em um teste de avaliação.

A discriminação do item, 0,119, está na faixa deficiente e, portanto, o item é passível de descarte.

O item apresentou dois distratores, (A) e (C), com coeficiente bisserial positivo, indicando que atraiu um percentual maior de estudantes de bom desempenho do que de baixo desempenho. Neste caso, é interessante rever as alternativas afim de detectar se as opções estão funcionando como “peguinhas” para os bons estudantes. Caso isso aconteça, sugere-se que as opções sejam reelaboradas.

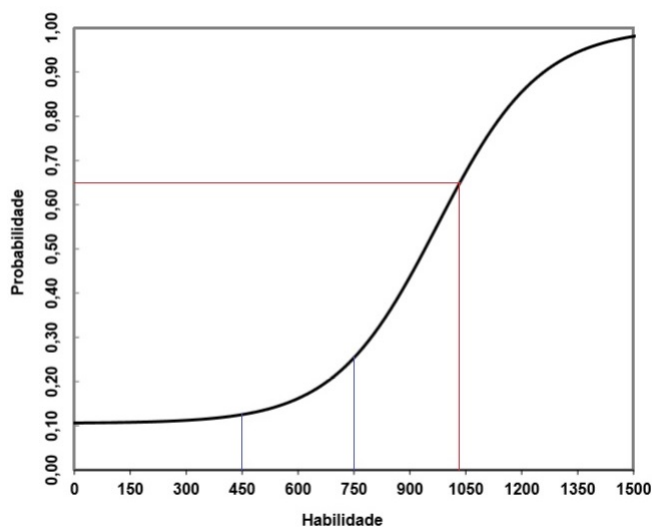
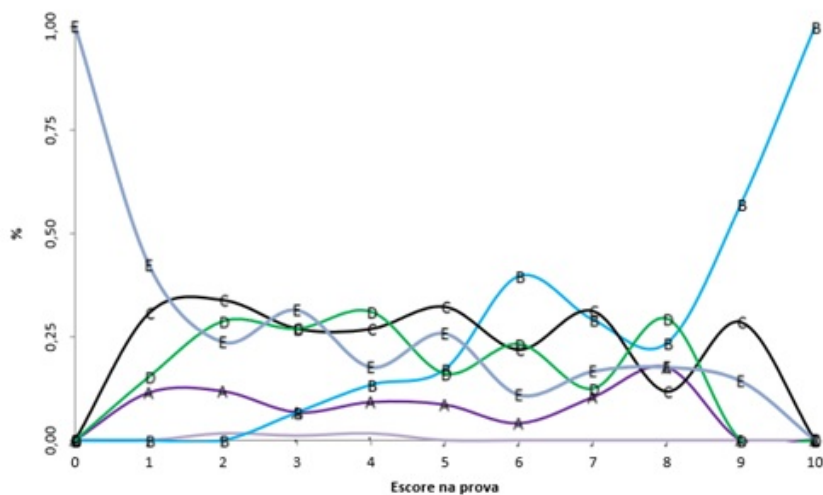
O item foi descartado para o processamento da TRI, pois não se comportou bem do ponto de vista dos pressupostos da teoria, apresentando discriminação e coeficiente bisserial negativos na primeira rodada da teoria clássica pelo *BLOG-MG*.

Questão 10

Sempre que Yurika abastece seu carro, ela enche o tanque e anota a data, a quilometragem marcada no painel e a quantidade de litros de combustível colocada. Na tabela estão os dados registrados por Yurika em dois abastecimentos consecutivos. Quantos quilômetros por litro, aproximadamente, fez o carro de Yurika nesse período?

data	km	litros
⋮	⋮	⋮
01/02	35 723	32,5
07/02	36 144	43,0
⋮	⋮	⋮

- a. 5,6
- b. 9,8
- c. 11,1
- d. 12,9
- e. 40,1



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	B	0,170	0,326	0,489		
	p_A	p_B	p_C	p_D	p_E	p_.
	0,084	0,170	0,279	0,234	0,223	0,007
	Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
	0,024	0,489	-0,064	-0,105	-0,199	-0,223
TRI	a	b	c			
	0,004258	937,1023029	0,10565			

Pela TCT, o item apresentou grau de dificuldade alto, com apenas 17% de acerto. O gabarito foi uma das alternativas menos escolhidas pelos respondentes, superando apenas a opção incorreta (A). Os distratores (C), (D) e (E) foram as alternativas que apresentaram maior quantidade de marcações.

A discriminação do item, 0,326, está na faixa boa, porém, passível de aprimoramento. O bom poder discriminativo do item pode ser observado na AGI, quando a linha correspondente ao gabarito (B) cresce com aumento da habilidade.

O distrator (A) apresentou coeficiente bisserial positivo, indicando que uma quantidade maior de alunos que foi bem na prova escolheu essa opção de resposta do que a

quantidade de alunos com baixo desempenho. Fato que é esperado que aconteça apenas com o gabarito.

Pela TRI, o item mostrou-se difícilimo e com baixa discriminação. Nota-se na CCI que a inclinação da curva está deslocada para a direita, e que todos os estudantes da amostra tiveram probabilidade inferior a 30% de acertar o item. Infere-se dos resultados apresentados que, para ter uma elevada probabilidade de acerto neste item, o estudante deve ter proficiência igual ou superior a 970. Pela CCI, observa-se que o item classificou-se como item âncora de um nível superior a 1000. De fato, vale lembrar que nenhum estudante do presente estudo obteve tal proficiência, sugerindo que não houve domínio do conhecimento avaliado pelo item.

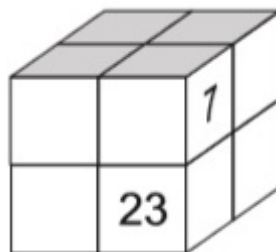
Apesar da alta dificuldade do item, o parâmetro de acerto ao acaso 0,10565 revela que estudantes de baixo desempenho não foram atraídos pela resposta correta.

Questão 11

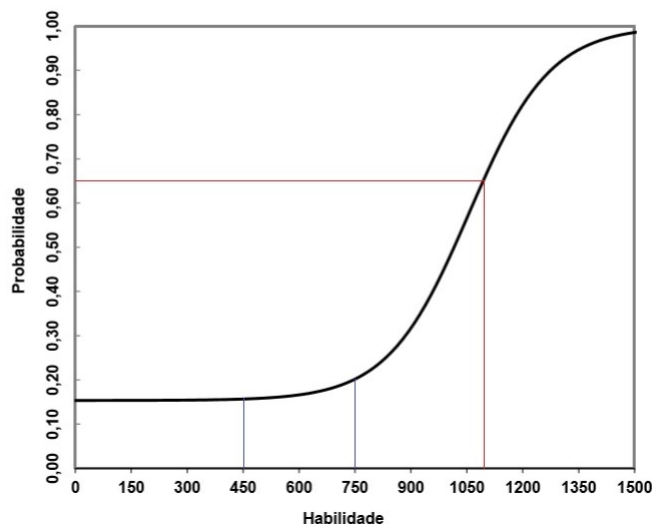
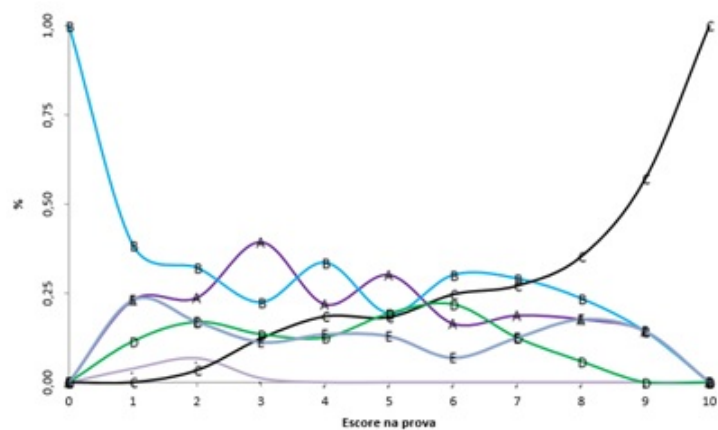
Todos os números de 1 a 24 devem ser escritos nas faces de um cubo, obedecendo-se às seguintes regras:

- em cada face devem ser escritos quatro números consecutivos;
- em cada par de faces opostas, a soma do maior número de uma com o menor número da outra deve ser igual a 25.

Se os números 7 e 23 estiverem escritos no cubo como na figura, qual é o menor número que pode ser escrito na face destacada em cinza?



- a. 1
- b. 5
- c. 9
- d. 11
- e. 17



TCT	Gabarito	Dificuldade	Discriminação			Bisserial
	C	0,178	0,218	0,372		
	p_A	p_B	p_C	p_D	p_E	p_.
	0,251	0,279	0,178	0,152	0,129	0,011
	Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
	-0,119	-0,076	0,372	-0,003	-0,086	-0,492
	TRI			a	b	c
				0,005395	1055,728521	0,15347

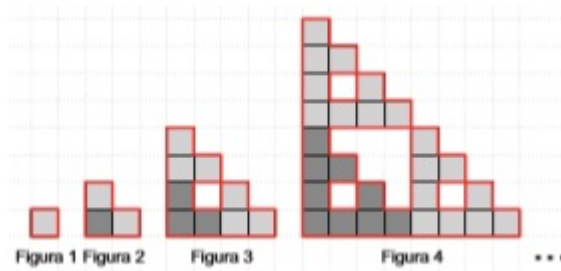
Pela TCT, a análise dos percentuais de respostas às alternativas revela que apenas 17,8% dos estudantes escolheram a opção correta (C), sugerindo que o item se revelou difícil para eles. Um quantitativo elevado de estudantes, 25,1% e 27,9%, preferiram as alternativas incorretas (A) e (B), respectivamente. Pela AGI é possível perceber que esses dois distratores atraíram a maioria dos estudantes com escore inferior a 7 pontos. Os alunos com melhor desempenho na prova acertaram o item, embora o poder de discriminação do item, 0,218, estar na faixa marginal. Conforme o esperado, os bisseriais de todos os distratores foram negativos, indicando que alunos com menor desempenho foram mais atraídos para essas alternativas do que estudantes com maior desempenho.

Pela TRI, o item revelou-se muito difícil, com parâmetro de dificuldade igual a 1055,728521 e com baixo poder de discriminação, 0,005395. Nota-se, no intervalo de proficiência onde

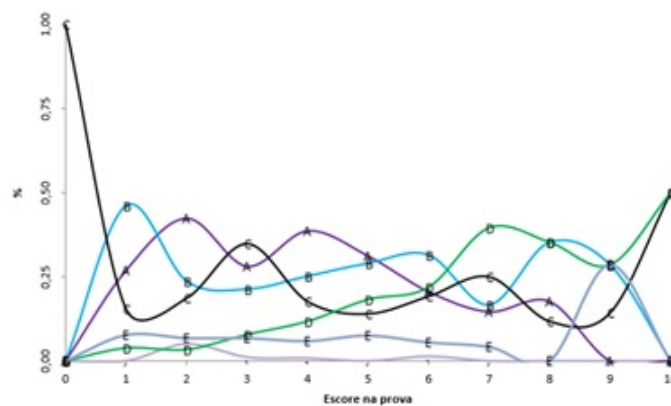
localizam-se os estudantes da amostra, que a probabilidade de acertar o item foi praticamente a mesma. Assim, não foi possível perceber o formato de sigmóide desejado para um bom item. Caracterizou-se como item âncora de um nível extremamente alto de proficiência no qual não se localizam alunos neste estudo, sugerindo que entre os alunos pesquisados nenhum demonstrou ter domínio do conhecimento necessário para a compreensão e resolução do item.

