

Katia Neves Pedroza

**COMPARABILIDADE DAS AVALIAÇÕES EDUCACIONAIS
EM PAPEL E EM COMPUTADOR: UM ESTUDO DO PISA 2012**

Dissertação submetida ao Programa de Pós-Graduação em Métodos e Gestão em Avaliação da Universidade Federal de Santa Catarina para obtenção do Grau de Mestre em Métodos e Gestão em Avaliação

Orientador: Prof. Dr. Adriano Ferreti Borgatto

Florianópolis
2015

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Pedroza, Katia Neves
Comparabilidade das avaliações educacionais em papel e
em computador : um estudo do PISA 2012 / Katia Neves
Pedroza ; orientador, Adriano Ferreti Borgatto -
Florianópolis, SC, 2015.
151 p.

Dissertação (mestrado profissional) - Universidade
Federal de Santa Catarina, Centro Tecnológico. Programa de
Pós-Graduação em Métodos e Gestão em Avaliação.

Inclui referências

1. Métodos e Gestão em Avaliação. 2. Comparabilidade. 3.
Testes em papel e em computador. 4. Programa Internacional
de Avaliação de Estudantes. 5. Teoria da Resposta ao Item.
I. Borgatto, Adriano Ferreti . II. Universidade Federal de
Santa Catarina. Programa de Pós-Graduação em Métodos e
Gestão em Avaliação. III. Título.

Katia Neves Pedroza

**COMPARABILIDADE DAS AVALIAÇÕES EDUCACIONAIS
EM PAPEL E EM COMPUTADOR: UM ESTUDO DO PISA 2012**

Esta Dissertação foi julgada adequada para obtenção do Título de Mestre Profissional e aprovada em sua forma final pelo Programa de Pós-Graduação em Métodos e Gestão em Avaliação.

Florianópolis, 17 de agosto de 2015.

Prof. Renato Cislighi, Dr.
Coordenador do Programa

Banca Examinadora:

Prof. Adriano Ferreti Borgatto, Dr.
Orientador
Universidade Federal de Santa Catarina

Frederico Neves Condé, Dr.
Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

Prof. Rafael Tezza, Dr.
Universidade do Estado de Santa Catarina

Prof. Dalton Francisco de Andrade, Dr.
Universidade Federal de Santa Catarina

Ao meu pai,
José Pedroza Nunes,
que com a sua admirável essência
sempre me ensina ser eterna a memória do amor.

À minha mãe,
Marleide Silveira Neves,
meu maior exemplo de força e fé,
de quem recebo o incentivo para duvidar do impossível.

AGRADECIMENTOS

A Deus, pela força que me concede a cada manhã.

Ao professor Adriano Ferreti Borgatto, pela oportunidade ímpar de ter um orientador tão comprometido, cuja sabedoria e disposição em ensinar contribuíram imensamente para o meu crescimento pessoal e profissional. Sua confiança e comentários tranquilizadores foram fundamentais para a conclusão desta dissertação.

Aos professores Dalton Francisco de Andrade, Rafael Tezza e Frederico Neves Condé pela valiosa contribuição à versão final deste trabalho.

Aos demais docentes do PPGMGA, à Manoella e à Katiana, bem como a todos os colegas de curso – José Henrique, Carlos Eduardo, Marcos, Denys, Téo, Rosi, Thiago, Clara, Simone, Léo, Rosa, Rafael, Diego, Ioná e Joel – que tornaram essa caminhada conjunta mais divertida e prazerosa.

Às minhas duas grandes famílias, divididas entre Brasília e Salvador, mas únicas no meu coração. À minha mãe, Marleide e às minhas irmãs, Mariana, Juliana e Pillar, por me motivarem a ser sempre melhor. Ao meu pai, José Pedroza e à tia Delma, por constituírem a base que nos une aos meus irmãos, Ciro e Gustavo e que se estendeu às minhas cunhadas, Larissa e Nathalie, a Rita e aos meus sobrinhos, Bia e Caio, que tanto alegam a minha vida. Agradeço por me sustentarem com infinito amor.