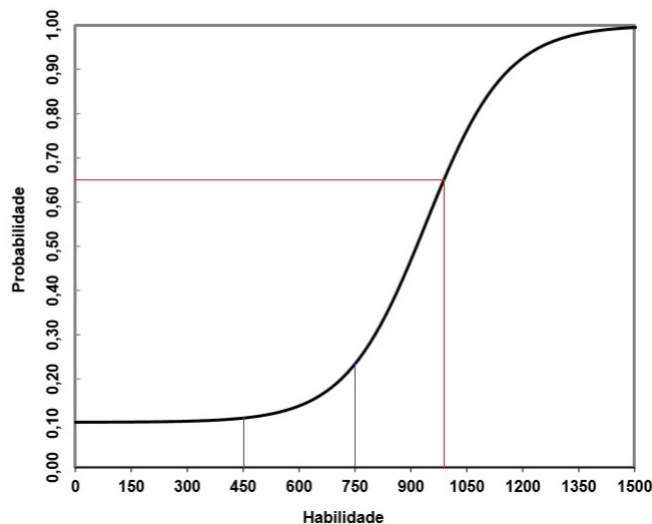
Questão 12

Começando com um quadrado de 1 cm de lado, formamos uma sequência de figuras, como na ilustração. Cada figura, a partir da segunda, é formada unindo-se três cópias da anterior. Os contornos destacados em vermelho das quatro primeiras figuras medem, respectivamente, 4 cm, 8 cm, 20 cm e 56 cm. Quanto mede o contorno da Figura 6?



- a. 88 cm
- b. 164 cm
- c. 172 cm
- d. 488 cm
- e. 492 cm





TCT	Gabarito	Dificuldade	Discriminação	Bisserial
	D	0,159	0,242	0,412

p_A	p_B	p_C	p_D	p_E	p_.
0,294	0,264	0,208	0,159	0,064	0,011

Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
-0,192	-0,022	-0,058	0,412	-0,019	-0,252

TRI	a	b	c
	0,005445	939,9269265	0,10197

Pela TCT, o item revelou-se difícil, com apenas 15,9% de acerto. Muitos estudantes, 29,4%, 26,4% e 20,8%, escolheram inadequadamente as opções (A), (B) e (C), respectivamente. A alternativa (E) foi preferência de apenas 6,4% dos estudantes, fato que sugere uma investigação sobre a plausibilidade desse distrator.

O poder de discriminação do item, 0,242, pertence ao intervalo característico de itens marginais, sujeitos à reelaboração. A AGI apresentou-se confusa em sua totalidade. 50% dos estudantes de melhor desempenho na prova escolheram a alternativa incorreta (C) como opção de resposta e os outros 50% acertaram o item.

Em relação ao coeficiente bisserial das alternativas, o gabarito apresentou valor positivo, enquanto as demais opções apresentaram valor negativo, conforme esperado.

Os parâmetros obtidos utilizando-se a TRI revelaram que o item teve discriminação baixa e altíssima dificuldade, que podem ser observados na forma da curva apresentada na CCI. Nota-se que a inclinação da curva está deslocada à direita e que todos os estudantes da amostra tiveram entre 10 e 25% de chance de acertar o item. O parâmetro acerto ao acaso ficou abaixo do permitido, revelando que alunos com baixa proficiência não foram atraídos para a alternativa correta.

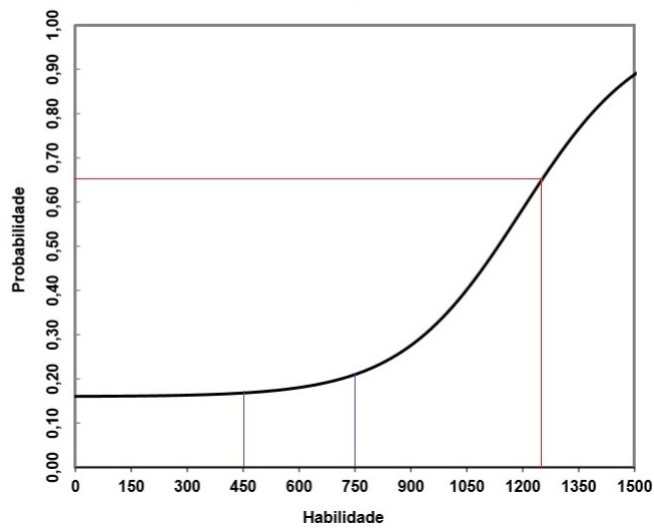
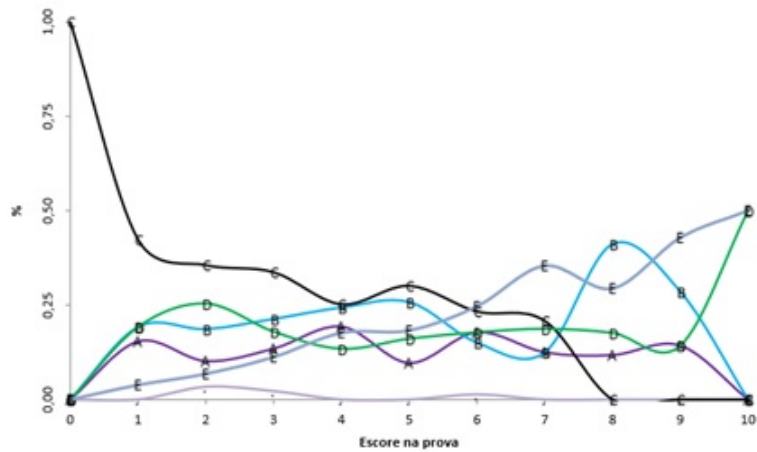
A partir da análise de desempenho, verificou-se que o item estaria inserido em um nível próximo de 1000 da escala de proficiência. Como no presente estudo a proficiência máxima obtida foi 748,69, portanto, infere-se que nenhum estudante da amostra desenvolveu plenamente a habilidade avaliada no item.

Questão 13

Na conta indicada a seguir, as letras X, Y e Z representam algarismos distintos. Qual é o algarismo representado pela letra Z?

$$\begin{array}{r}
 XXXX \\
 YYY \\
 + ZZZ \\
 \hline
 YXXXX
 \end{array}$$

- a. 1
- b. 3
- c. 5
- d. 6
- e. 8



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	E	0,182	0,214	0,330		
	<u>p_A</u>	<u>p_B</u>	<u>p_C</u>	<u>p_D</u>	<u>p_E</u>	<u>p_.</u>
	0,142	0,213	0,277	0,176	0,182	0,009
	<u>Bis_A</u>	<u>Bis_B</u>	<u>Bis_C</u>	<u>Bis_D</u>	<u>Bis_E</u>	<u>Bis_.</u>
	0,001	0,005	-0,226	-0,027	0,330	-0,240

TRI	a	b	c
	0,003633	1196,581045	0,15988

A análise do resultado, pela TCT, mostra que o item revelou-se difícil para os respondentes, já que apenas 18,2% foram capazes de respondê-lo corretamente. A opção (C) foi a preferência de um quantitativo elevado de estudantes, 27,7%, porém, observa-se na AGI que os que optaram por essa alternativa tiveram baixo desempenho na prova como um todo. Já os distratores (A) e (B) atraíram muitos estudantes de bom desempenho pois apresentaram bisseriais positivos. O item, de maneira geral, apresentou discriminação marginal, sendo passível de reelaboração.

Pela TRI, o item apresentou dificuldade elevadíssima, ($b = 1196,581$), e baixa discriminação, ($a = 0,003633$). O formato de sigmóide esperado para bons itens quase não pode ser observado. No intervalo de habilidade (450, 750), a curva se aproximou de uma reta horizontal, indicando que todos os estudantes da amostra tiveram praticamente a mesma probabilidade de acertar o item, inferior a 20%. O parâmetro acerto ao acaso ficou dentro do limite permitido, sugerindo que o gabarito não atraiu alunos com baixo desempenho. Observa-se na CCI que o item é característico de um nível de proficiência elevadíssimo, superior a 1000. Neste estudo, como nenhum aluno obteve esse nível de proficiência, o comportamento do item sugere que o conteúdo não foi compreendido pelos estudantes.

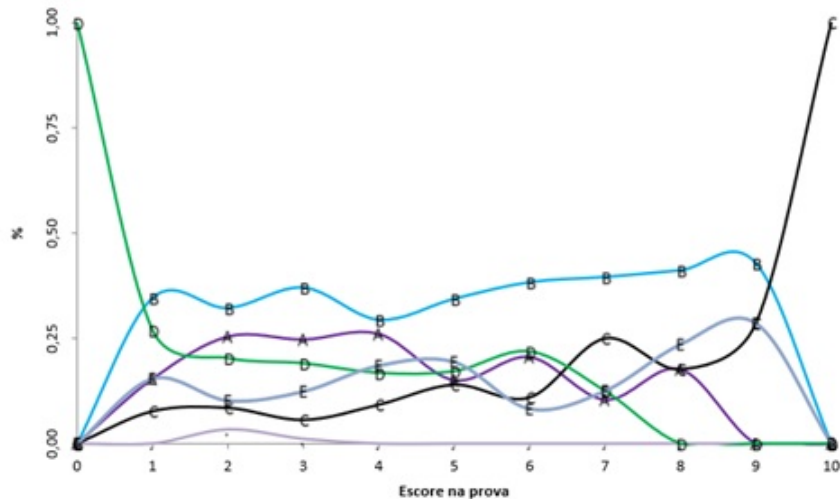
Questão 14

Rosane percebeu que seu antigo relógio de parede tinha parado às 9 horas. Ela deu corda no relógio, colocando-o para funcionar sem acertar o horário, e foi imediatamente ao mercado. Chegou ao mercado às 10 horas e 10 minutos. Fez suas compras em 1 hora e voltou para casa. Entrando em casa, notou que o relógio de parede marcava 10 horas e 40 minutos. Se Rosane realizou os percursos de ida e volta ao mercado em tempos iguais, a que horas ela entrou em casa?



- a. 10 horas e 50 minutos

- b. 11 horas e 10 minutos
- c. 11 horas e 30 minutos
- d. 11 horas e 40 minutos
- e. 11 horas e 50 minutos



TCT	Gabarito	Dificuldade	Discriminação	Bisserial
	C	0,118	0,115	0,273

p_A	p_B	p_C	p_D	p_E	p_.
0,204	0,346	0,118	0,178	0,148	0,006

Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
-0,127	0,048	0,273	-0,144	0,035	-0,389

Pela TCT, o item apresentou dificuldade elevadíssima, já que apenas 11,8% dos estudantes responderam corretamente o item. Pelo poder de discriminação 0,115, o item revelou-se como deficiente, sujeito à rejeição. Os coeficientes bisseriais dos distratores (B) e (E) foram positivos, indicando que muitos estudantes de bom desempenho foram atraídos para essas opções incorretas. Nestes casos, é interessante reanalisar os distratores a fim de evitar “peguinhas”.