Ao Inep e, principalmente, à COGEP e à DAEB que de diferentes maneiras incentivaram a realização deste mestrado. Aos colegas de trabalho que direta ou indiretamente colaboraram com palavras de estímulo e discussões enriquecedoras. Aos gestores antigos e aos atuais, Marcelo Cortes, David Lima, Ticiane Marassi, Robert Lassance, Alexandre dos Santos e Patrícia Vieira que se mostraram solidários e me apoiaram, sobretudo na fase de escrita que parecia não chegar ao fim. À equipe do PISA 2012, João Bacchetto, Pedro Henrique e Meiry Furusato, que desde o início compartilharam comigo o interesse e entusiasmo com esse projeto. Aos colegas Marina Barbosa, Wallace Nascimento, João Horta e Marco César que se colocaram à disposição para me auxiliar com a Matemática e a Gabriela Barros que gentilmente me ajudou com o SPSS. Àqueles que fizeram e/ou ainda fazem parte da minha rotina diária, tornando-a mais leve e descontraída: Bartira Neri, Cléia Macedo, Margareth Cerqueira, Pedro Cayres, Flávia Ghignone, Alexandre Jaloto, André Victor, Lenice Medeiros, Flávia Costa, Deise

Guedes, Tatiane Viana, Ellen Lima e Raíssa Ortega. O meu especial agradecimento às amigas Danielle Costa e Viviane Fernandes, não apenas pela leitura do texto e aprendizagem acadêmica, mas principalmente por terem estado pacientemente ao meu lado nos bons e maus momentos, sempre com ouvidos atentos, abraços a postos, palavras de ânimo e lições de afeto, tão importantes para chegar até aqui. Da mesma forma, agradeço também à amiga Suellen Fachinetto pela parceria inestimável, que vivenciando as mesmas angústias e anseios, por diversas vezes com doçura e generosidade amenizou as minhas atividades.

Ao Joel Ribeiro, Beatriz Arcoverde e Clodoaldo da Silva, por me oferecerem suporte ao longo desses anos, minimizando assim algumas das preocupações que do contrário eu teria.

Aos queridos amigos, de perto e de longe, que há anos dividem comigo os sorrisos e lágrimas de cada conquista. A Ana Villela, Thaís Francis e Renato Oliveira, que mesmo quando distantes não teriam como se fazer mais presentes e que através do cuidado e carinho constantes me dedicam a compreensão necessária nos momentos mais difíceis. Ao Rodrigo Morales, que mesmo estando do outro lado do mundo me proporciona longas conversas e conselhos sensatos quando mais preciso. A Viviani Ramos, Soraia Costa, Eduardo Risi, Roberta Timponi, Tula Bisol, Carolina Mayumi, Frederico Cintra, Igor Orlandi e Gisele Landim, que de distintas formas me dão a certeza de poder contar sempre com um ombro amigo.

A todos, a minha mais sincera gratidão.

“Somos assim: sonhamos o voo, mas tememos a altura.
Para voar é preciso ter coragem para enfrentar o terror do vazio.
Porque é só no vazio que o voo acontece.
O vazio é o espaço da liberdade, a ausência de certezas.
Mas é isso o que tememos: o não ter certezas”

Os irmãos Karamazov,
Fiódor Dostoiévski, 1879.

RESUMO

O presente trabalho teve como objetivo comparar as avaliações educacionais em papel e em computador tendo como base as provas de Matemática do PISA 2012. Mais precisamente, analisou o efeito do modo de aplicação nos dois lados do processo: no instrumento e no avaliando, pelo estudo da dimensionalidade das provas e da comparação tanto das médias de grupos específicos, quanto do nível de dificuldade dos itens. Para as duas primeiras análises, foram utilizadas as respostas dos 1.390 alunos brasileiros que fizeram parte da amostra e que realizaram a prova de Matemática nos dois formatos. A avaliação do número de dimensões foi feita pelo método de análise fatorial de informação completa e o ajuste dos itens aos fatores revelou um conjunto predominantemente unidimensional, indicando a equivalência das provas nesse quesito. A comparação das médias considerou características individuais dos alunos e das escolas: posse de computadores, sexo, ano de ensino e localização da escola (urbana/rural; região/estado), demonstrando o impacto positivo das provas em computador no desempenho dos alunos de quase todos os grupos avaliados. A discussão do nível de dificuldade dos itens foi realizada através de uma análise quali-quantitativa, que levou em conta o posicionamento dos alunos e dos itens na escala geral de Matemática, bem como a dificuldade dos itens em papel e em computador da categoria de conteúdo *Espaço e Forma* classificados na mesma categoria de processo e contexto. O posicionamento dos 125 itens (84 em papel e 41 em computador) que fizeram parte do estudo indicou que, em geral, a prova em computador foi mais difícil que a prova em papel e que um número muito reduzido de itens em computador estavam adequados ao nível de proficiência da maioria dos alunos brasileiros, sugerindo uma possível superestimação das médias de proficiência nesse modo de aplicação, com impacto ainda maior nos grupos de baixo desempenho. As características distintas dos itens não permitiram obter resultados mais conclusivos quanto à comparabilidade dos dois modos de aplicação.