Observa-se na AGI que, em todas as faixas de habilidades, exceto a proficiência 10, a alternativa (B) foi a preferida pelos estudantes, sendo escolhida por um total de 34,6% dos respondentes. Apesar disso, todos os estudantes com melhor desempenho no teste responderam corretamente o item, optando pelo gabarito (C).

O item foi descartado para o processamento da TRI, pois não se comportou bem do ponto de vista dos pressupostos da teoria, apresentando discriminação e coeficiente bisserial negativos na primeira rodada da teoria clássica pelo *BILOG-MG*.

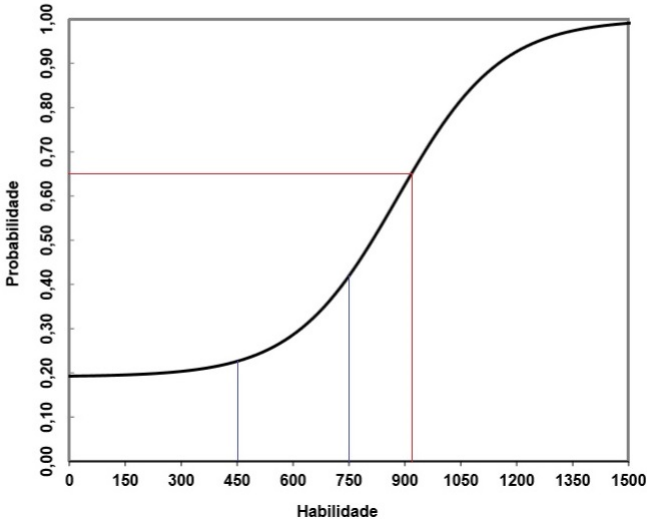
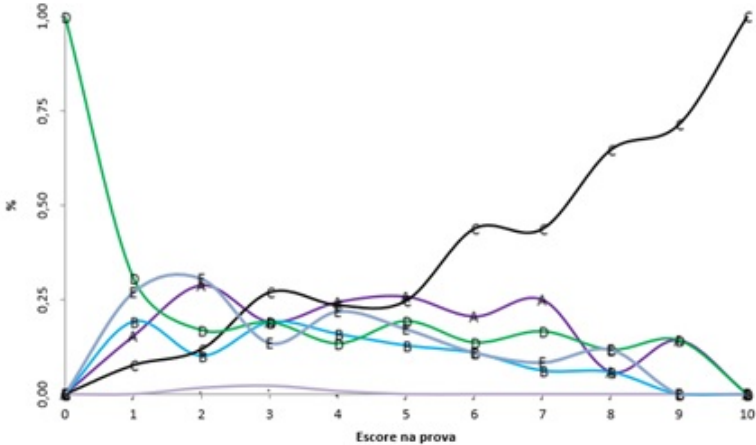
Questão 15

Télio comprou laranjas, maçãs e uvas no mercado. O preço por quilograma de cada fruta está na tabela abaixo. Metade do peso total da compra era de maçãs e o peso das

uvas era o dobro do peso das laranjas. Se Télió gastou R\$ 38,00, quantos quilogramas de frutas ele comprou?

Preços (R\$) por quilograma	
Maçã	3,00
Uva	4,00
Laranja	2,00

- a. 10
- b. 11
- c. 12
- d. 13
- e. 14



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	C	0,290	0,294	0,343		
	p_A	p_B	p_C	p_D	p_E	p_.
	0,225	0,133	0,290	0,170	0,174	0,007
	Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
	-0,037	-0,145	0,373	-0,102	-0,210	-0,272
	TRI		a	b	c	
			0,004236	879,6523472	0,19127	

Pela TCT, o item revelou-se difícil já que apenas 29% dos estudantes pesquisados responderam corretamente à questão proposta. Observa-se que o poder de discriminação do item, 0,294, caracteriza o item como marginal, sujeito à reelaboração.

Como nenhuma das opções incorretas apresentou coeficiente bisserial positivo, infere-se que os alunos que erraram o item foram, em sua maioria, os de menor desempenho na prova como um todo. Esse fato pode ser observado na AGI, já que a curva referente ao gabarito cresce com o aumento da proficiência, diferentemente do que acontece com os distratores.

O percentual de marcação do item foi consideravelmente equilibrado entre as cinco opções de respostas, sendo o gabarito o preferido dos estudantes e a alternativa (B) a menos escolhida.

Pela TRI, o item apresentou-se muito difícil, indicando que para ter probabilidade alta de acerto ao item, o estudante deve ter proficiência mínima de 879,6523471.

Observa-se, na CCI, que alunos com proficiência inferior a 600 tem praticamente a mesma probabilidade de acerto, aproximadamente 20%. Já entre os alunos pesquisados, as proficiências mais altas estão próximas ao valor 750, com probabilidade de acerto em torno de 40%. Infere-se então que nem mesmo os alunos com proficiências mais elevadas têm probabilidade alta de acerto, fato característico de um item com baixo poder de discriminação.

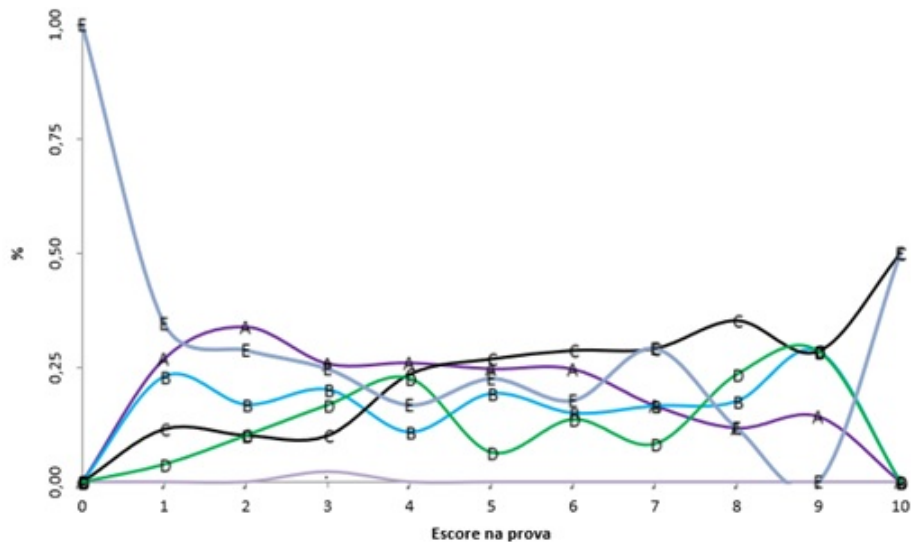
Pela CCI, é possível perceber que o item é característico de um nível superior a 900 na escala de proficiência. Entre os alunos pesquisados, nenhum atingiu esse nível de habilidade, o que sugere, portanto, que o item não cumpriu seu propósito básico de uma avaliação educacional, ou seja, servir para interpretar o significado do escore e para comparar o escore do estudante com os demais indivíduos do grupo de respondentes.

Questão 16

A mãe de Lúcia pediu para ela não comer mais de 10 docinhos por dia. Além disso, se em um dia ela comer mais de 7 docinhos, nos dois dias seguintes não poderá comer mais de 5 docinhos em cada dia. Qual é o maior número de docinhos que Lúcia pode comer durante um período de 29 dias seguidos, obedecendo ao pedido de sua mãe?

a. 203

- b. 204
- c. 206
- d. 213
- e. 290



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	C	0,215	0,196	0,256		
	<u>p_A</u>	<u>p_B</u>	<u>p_C</u>	<u>p_D</u>	<u>p_E</u>	<u>p_.</u>
	0,249	0,167	0,215	0,140	0,225	0,004
	<u>Bis_A</u>	<u>Bis_B</u>	<u>Bis_C</u>	<u>Bis_D</u>	<u>Bis_E</u>	<u>Bis_.</u>
	-0,128	-0,020	0,256	0,021	-0,105	-0,252

Os resultados apresentados pela TCT sugerem que os estudantes sentiram dificuldade ao responder o item, já que apenas 21,5% optaram pelo gabarito, a alternativa (C).

Em relação à proporção de marcações, verifica-se que os estudantes ficaram divididos entre as alternativas (A) e (E), com 24,9% e 22,5% de marcações, respectivamente. Nota-se que esses distratores são opções plausíveis, tendo em vista que a escolha pela alternativa (A) significa que o estudante realizou o cálculo $7 \times 29 = 203$ e, para marcar a alternativa (E), o estudante provavelmente não interpretou corretamente o problema, e apenas calculou $10 \times 29 = 290$. Observa-se, na AGI, que 50% dos estudantes com desempenho mais elevado no teste também cometeram esse erro, optando inadequadamente pela alternativa (E).

Apesar da baixa quantidade de marcações, apenas 14%, o distrator (D) atraiu um número maior de alunos de bom desempenho do que alunos com baixos escores, apresentando coeficiente bisserial positivo. Observa-se, na AGI, que esse distrator foi o mais escolhido por estudantes com nota 9, em uma escala de 0 a 20 pontos possíveis. Os demais distratores apresentaram coeficiente bisserial negativo.

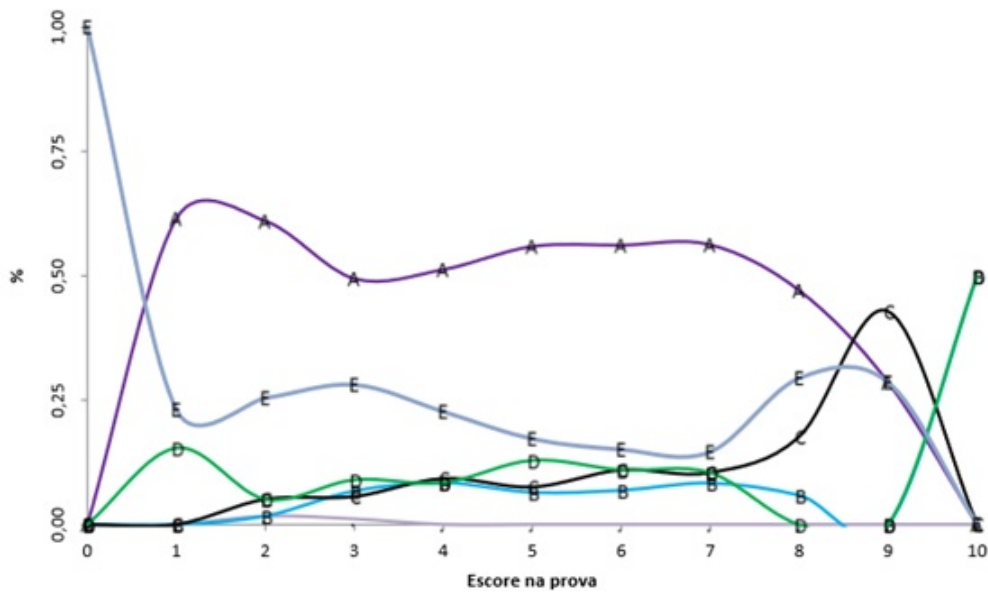
De maneira geral, pelo baixo poder de discriminação, o item revelou-se deficiente, sujeito à rejeição.