Palavras-chave: Comparabilidade. Teste em papel. Teste em computador. PISA 2012. Teoria da resposta ao item.

ABSTRACT

This study aimed to verify the comparability of educational assessments in both paper-based and computer-based formats based on the evidence of PISA 2012 Mathematics tests. More precisely, it examined the effect of the application mode on both sides of the process: in the instrument and in the examinees by the study of the dimensionality of the tests and by the comparison of the specific groups means and the level of items difficulty. For the first two analyzes, the responses of the 1.390 Brazilian students who were part of the sample and did the Mathematics tests in both formats were used. The evaluation of the number of dimensions was taken by the full information factor analysis and the adjustment of the items to the factors revealed a predominantly unidimensional set, indicating the equivalence of the tests in this matter. The comparison of means considered individual characteristics of the students and the schools: computers possession, sex, school year and school location (urban/rural, region/state), demonstrating the positive impact of the computer tests in students' performance in almost all groups evaluated. The discussion about the items difficulty level was accomplished through a qualitative and quantitative analysis, which considered the position of students and items in the Mathematics general scale as well as the difficulty of the paper and computer items of the content category *Space and Shape* classified under the same category of process and context. The position of the 125 items (84 in paper and 41 in computer) that were part of the study indicated that, in general, the computer-based test was more difficult than the paper-based test and that a very small number of items in computer were adequate to the proficiency level of most Brazilian students, suggesting a possible overestimation of proficiency average in this application mode, with even greater impact on the underperforming groups. The items' distinct characteristics did not allow to obtain more conclusive results regarding the comparability of the two application modes.

Keywords: Comparability. Computer-based test. Paper-based test. PISA 2012. Item response theory.

LISTA DE FIGURAS

| | |
|---|-----|
| Figura 1. Modelo de letramento em Matemática na prática. | 44 |
| Figura 2. Questões do questionário do aluno utilizadas para verificar o impacto da posse e uso de computadores nas médias de desempenho do Brasil nas provas de Matemática em papel e em computador do PISA 2012..... | 91 |
| Figura 3. Mapa dos itens de Matemática do PISA 2012 aplicados no Brasil. | 101 |
| Figura 4. Mapa dos itens de Matemática do PISA 2012 classificados na categoria de conteúdo Espaço e Forma. | 104 |
| Figura 5.1. Item CM020Q01. | 107 |
| Figura 5.2. Item CM020Q02. | 108 |
| Figura 5.3. Item CM020Q03. | 109 |
| Figura 5.4. Item CM020Q04. | 110 |
| Figura 6. Item PM934Q02. | 113 |