O item foi descartado para o processamento da TRI, pois não se comportou bem do ponto de vista dos pressupostos da teoria, apresentando discriminação e coeficiente bisserial negativos na primeira rodada da teoria clássica pelo *BILOG-MG*.

Questão 17

Gustavo possui certa quantidade de moedas de 1, 10, 25 e 50 centavos, tendo pelo menos uma de cada valor. É impossível combiná-las de modo a obter exatamente 1 real. Qual é o maior valor total possível para suas moedas?

- a. 86 centavos
- b. 1 real e 14 centavos
- c. 1 real e 19 centavos
- d. 1 real e 24 centavos
- e. 1 real e 79 centavos



TCT	Gabarito	Dificuldade	Discriminação	Bisserial
	C	0,084	0,084	0,238

p_A	p_B	p_C	p_D	p_E	p_.
0,537	0,064	0,084	0,096	0,215	0,004

Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
-0,049	0,135	0,238	0,018	-0,114	-0,343

Segundo os resultados apresentados pela TCT, entre todos os 20 itens da prova, esse foi o item com o grau de dificuldade mais elevado, uma vez que apenas 8,4% dos estudantes

responderam corretamente. Observa-se pela proporção de marcações que a alternativa (A) foi escolhida por um quantitativo bem elevado de estudantes. Provavelmente, os alunos foram atraídos para esse distrator devido à falha nos aspectos formais de elaboração, afinal, as opções devem ser construídas com extensão e estruturas semelhantes e de forma relativamente homogênea. Nota-se que a alternativa (A) é a única que tem sua estrutura diferenciada, formada apenas por uma quantidade inferior a um real, enquanto as outras alternativas são compostas de uma quantidade superior a um real. Além disso, o valor apresentado nessa alternativa é a soma dos valores encontrados no enunciado. Desse modo, torna-se uma opção muito atrativa para os estudantes que não compreenderam a situação-problema proposta.

Em relação ao poder de discriminação, o item foi considerado o menos discriminativo da prova, e sujeito à rejeição. Observa-se pelos coeficientes bisseriais dos distratores, que as alternativas (B) e (D) apresentaram valores positivos, revelando que alunos com alto desempenho foram atraídos para essas respostas.

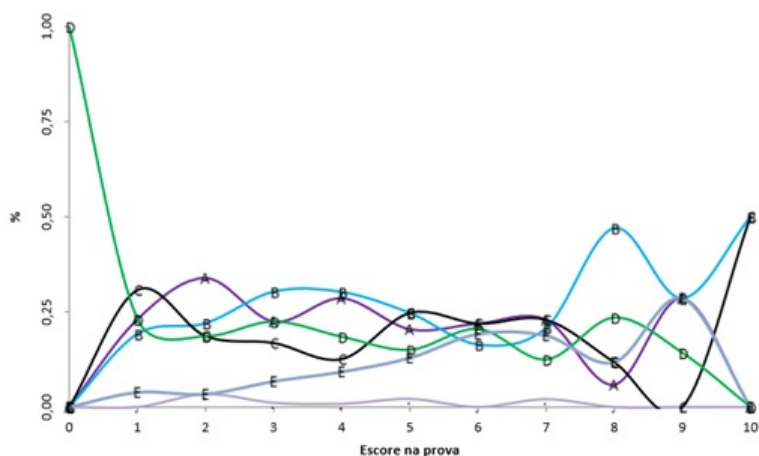
Pelos resultados expostos, infere-se que os estudantes pesquisados não compreenderam a situação-problema proposta. Além disso, os valores observados dos parâmetros sugerem que item não cumpriu sua finalidade e não deveria figurar em uma prova que objetiva avaliar estudantes.

Assim como ocorreu no item anterior, este item também foi descartado para o processamento da TRI pelos mesmos motivos citados no *Item 16*.

Questão 18

O número 2014 tem quatro algarismos distintos, um ímpar e três pares, sendo um deles 0. Quantos números possuem exatamente essas características?

- a. 60
- b. 180
- c. 360
- d. 420
- e. 540



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	E	0,110	0,132	0,262		
	<u>p_A</u>	<u>p_B</u>	<u>p_C</u>	<u>p_D</u>	<u>p_E</u>	<u>p_.</u>
	0,242	0,257	0,191	0,187	0,110	0,013
	<u>Bis_A</u>	<u>Bis_B</u>	<u>Bis_C</u>	<u>Bis_D</u>	<u>Bis_E</u>	<u>Bis_.</u>
	-0,099	0,011	0,004	-0,076	0,262	-0,082

Pela TCT, o item classifica-se como muito difícil e de baixa discriminação. Observa-se pelo percentual de marcações que apenas 11% dos estudantes chegaram à resposta correta, e que, apesar de correto, o gabarito (E) foi a alternativa menos atraente.

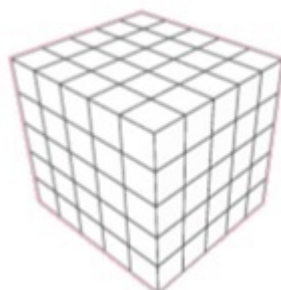
Os coeficientes bisseriais das alternativas incorretas (B) e (C) foram positivos, indicando que essas opções atraíram uma quantidade maior de estudantes de bom desempenho do que de baixo desempenho no teste como um todo.

Espera-se de um item que quanto maior o escore no teste, maior a proporção de acerto. Observando a AGI, nota-se que as curvas apresentam-se “emboladas”, e que em cada escore tem-se uma proporção de alunos bem distribuída entre as cinco opções de resposta. Nenhuma das curvas correspondente às opções da AGI tem um aumento significativo da probabilidade de acerto ao item quando se aumenta o desempenho na prova, portanto, infere-se que o item quase não discrimina. Portanto, infere-se que o item quase não discrimina, sendo considerado marginal, que deveria ser rejeitado.

Novamente tem-se um item descartado para que a TRI possa ser processada, pois não se comportou bem do ponto de vista dos pressupostos da teoria, apresentando discriminação e coeficiente bisserial negativos na primeira rodada da teoria clássica pelo *BILOG-MG*.

Questão 19

Um cubo de madeira foi pintado de vermelho e depois cortado em n^3 cubinhos iguais, $n > 2$. Alguns desses cubinhos ficaram sem nenhuma face pintada e outros com uma, duas ou três faces pintadas. Se o número de cubinhos sem nenhuma face pintada é igual ao número de cubinhos com exatamente uma face pintada, qual é o valor de n ?



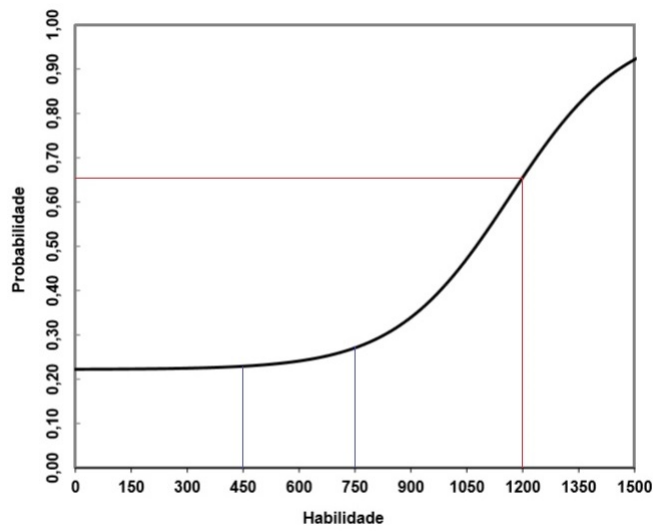
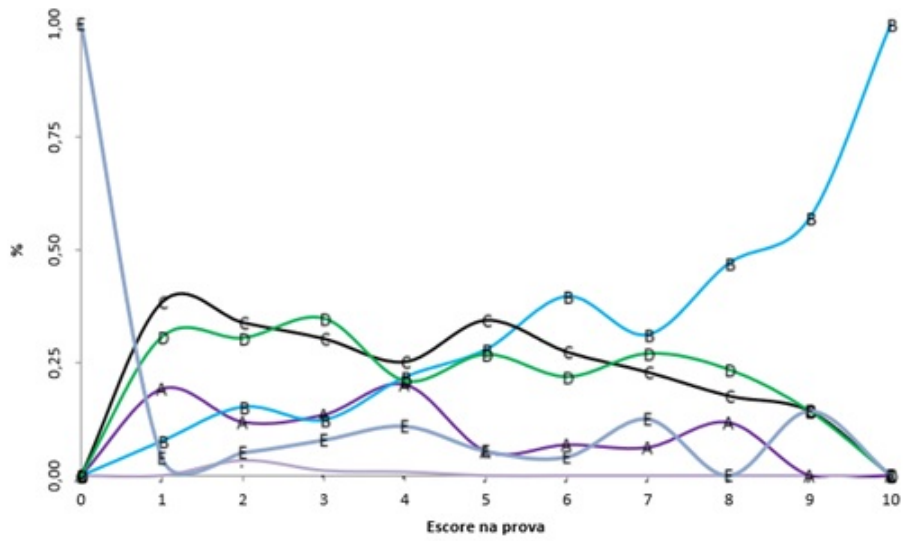
a. 7

b. 8

c. 9

d. 10

e. 11



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	B	0,247	0,269	0,341		
	p_A	p_B	p_C	p_D	p_E	p_.
	0,118	0,247	0,288	0,264	0,075	0,007

Bis_A	Bis_B	Bis_C	Bis_D	Bis_E	Bis_.
-0,172	0,341	-0,105	-0,090	-0,012	-0,321

TRI	a	b	c
	0,003834	1164,685562	0,22201

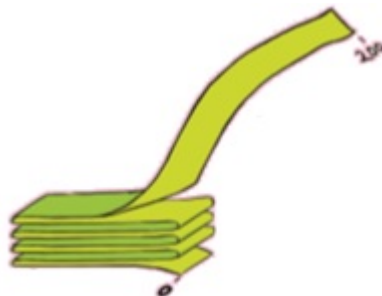
Pela TCT, o item revelou-se difícil, já que apenas 24,7% dos estudantes optaram pelo gabarito, a alternativa (B). Observa-se na proporção de marcações que a maioria dos estudantes optou pelos distratores (C) e (D). Apesar de terem obtido uma quantidade significativa de marcações, os coeficientes bisseriais desses distratores foram negativos, revelando que essas alternativas atraíram, em sua maioria, estudantes com baixo desempenho na prova.

Já o gabarito (B) apresentou coeficiente bisserial positivo, indicando que atraiu uma quantidade alta de alunos com bom desempenho na prova como um todo, conforme o que é desejado para uma opção correta. Apesar disso, em relação à discriminação, o item classificou-se como marginal, sujeito à reelaboração.