LISTA DE QUADROS E GRÁFICOS

| | |
|---|-----|
| Quadro 1. Escala de proficiência em Matemática. | 46 |
| Gráfico 1. Diagrama de dispersão entre as médias dos estados brasileiros nas provas de Matemática em papel e em computador do PISA 2012.. | 99 |
| Gráfico 2. Curvas de informação das provas de Matemática em papel e em computador aplicadas no Brasil no PISA 2012. | 102 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1. Médias nas avaliações de Matemática e Leitura em papel (PB) e em computador (CB) e diferença entre elas por país/economia – PISA 2012..... | 42 |
| Tabela 2. Desenho e análise do estudo por tipo de questão de comparabilidade | 58 |
| Tabela 3. Rotação dos blocos para composição dos cadernos da prova em papel do PISA 2012 e número de respondentes considerados para efeitos deste estudo em cada um dos cadernos aplicados no Brasil. | 69 |
| Tabela 4. Rotação dos blocos para a composição dos formulários da prova em computador do PISA 2012 e número de respondentes considerados para efeitos deste estudo em cada um dos formulários aplicados no Brasil. | 71 |
| Tabela 5. Distribuição dos itens de Matemática em papel e em computador do PISA 2012 aplicados no Brasil, de acordo com as categorias da matriz de letramento em Matemática e do tipo de resposta. | 79 |
| Tabela 6. Percentual de respostas ao último item de Matemática dos cadernos em papel e formulários em computador aplicados no Brasil no PISA 2012. | 82 |
| Tabela 7. Comparação dos modelos de uma e duas dimensões. | 83 |
| Tabela 8. Análise fatorial dos itens de Matemática do PISA 2012 | 85 |
| Tabela 9. Distribuição dos itens nas duas dimensões de acordo com o tipo de resposta, conteúdo, contexto e processo. | 88 |
| Tabela 10. Médias de desempenho dos alunos brasileiros nas provas de Matemática em papel e em computador do PISA 2012 em função das respostas a duas perguntas do questionário do aluno. | 92 |
| Tabela 11. Médias de desempenho dos alunos nas provas de Matemática em papel e em computador do PISA 2012 em função do sexo. | 93 |
| Tabela 12. Médias de desempenho dos alunos brasileiros nas provas de Matemática em papel e em computador do PISA 2012 em função do ano de ensino..... | 95 |

Tabela 13. Médias de desempenho dos alunos nas provas de Matemática em papel e em computador do PISA 2012 em função da localização das escolas. 96

Tabela 14. Médias de desempenho dos alunos nas provas de Matemática em papel e em computador do PISA 2012 em função da região/estado do Brasil 97

LISTA DE ABREVIATURAS E SIGLAS

| | |
|-------------------|--|
| ACER | - <i>Australian Council for Educational Research</i> |
| AIC | - Critério de Informação de Akaike (<i>Akaike Information Criterion</i>) |
| BIC | - Critério de Informação Bayesiano (<i>Bayesian Information Criterion</i>) |
| CAT | - Teste Adaptativo Computadorizado (<i>Computer adaptive test</i>) |
| CBAL | - <i>Computer-based assessment literacies</i> |
| CBAM | - <i>Computer-based assessment of Mathematics</i> |
| CBT | - <i>Computer-based test</i> |
| CESPE/ UnB | - Centro de Seleção e Promoção de Eventos da Universidade de Brasília |
| DIF | - Comportamento Diferencial do Item |
| DRA | - <i>Digital Reading Assessment</i> |
| EM | - <i>Expectation-Maximization</i> |
| ENCCEJA | - Exame Nacional para Certificação de Competências de Jovens e Adultos |
| ENEM | - Exame Nacional do Ensino Médio |
| FIFA | - <i>Full Information Factor Analysis</i> |
| IDH | - Índice de Desenvolvimento Humano |
| INEP | - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira |
| NAEP | - <i>National Assessment of Educational Progress</i> |
| OCDE | - Organização para Cooperação e Desenvolvimento Econômico |
| OECD | - <i>Organisation for Economic Co-operation and Development</i> |
| PBT | - <i>Paper-based test</i> |
| PISA | - Programa Internacional de Avaliação de Estudantes (<i>Programme for International Student Assessment</i>) |
| SAEB | - Sistema de Avaliação da Educação Básica |
| TAO TM | - <i>Testing Assisté par Ordinateur</i> |
| TCS | - <i>Target Cluster Size</i> |
| TCT | - Teoria Clássica dos Testes |
| TEA | - <i>Texas Education Agency</i> |