Na TRI, o item foi classificado como muito difícil, sugerindo que apenas alunos com proficiência superior a 1100 tem probabilidade razoável de acerto. Em relação à discriminação, observa-se que a leve inclinação da curva está deslocada para direita. Nota-se que um deslocamento significativo no eixo da habilidade não afeta a probabilidade de acerto do item. Neste estudo, estudantes menos habilidosos, com proficiência próxima a 450 pontos, tiveram praticamente a mesma probabilidade de acertar o item do que os melhores estudantes, com proficiências de aproximadamente 750 pontos. Portanto, infere-se que o item é bem pouco discriminativo.

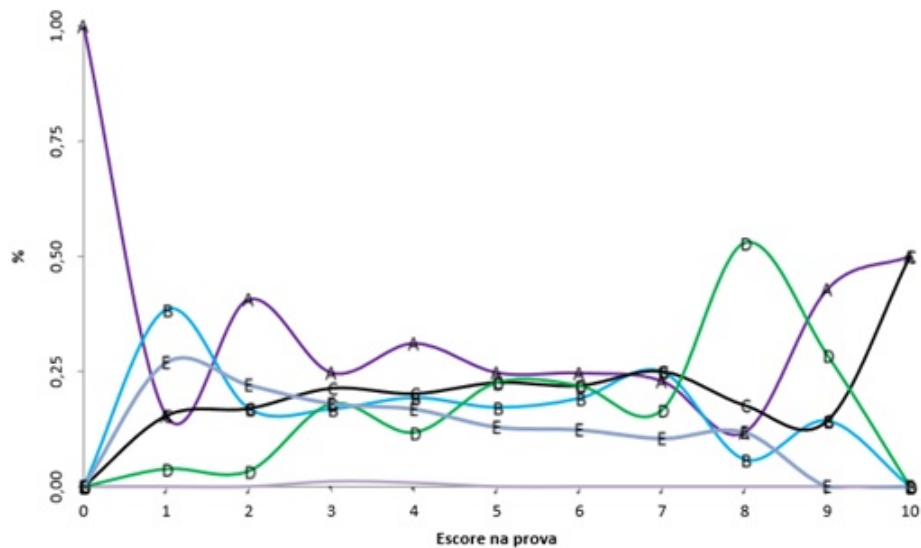
Questão 20

Rodrigo brinca com uma fita de dois metros, com marcas de centímetro em centímetro. Começando pela ponta de marca 0 cm, ele dobra a fita várias vezes em zigue-zague, como na figura, sobrepondo pedaços de fita de mesmo tamanho até dobrar um último pedaço, que pode ser menor do que os demais. Ele observa que as marcas de 49 cm e de 71 cm ficaram sobrepostas em pedaços vizinhos. Ele observa também que a marca de 139 cm ficou alinhada com elas. Com qual marca do penúltimo pedaço a ponta final da fita ficou sobreposta?



- a. 160 cm
- b. 176 cm

- c. 184 cm
- d. 190 cm
- e. 196 cm



TCT	Gabarito	Dificuldade	Discriminação	Bisserial		
	D	0,167	0,130	0,264		
	<u>p_A</u>	<u>p_B</u>	<u>p_C</u>	<u>p_D</u>	<u>p_E</u>	<u>p_.</u>
	0,273	0,191	0,208	0,167	0,157	0,004
	<u>Bis_A</u>	<u>Bis_B</u>	<u>Bis_C</u>	<u>Bis_D</u>	<u>Bis_E</u>	<u>Bis_.</u>
	-0,067	-0,059	0,062	0,264	-0,180	-0,162

Pelos resultados obtidos pela TCT, infere-se que os estudantes apresentaram muita dificuldade na compreensão e resolução do item, já que apenas 16,7% deles responderam corretamente.

Nota-se na proporção de marcações que a alternativa incorreta (A) foi a mais atraente, sendo escolhida por um quantitativo elevado de estudantes, gerando 27,3% das marcações, superando até mesmo a quantidade de marcações do gabarito, a letra (D). Nesses casos, sugere-se uma nova análise desse distrator, com finalidade de revelar conclusões importantes acerca do aprendizado dos estudantes ou até mesmo uma falha na elaboração da opção, como a presença de um possível “peguinha”.

O item quase não discriminou, sendo considerado deficiente e sujeito à rejeição. Observa-se esse baixo poder de discriminação na AGI quando a linha referente ao gabarito decresce após o escore 8. Diferentemente do esperado, todos os alunos com notas mais altas no teste erraram o item, marcando inadequadamente as alternativas (A) e (C).

Por apresentar coeficiente bisserial positivo, infere-se que o distrator (C) atraiu muitos alunos de bom desempenho na prova como um todo.

A análise desses resultados sugere que o aprendizado dos alunos está distante do desejado e que o item não está adequado para fins de avaliação educacional em larga escala.

O item foi descartado para o processamento da TRI, pois não se comportou bem do ponto de vista dos pressupostos da teoria, apresentando discriminação e coeficiente bisserial negativos na primeira rodada pela teoria clássica pelo *BILOG-MG*.

4.3 Observações Gerais

A análise psicométrica, quantitativa, de cada item feita anteriormente é repleta de informações valiosas acerca dos parâmetros usualmente associados a cada um deles, servindo de ponto de partida para a análise pedagógica, qualitativa, do comportamento dos estudantes frente aos desafios propostos nas situações-problema apresentadas. Por exemplo, a informação de que uma proporção bem maior de estudantes escolheu uma opção incorreta em vez da correspondente ao gabarito, leva à reflexão acerca das aprendizagens dos conceitos associados à solução do problema proposto e dos possíveis erros cometidos por eles.

No entanto, as inferências somente poderão ser feitas se os itens forem considerados bons, de acordo com parâmetros já assumidos pela literatura da área. A análise acima revela que, infelizmente, boa parte da prova da Primeira Fase da OBMEP de 2014 revelou-se inadequada para a população que foi submetida ao teste na escola pesquisada. Há necessidade de se investigar melhor se o problema se refere à amostra da pesquisa ou à própria validade do teste para os propósitos estabelecidos. Quem garante a qualidade da medida é a qualidade do instrumento.

5 Considerações Finais

5.1 O Uso da TRI na Seleção de Estudantes para a Segunda Fase da OBMEP

O critério adotado pelas escolas participantes na seleção dos estudantes para a realização da Segunda Fase da OBMEP é o da dificuldade dos itens da prova, ou seja, os estudantes que acertaram mais itens da prova são selecionados para a próxima etapa. Além disso, caso haja alunos com mesma pontuação, cada escola determina e divulga um critério de desempate a ser utilizado, por exemplo, comparação de notas de bimestres anteriores, idade e comportamento dos estudantes.

Nota-se que, de maneira geral, as escolas utilizam um critério fundamentado na TCT. Tendo em vista que a TRI leva em consideração outros parâmetros relacionados ao teste, como discriminação e dificuldade de cada item, estimativa de acerto ao acaso e a proficiência do aluno, além de proporcionar uma análise pedagógica do desempenho, subentende-se que esse procedimento seja mais completo e justo para a seleção dos estudantes que vão prosseguir nas demais fases da Olimpíada.

Dessa forma, com a finalidade de detectar possíveis divergências nos resultados de classificação, o presente estudo apresenta, na **Tabela 14**, uma contraposição entre os estudantes selecionados pela TCT e os estudantes que apresentaram proficiência mais elevadas e que seriam selecionados para prosseguir, caso o critério de classificação utilizado fosse a TRI.

Tabela 14: Relação dos estudantes classificados para Segunda Fase da OBMEP pelo método da TCT e que divergem dos estudantes classificados pela TRI.

Estudantes classificados pela TCT	Escore total* (TCT)	Proficiência	Estudantes classificados pela TRI	Escore total (TCT)	Proficiência
515	10	672,8557	507	9	748,6854
527	10	627,6668	520	7	722,3995
507	9	748,6854	510	8	697,1439
509	9	537,6714	336	7	696,8356
513	9	538,2048	519	8	681,6104
517	9	522,6274	416	7	679,0807
531	9	614,8441	515	10	672,8557
534	9	549,1446	525	7	672,8557
501	8	632,6838	207	6	656,7199
504	8	522,4757	229	6	651,3551
505	8	578,6557	524	7	640,1977
506	8	606,5451	528	8	637,3802
508	8	563,7891	501	8	632,6838
510	8	697,1439	527	10	627,6668
511	8	511,359	529	7	620,4936
512	8	482,8198	110	7	616,9934
514	8	561,7967	531	9	614,8441
516	8	507,9593	109	7	610,5226

Estudantes classificados pela TCT	Escore total* (TCT)	Proficiência	Estudantes classificados pela TRI	Escore total (TCT)	Proficiência
518	8	500,8463	506	8	606,5451
519	8	681,6104	335	7	605,7496
521	8	553,5041	37	6	603,8407
526	8	547,1825	238	6	602,178
528	8	637,3802	49	6	600,9136
532	8	503,4299	500	7	594,9741
533	8	555,5987	432	9	592,9354
498	7	550,5203	53	7	592,112
499	7	524,1578	107	5	592,112
500	7	594,9741	393	5	592,112
502	7	577,1259	48	6	584,4179
503	7	520,829	116	6	583,1861
520	7	722,3995	360	7	582,2564
522	7	557,3479	444	7	579,0387
523	7	564,8752	55	6	578,6557
524	7	640,1977	505	8	578,6557
525	7	672,8557	502	7	577,1259
529	7	620,4936	27	7	576,8673
530	7	509,6664	332	6	575,4388

* valor máximo possível no teste via TCT é de 20 pontos

Por meio da análise dos resultados apresentados na **Tabela 14**, nota-se que, dos estudantes relacionados, 61,76% (21) deles divergem entre um processo de seleção e outro, ou seja, se o critério adotado para seleção dos estudantes aprovados para a Segunda Fase da OBMEP utilizasse fundamentos da TRI e não da TCT, os estudantes selecionados não seriam os mesmos, em sua maioria, portanto, deveriam ser substituídos. Os estudantes cuja numeração foi destacada em negrito seriam selecionados por ambos os critérios.

Nota-se que os critérios são semelhantes para alunos com notas 10 (em um total de 20 pontos), dado que todos os candidatos selecionados com esta nota também seriam selecionados caso o critério utilizado fosse fundamentado na TRI. A relação dos candidatos difere a partir da nota 9. Embora seis estudantes tenham sido selecionados com essa nota, apenas metade deles alcançaram um nível de proficiência suficiente para ser aprovado caso o critério de seleção fosse alterado.

É interessante observar que os dois estudantes com maiores escores na prova não obtiveram as maiores proficiências, ou seja, os melhores estudantes pela TCT não foram os mesmos na análise via TRI. Além disso, alunos com desempenho razoável na prova de maneira geral, seriam selecionados pela TRI pois alcançaram proficiência elevadas em relação aos demais estudantes da escola. Por exemplo, os estudantes 107 e 393 que acertaram somente um quarto da prova e os estudantes 37, 48, 49, 55, 116, 207, 229, 238, 332 que obtiveram nota 6 não foram selecionados, apesar de terem alcançado proficiência suficiente para realizarem a segunda fase da avaliação.

É possível observar também que o critério de desempate utilizado pela escola para selecionar alguns estudantes com nota 7 para a Segunda Fase mostrou-se inadequado,

tendo em vista que os estudantes 336, 416, 110, 109, 335, 53, 360, 444 e 27 alcançaram essa pontuação e por terem proficiências elevadas, deveriam ter sido selecionados, e no entanto não foram.

Nas avaliações de larga escala, como SAEB e Prova Brasil, os itens que não se comportam bem via análise psicométrica são retirados da prova e do cálculo das proficiências dos alunos. Se adotássemos o mesmo procedimento e calculássemos as notas (escores brutos) dos estudantes considerando apenas os 11 itens que foram considerados para análise via TRI, obteríamos os resultados explicitados na **Tabela 15**, na qual é possível observar uma comparação entre o escore obtido pelos estudantes classificados para a Segunda Fase e o escore que esses estudantes obteriam utilizando esse novo procedimento.

Tabela 15: Comparação de escores obtidos pelos estudantes selecionados na prova com todos os itens e na prova considerando apenas os 11 itens que possibilitaram o processamento dos dados, via TRI.

Estudantes classificados pela TCT	Escore total* (TCT)	Escore obtido nos 11 itens selecionados** (TCT)	Estudantes classificados pela TCT	Escore total* (TCT)	Escore obtido nos 11 itens selecionados** (TCT)
515	10	7	519	8	8
527	10	7	521	8	4
507	9	8	526	8	5
509	9	5	528	8	6
513	9	4	532	8	2
517	9	4	533	8	5
531	9	6	498	7	4
534	9	6	499	7	4
501	8	5	500	7	5
504	8	4	502	7	6
505	8	5	503	7	5
506	8	6	520	7	7
508	8	6	522	7	4
510	8	7	523	7	5
511	8	4	524	7	6
512	8	4	525	7	7
514	8	6	529	7	5
516	8	4	530	7	4
518	8	4			

* valor máximo possível no teste via TCT é de 20 pontos;

** valor máximo possível no teste via TCT, nessa situação, é de 11 pontos.

Na **Tabela 15**, infere-se que alguns estudantes selecionados para a Segunda Fase da OBMEP acertaram itens que não tiveram muito significado do ponto de vista avaliativo, especialmente no que se refere à discriminação, e, caso esses itens fossem retirados da análise, a nota não seria suficiente para que esses alunos fossem selecionados para próxima fase da prova. Por exemplo, o aluno 532 ficaria apenas com dois pontos, considerando uma pontuação máxima possível de 11.

Se esse procedimento fosse adotado, a escola pesquisada aprovaria para a Segunda Fase da OBMEP todos os estudantes que atingiram escore 8, 7 ou 6, além de alguns estudantes

com nota 5, que seriam classificados depois de definido algum critério de desempate. Desse modo, observa-se, na **Tabela 15**, que pelo menos 13 estudantes deveriam ser substituídos. Portanto, haveria uma alteração de pelo menos 35% dos estudantes que seguiriam para a fase seguinte da OBMEP.

Ao se utilizar a TCT como critério de seleção e um critério de desempate subjetivo, definido por coordenadores das diferentes escolas participantes, estudantes com proficiências maiores podem ser prejudicados e preteridos por outros de menor conhecimento. Além disso, o processo manual de correção pode apresentar algumas falhas. No presente estudo, a correção eletrônica revelou um equívoco ocorrido no processo manual de correção das provas da escola, uma vez que, o estudante número 432 obteve nota 9 e não foi indicado para a segunda etapa da avaliação.

Segundo Maia (2009), “o uso de instrumentos seletivos inadequados elege candidatos menos aptos, em detrimento de outros mais capazes e merecedores da vaga”. Sugere-se, então, o uso da Teoria de Resposta ao Item - TRI para a classificação dos estudantes para a Segunda Fase da OBMEP. No entanto, o método sugerido necessita de uma quantidade expressiva de participantes, o que pode impossibilitar a análise em escolas de pequeno porte. Nesses casos, uma possível solução seria reunir os dados das escolas localizadas em um mesmo município e analisá-los em conjunto.

5.2 Sugestões de Aprimoramento do Processo

Naturalmente, a análise aqui feita está longe de ser exaustiva e precisa ser aprofundada, tanto no que se refere a outros indicadores estatísticos relacionados à validade dos testes, quanto a outras amostras de indivíduos submetidos ao processo. No entanto, ela já sinaliza que a Primeira Fase da OBMEP precisa ser repensada, caso seja de interesse dos organizadores oferecer para as escolas um conjunto de itens capaz de propiciar *feedback* em termos de aprendizagem dos estudantes, e para garantir uma seleção justa dos alunos que irão prosseguir na Olimpíada.

Quando se busca compreender os motivos que levaram ao comportamento dos itens evidenciado pela análise aqui descrita, surge logo a inquietação da pesquisadora quando em entrevista com um coordenador da OBMEP foi informada que a prova não é elaborada a partir de nenhuma matriz de referência.

Segundo Rabelo (2013), a matriz de referência é um instrumento que traz a fundamentação teórico e metodológica do processo de avaliação de larga escala. Desse modo, a inexistência da matriz pode estar colaborando para a elaboração de um instrumento que faz uma avaliação dispersa, não focada, sem deixar claro se o propósito é a criação de um instrumento focado no domínio de determinados conteúdos, não especificados em nenhum local de domínio público, ou no desenvolvimento de certas habilidades mentais importantes para o desenvolvimento do pensamento matemático, tais como as capacidades de se fazer inferências dedutivas, indutivas e analógicas, entre outras.

A elaboração de uma matriz de referência para as provas da OBMEP possibilitaria uma análise mais completa dos resultados e um *feedback* mais consistente para os estudantes e professores. Seria possível dar sentido pedagógico aos resultados numéricos de proficiências encontrados pela TRI. Portanto, o aluno receberia um boletim de desempe-

nho não apenas com sua proficiência e seu posicionamento perante os outros estudantes da escola mas também com a descrição de cada nível de proficiência em termos de aprendizagem, isto é, com o significado prático dos valores alcançados na escala. Desse modo, seria possível determinar e retornar a cada estudante quais os conceitos ele já construiu, quais conceitos ainda estão em desenvolvimento e quais ainda não foram alcançados, traçando-se uma radiografia aproximada acerca do aprendizado de cada aluno.

Além da falta da matriz de referência, outro fator impossibilitou usar uma das maiores potencialidades da TRI, que é descrever o significado das faixas de proficiência da escala: a extrema dificuldade dos itens da avaliação. Todos os itens da prova foram representativos de níveis extremos da escala de proficiência, e no presente estudo, nenhum aluno atingiu esses níveis.

De modo geral, o presente estudo sugere que essa prova não se revelou com as características suficientes e adequadas para que seja feito um trabalho pedagógico dos estudantes, o que não significa que outras provas da OBMEP não tenham esse propósito. Dessa maneira, infere-se que, pelo menos no que se refere à Primeira Fase de 2014, a prova aplicada pela OBMEP não possibilitou que fosse cumprido plenamente os objetivos por ela proposto, revelando-se apenas como um processo de descoberta de novos talentos matemáticos, caso nenhum deles tenham ficado para trás pelas razões de escolha aqui apontadas.

Sugere-se que estudos mais aprofundados sejam feitos com uma amostra maior, que seja representativa da população que é submetida ao teste. Dessa maneira, seria possível concluir em que medida os resultados aqui evidenciados são consequência do instrumento ou da amostra pesquisada.

De qualquer maneira, cabe acrescentar que a amostra estudada representa parte da população do Distrito Federal e pertence a uma escola que apresentou a média de proficiências em Matemática superior à média de proficiências em Matemática de todo país na Prova Brasil de 2013, segundo dados do Inep, apresentados na **Tabela 16**. Além disso, ao ampliar a amostra desse estudo para uma amostra com 1658 estudantes de quatro escolas de Regiões Administrativas distintas do Distrito Federal, os parâmetros via TCT não apresentaram variações significativas. Assim como na amostra inicial, os itens continuaram apresentando dificuldade elevada e baixo poder discriminativo, conforme pode ser observado nas AGI's em anexo (**Apêndice C**).

Tabela 16: Médias de Proficiência de alunos do 9º ano em Matemática na Prova Brasil - 2013.

	Valor da Proficiência Média
Escolas Federais do Brasil	321,45
Escolas Estaduais do Brasil	244,41
Escolas Municipais do Brasil	238,85
Total Brasil	242,35
Escolar Estaduais do DF	247,13
Total Estado	247,86
Escola Pesquisada	244,99

Fonte: *www.inep.gov.br*

Por fim, acredita-se que com a ampliação desse estudo e com as modificações aqui sugeridas haja um melhor envolvimento das escolas e dos estudantes com a Olimpíada e uma estimação mais precisa do efeito da OBMEP no desempenho dos estudantes em Matemática.

Referências

- [1] ALVES, W.J.S. **O Impacto da Olimpíada de Matemática em Alunos da Escola Pública**. 2010. 92p. Dissertação (Mestrado Profissional em Ensino de Matemática). Pontifícia Universidade Católica de São Paulo - PUC-SP, São Paulo, 2010
- [2] ANDRADE, D.F.; BORGATTO, A.F. **Análise Clássica de Testes com Diferentes Graus de Dificuldade**. Estudos em Avaliação Educacional, São Paulo, V. 23, nº 52, Maio 2012
- [3] ANDRADE, D.F.; KARINO, C.A. **Nota Técnica**. Disponível em http://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_tri_enem_18012012.pdf, acessado em 3/2/2014
- [4] ANDRADE, D.F.; TAVARES, H.R.; VALLE, R.C. **Introdução à Teoria de Resposta ao Item: Conceitos e Aplicações**. SINAPE, 2000.
- [5] BARDIM, L. **Análise de Conteúdo**. Lisboa: Edições 70, 1977
- [6] CAMACHO, C. **Dez Anos de OBMEP**. Editorial Site OBMEP, disponível em <http://www.obmep.org.br/editorial.html>, acessado em 8/2/2015
- [7] GOMES, L.S. **A Teoria de Resposta ao Item na Avaliação em Larga Escala: Um Estudo sobre o Exame Nacional de Acesso do Mestrado Profissional em Matemática em Rede Nacional**. PROFMAT, 2012
- [8] LANDIM, C. (2013) *apud* FARJADO, V. **Cidades pequenas dominam 'ranking' de campeões em matemática**. Disponível em <http://server22.obmep.org.br:8080/media/servicos/recursos/314918.o>, acessado em 15/1/2015
- [9] LANDIM, C. (2014) *apud* MARQUES, C. **Olimpíada Brasileira de Matemática das Escolas Públicas completa dez anos comemorando resultados**. Disponível em http://www.brasilpost.com.br/2014/03/24/olimpiada-matematica-dez-anos_n_5017813.html, acessado em 12/2/2015
- [10] LIBANEO, J.C. et al. **Educação Escolar: Políticas, Estrutura e Organização**. 10ª ed., São Paulo: Cortez, 2011
- [11] MAIA, J.L. **Uso da Teoria Clássica dos Testes e da Teoria de Resposta ao Item na Avaliação da Qualidade Métrica de Testes de Seleção**. Programa de Pós-Graduação em Educação Brasileira. Universidade Federal do Ceará, Fortaleza, 2009
- [12] MARANHÃO, T.P.A. **Avaliação de Impacto da Olimpíada Brasileira de Matemática nas Escolas Públicas (OBMEP)**. CGEE, Série de Documentos Técnicos, nº 11, Julho 2011
- [13] PASQUALI, L. **Psicometria: Teoria dos Testes na Psicologia e na Educação**. Petrópolis: Vozes, 2003
- [14] **PORTAL BRASIL**. Disponível em <http://www.brasil.gov.br>, acessado em 28/1/2015