TIC - Tecnologia de Informação e Comunicação
TRI - Teoria da Resposta ao Item

SUMÁRIO

| | |
|--|-----------|
| 1 INTRODUÇÃO | 33 |
| 1.1 CONTEXTUALIZAÇÃO..... | 33 |
| 1.2 PROBLEMA DE PESQUISA | 34 |
| 1.3 OBJETIVOS | 35 |
| 1.3.1 <i>Objetivo Geral</i> | 35 |
| 1.3.2 <i>Objetivos Específicos</i> | 36 |
| 1.4 JUSTIFICATIVA | 36 |
| 1.5 LIMITES..... | 37 |
| 1.6 ESTRUTURA DO TRABALHO..... | 38 |
| 2 FUNDAMENTAÇÃO TEÓRICA..... | 39 |
| 2.1 O PISA..... | 39 |
| 2.1.1 <i>A avaliação de Matemática no PISA</i> | 43 |
| 2.2 A TEORIA DA RESPOSTA AO ITEM (TRI)..... | 49 |
| 2.2.1 <i>A unidimensionalidade dos testes</i> | 51 |
| 2.3 AS AVALIAÇÕES EM COMPUTADOR | 52 |
| 2.3.1 <i>Vantagens e desvantagens dos CBTs</i> | 55 |
| 2.3.2 <i>Métodos e modelos para comparabilidade</i> | 57 |
| 3 METODOLOGIA | 63 |
| 3.1 CARACTERIZAÇÃO DA PESQUISA | 63 |
| 3.2 CARACTERIZAÇÃO DO PISA 2012..... | 64 |
| 3.2.1 <i>Escopo e amostragem</i> | 64 |
| 3.2.2 <i>Aplicação e estrutura do material cognitivo</i> | 67 |
| 3.2.3 <i>O PISA e a TRI</i> | 73 |
| 3.3 METODOLOGIA PARA COMPARABILIDADE DOS TESTES | 76 |
| 3.3.1 <i>Análise da dimensionalidade</i> | 76 |
| 3.3.2 <i>Cálculo das médias</i> | 78 |
| 3.3.3 <i>Comparação do posicionamento dos itens</i> | 78 |
| 4 RESULTADOS E DISCUSSÃO | 81 |
| 4.1 ANÁLISE DA DIMENSIONALIDADE DAS PROVAS | 81 |
| 4.1.1 <i>Análise exploratória</i> | 81 |
| 4.1.2 <i>Determinação do modelo</i> | 83 |
| 4.1.3 <i>Análise fatorial</i> | 84 |
| 4.2 COMPARAÇÃO DAS MÉDIAS | 90 |
| 4.3 ANÁLISE DAS PROVAS..... | 100 |
| 4.3.1 <i>Comparação do posicionamento dos itens</i> | 103 |

| | |
|---|------------|
| 5 CONCLUSÃO | 115 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 119 |
| APÊNDICE A | 129 |
| APÊNDICE B | 131 |
| ANEXO | 133 |

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

As avaliações externas em larga escala possuem posição de destaque no cenário educacional brasileiro, fornecendo diagnóstico e subsídios para a implementação ou manutenção de políticas através de um monitoramento contínuo do sistema educacional (KLEIN & FONTANIVE, 1995). Castro (2000, p. 121) afirma que elas “cumpram um papel estratégico para o planejamento e desenho prospectivo de cenários, auxiliando enormemente a formulação de novas políticas e programas que possam responder às tendências de mudanças observadas”.

Nesse sentido, não podemos negar as mudanças profundas que tem ocorrido no processo de ensino-aprendizagem provocadas pelo consumo crescente das Tecnologias de Informação e Comunicação (TICs), que pode ainda trazer outros benefícios. Além de possibilitar novas formas de aprendizado, dentro e fora de sala, a facilidade e rapidez com que as informações são transmitidas e recebidas transformou a relação educador/educando e as habilidades necessárias para que os jovens possam ter plena participação na sociedade (OECD, 2005).

Para Pellegrino e Quellmalz (2010), as tecnologias expandiram não só a natureza da argumentação e o uso de evidências, mas também os fenômenos que podem ser investigados pela representação e manipulação de domínios, sistemas, modelos e dados de uma forma que não era possível anteriormente. Eles observam que a transição de um modelo estático para um modelo dinâmico tem consequência na forma como as disciplinas podem ser ensinadas, ao mesmo tempo em que uma nova geração de avaliações está em vias de transformar o que, como, quando, onde e por que uma avaliação ocorre.