- [15] **PORTAL INEP**. Disponível em <http://www.inep.gov.br>, acessado em 1/3/2015
- [16] **PORTAL OBM**. Disponível em <http://www.obm.org.br>, acessado em 10/1/2015
- [17] **PORTAL OBMEP**. Disponível em <http://www.obmep.org.br>, acessado em 8/1/2015
- [18] **PORTAL QEDU**. Disponível em <http://www.qedu.org.br>, acessado em 8/3/2015
- [19] RABELO, M. **Avaliação Educacional: Fundamentos, Metodologia e Aplicações no Contexto Brasileiro**. Rio de Janeiro. SBM, 2013
- [20] **Resultados Prova Brasil 2013**. Disponível em <http://provabrazil.inep.gov.br/resultados-2013>, acessado em 11/4/2015
- [21] THURSTONE, L.L. **Attitudes can be measured**. American Journal of Sociology, 33, 529-554 *apud* PASQUALI, L.L.; PRIMI, R. **Fundamentos da Teoria da Resposta ao Item - TRI**. Avaliação Psicológica, Porto Alegre, V. 2, n^o 2, 2003
- [22] URBINA, S. **Fundamentos da Testagem Psicológica**. Artmed, 2004
- [23] VIANNA, H.M. **Introdução à Avaliação Educacional**. São Paulo: Ibrasa, 1989

Apêndice A: Termo de Consentimento Livre e Esclarecido*



Universidade de Brasília - UnB
Instituto de Exatas - IE
Departamento de Matemática - MAT
Mestrado Profissional em Matemática em Rede Nacional PROFMAT



TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Prezado(a) participante:

Sou estudante do curso de pós-graduação do programa PROFMAT - Mestrado Profissional em Matemática, do Departamento de Matemática da Universidade de Brasília. Estou realizando uma pesquisa sob supervisão do professor Mauro Luiz Rabelo, cujo objetivo é: Analisar os itens e as respostas dos estudantes do 8º e 9º ano do Ensino Fundamental às provas da primeira fase da OBMEP como meio para oferecer um feedback ao aluno e promover a valorização dessa avaliação.

Sua participação envolve ceder o uso dos gabaritos respondidos pelos estudantes do 8º e 9º ano do ensino fundamental às provas da OBMEP de 2014 e possibilitar a divulgação do resultado da pesquisa e as propostas de ação ao grupo de docentes.

A participação nesse estudo é voluntária e se você decidir não participar ou quiser desistir de continuar em qualquer momento, tem absoluta liberdade de fazê-lo.

Na publicação dos resultados desta pesquisa, sua identidade será mantida no mais rigoroso sigilo. Serão omitidas todas as informações que permitam identificá-lo, identificar a escola e/ou identificar os estudantes em estudo.

Mesmo não tendo benefícios diretos em participar, indiretamente você estará contribuindo para a compreensão do fenômeno estudado e para a produção de conhecimento científico.

Quaisquer dúvidas relativas à pesquisa poderão ser esclarecidas pelo pesquisador Ana Paula Lima Vilarinho por telefone 92222144 ou por e-mail: anaplilarinho@gmail.com.

Atenciosamente

Nome e assinatura do(a) estudante

Local e data

Nome e assinatura do professor orientador

Consinto em participar deste estudo e declaro ter recebido uma cópia deste termo de consentimento.

Nome e assinatura do participante

Local e data

*O termo anexado encontra-se sem preenchimento para que se mantenha o sigilo da escola participante.

Apêndice B e Anexo A: Requerimento e Aceite - CE-BRASPE



Universidade de Brasília - UnB
Instituto de Exatas - IE
Departamento de Matemática - MAT
Mestrado Profissional em Matemática em Rede Nacional - PROFMAT



Requerimento

Somos estudantes do curso de pós-graduação do programa PROFMAT - Mestrado Profissional em Matemática, do Departamento de Matemática da Universidade de Brasília. Estamos realizando dois estudos sob supervisão do professor Mauro Luiz Rabelo. O material de estudo são as avaliações da primeira fase da OBMEP 2014, nível 2, de escolas do Distrito Federal.

Vimos requerer junto ao Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos - CEBRASPE para que elabore, a partir dos dados em anexo, os seguintes documentos:

- AGI
- CCI
- DIFF por sexo e ano do aluno.

Desde já agradecemos pela contribuição.

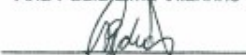
Quaisquer dúvidas relativas à pesquisa poderão ser esclarecidas pelas pesquisadoras:

Ana Paula Lima Vilarinho, e-mail: anaplvilarinho@gmail.com
ou Regiane Quezia Gomes da Costa, e-mail: regianegomez@gmail.com,
autoras dos respectivos estudos.

Atenciosamente



Ana Paula Lima Vilarinho



Regiane Quezia Gomes da Costa



Professor orientador

Local e data

Brasília, 19 de janeiro de 2015.

Prezado professor Paulo Portela,

De acordo com entendimentos mantidos entre a professora Maria Terezinha, o professor Mauro Rabelo, a professora Girlene, e as estudantes do curso de pós-graduação do programa PROFMAT, Ana Paula L. Vilarinho e Regiane Quezia G. da Costa, ficou acordado que a partir do dia 2 de fevereiro, as referidas estudantes deverão vir à Diretoria de Pesquisa em Avaliação, procurar a professora Girlene e ou Cacio para receber orientações de como elas próprias realizarão as análises desejadas.

Esclareço que não há no momento disponibilidade para que nossos estatísticos se encarreguem de gerar as AGIs, CCIs, e DIFF por sexo e ano, das respostas dos alunos avaliados na primeira fase da OBMEP 2014, nível 2, das escolas do DF.

No entanto, as duas alunas referidas acima estão autorizadas a utilizar um dos computadores da DPA, que possui instalados os programas necessários para realizar as análises pretendidas, para o que receberão orientações dos nossos estatísticos.

O tempo estimado para tais análises é de 15 dias.

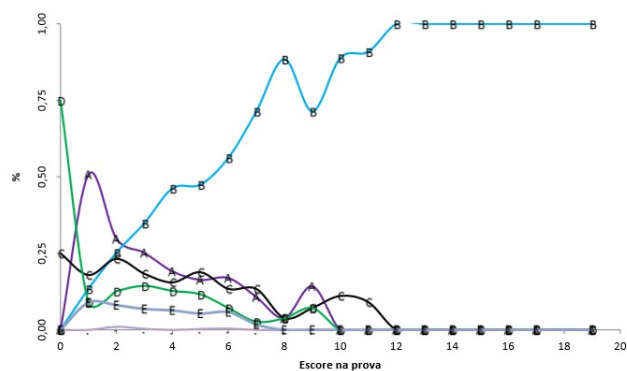
Atenciosamente,


Ana Paula L. Vilarinho
Coordenadora de Pesquisa
em Avaliação
CESPE/UnB

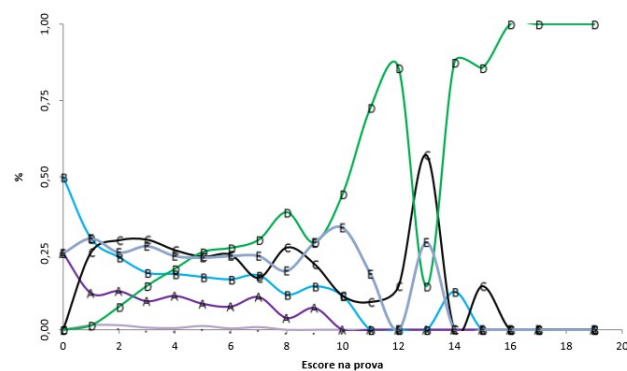
RECEBIDO
Em 19/01/15
Elizek

Apêndice C: AGI's dos Itens em uma Amostra Maior

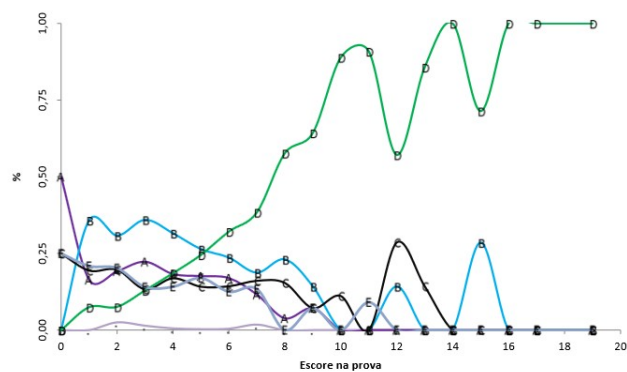
Item 1



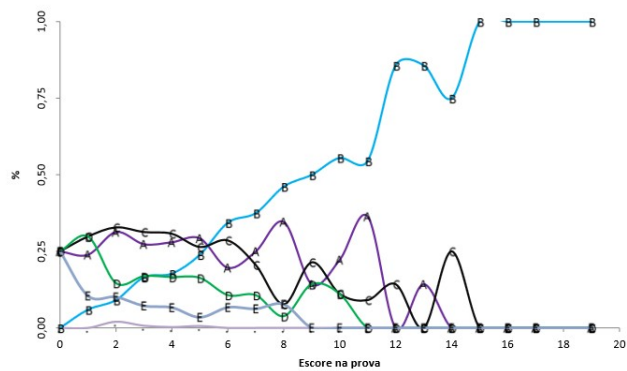
Item 2



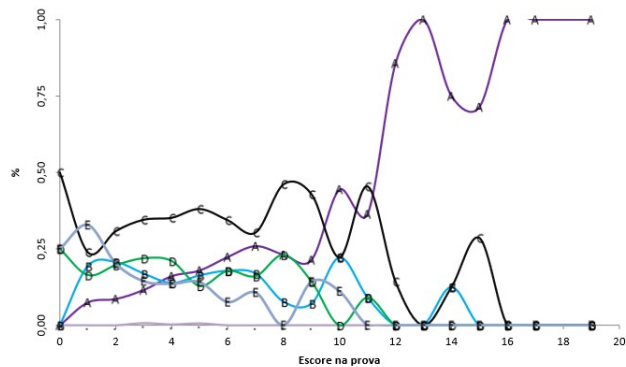
Item 3



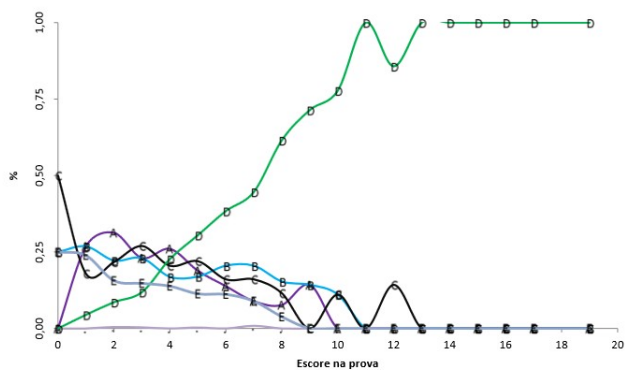
Item 4



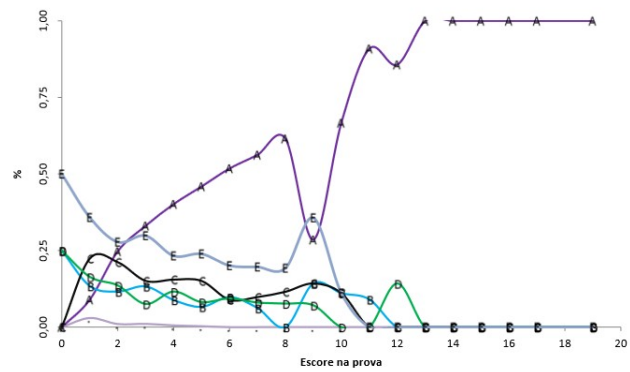
Item 5



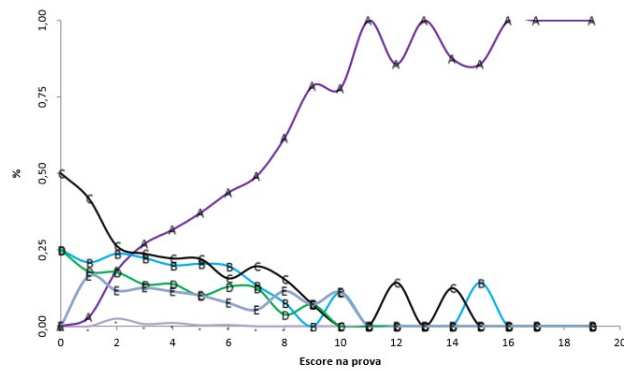
Item 6



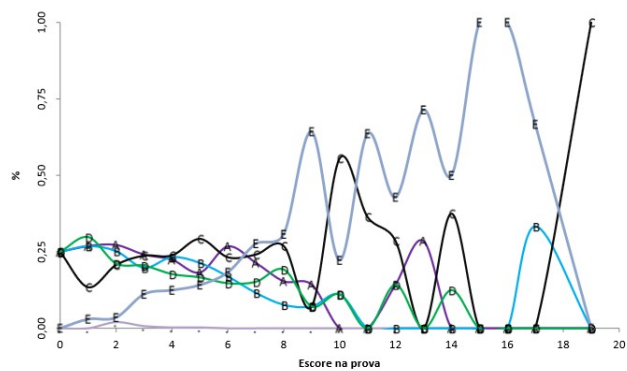
Item 7



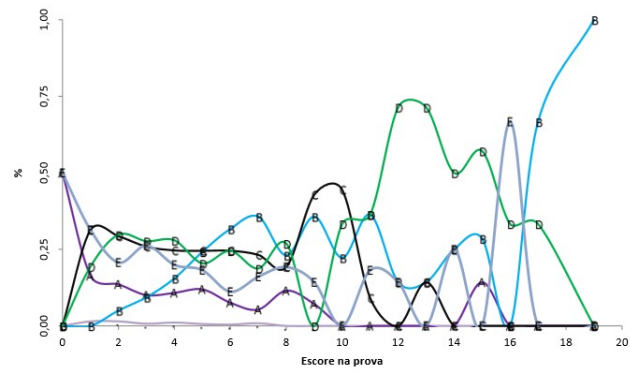
Item 8



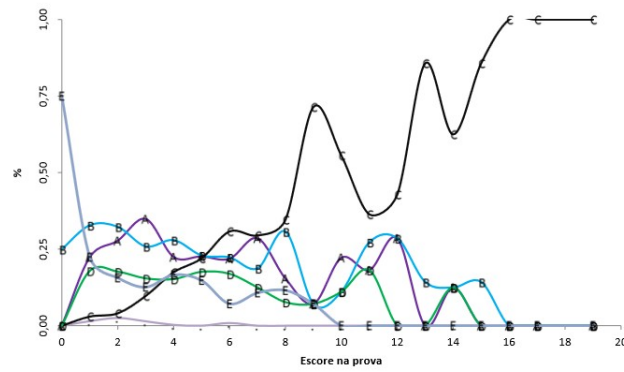
Item 9



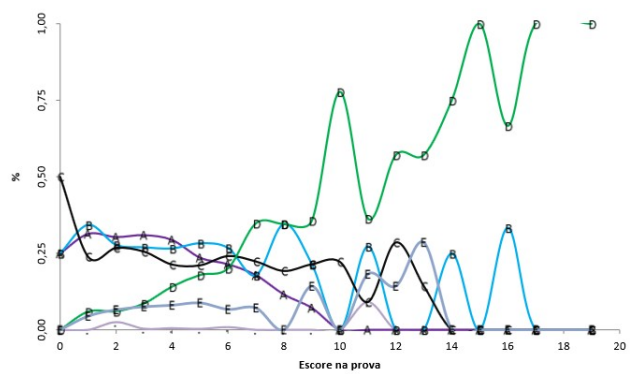
Item 10



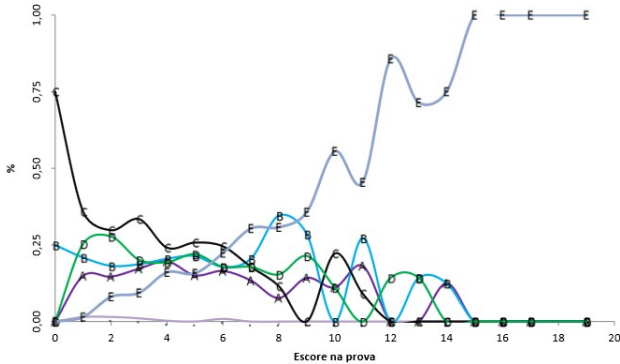
Item 11



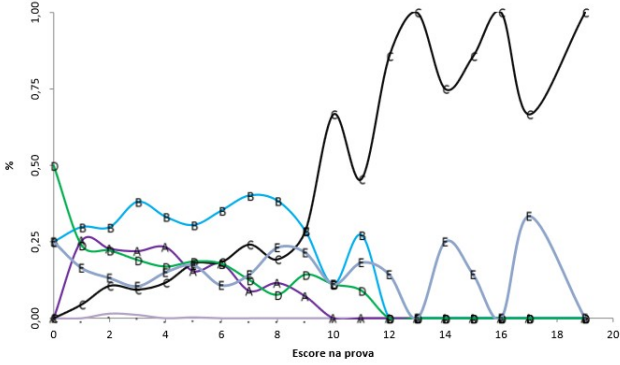
Item 12



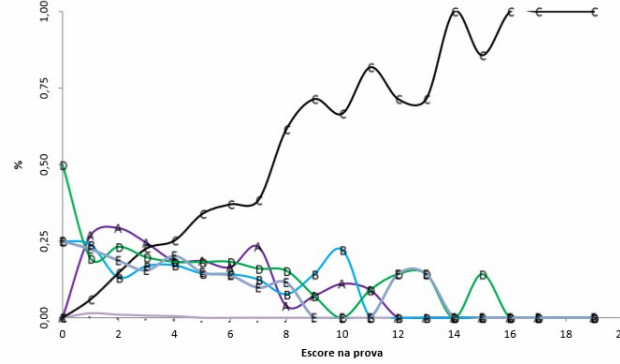
Item 13



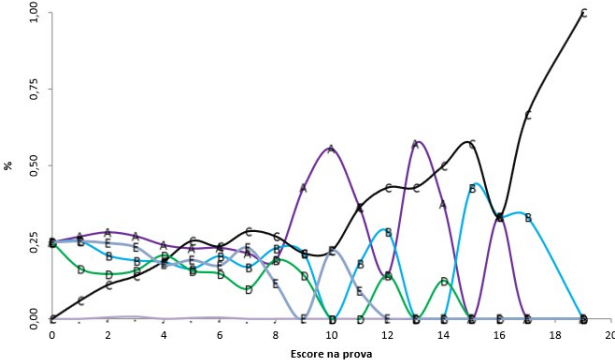
Item 14



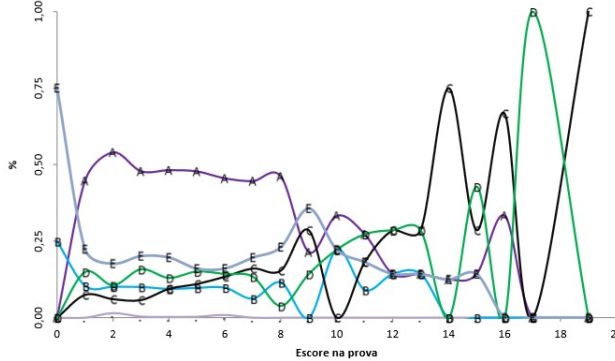
Item 15



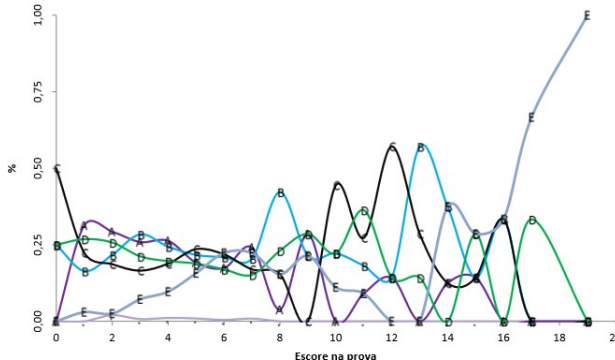
Item 16



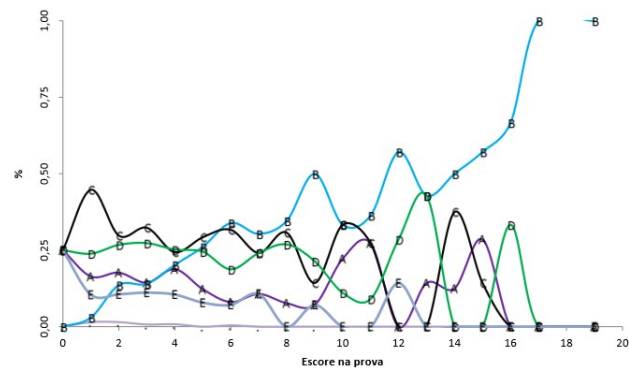
Item 17



Item 18



Item 19



Item 20