Bennett (2002) argumenta que a incorporação da tecnologia nas avaliações é inevitável porque tem se entrelaçado com o que e como os alunos aprendem e, assim, os meios usados para verificar os avanços devem manter o mesmo passo. No entanto, segundo Clarke e Dede (2010, p. 309), “apesar de quase três décadas de desenvolvimento nas TICs e uma geração de pesquisas em cognição e em novas estratégias pedagógicas, o campo da avaliação não progrediu muito além dos testes baseados em itens no papel, cujo modelo fundamental foi desenvolvido um século atrás”¹.

¹ Tradução livre. No original: “Despite almost three decades of advances in information and communications technology (ICT) and a generation of research on cognition and on new pedagogical strategies, the field of assessment has not

Mc Donald (2002) esclarece que essas mudanças têm ocorrido de forma gradual por várias razões, incluindo o conservadorismo inerente em muitos sistemas educacionais, às vezes o acesso restrito à tecnologia e a resistência a algumas limitações que os testes em computador impõem, como a dificuldade em corrigir questões abertas. Para o pesquisador, a natureza gradual dessas mudanças significa que em educação os testes em papel e em computador provavelmente irão coexistir em um futuro próximo, com algumas avaliações sendo aplicadas nos dois formatos, fazendo com que a equivalência entre elas seja um fator importante a ser considerado.

De acordo com Singleton (2001), nos estágios relativamente iniciais da história das avaliações em computador, a equivalência surgiu em grande parte como uma preocupação na conversão de testes já existentes, sendo, mais recentemente, nos testes projetados especificamente para esse novo formato, uma questão de validação (concorrente e/ou prevista) em relação à testes estabelecidos que medem características iguais ou similares ou em relação à critérios objetivos.

Em uma revisão de estudos de comparabilidade, Blazer (2010) aponta que quando consideradas características demográficas, habilidade com computadores, tipo de computadores, características dos testes, tipos de itens e área (conteúdo) avaliada, o modo de aplicação pode impactar no desempenho dos alunos. Segundo Choi, Kim e Boo (2003), a maioria desses estudos apresentam resultados ambíguos ou conflitantes, provavelmente devido à diferenças idiossincráticas em muitas variáveis.

1.2 PROBLEMA DE PESQUISA

Diante das suas inúmeras vantagens, os testes em computador têm sido cada vez mais utilizados em avaliações educacionais. Ghaderi et. al (2014) afirmam que esse possivelmente será o principal modo de aplicação no futuro. Todavia, Parshall et. al (2002) ponderam que, embora essas avaliações sejam frequentemente percebidas como o “estado da arte” ou instintivamente melhores que as avaliações tradicionais em papel, tais suposições não são precisas e muitos são os desafios inerentes à aplicação de um teste em computador, de forma que o desenvolvimento de um sistema de avaliação computadorizado não deve ser considerado fácil.

progressed much beyond paper-and-pencil item-based tests whose fundamental model was developed a century ago”.

Sem dúvida, o estabelecimento de avaliações educacionais em papel por si só já representa um enorme esforço para elaboração de instrumentos válidos e confiáveis, acrescentando-se ainda, de forma mais acentuada àquelas aplicadas em larga escala, fatores importantes relacionados ao sigilo e à logística. As dificuldades enfrentadas pelo Exame Nacional do Ensino Médio – ENEM em 2009, 2010 e 2011 são alguns exemplos, de sorte que pesquisadores como Alavarse e Melo (2012) defendem a utilização de testes adaptativos computadorizados (*computer adaptive test*) – CAT como uma potencial solução à esses problemas, não sem destacar a necessidade de serem conduzidos estudos em paralelo para responder aos questionamentos suscitados pela inclusão de tal inovação, dentre eles a comparabilidade dos dois modos de aplicação.

Para Mc Donald (2002), sempre que um teste em papel é convertido para o formato em computador ou quando são utilizados em paralelo indiscriminadamente, torna-se fundamental verificar a equivalência das pontuações e dos construtos, bem como a influência das diferenças individuais sobre ela, o que é particularmente importante nas avaliações educacionais de alto impacto e para fins de diagnóstico, em que são estipuladas pontuações de corte.

Embora sejam encontrados na literatura internacional estudos que exploram diferentes aspectos relacionados ao tema, Dillon (1992 *apud* Al-Amri, 2009) argumenta não ser apropriada a suposição de comparabilidade sem uma investigação apropriada em um contexto específico. Assim, é imperioso salientar que, no Brasil, as avaliações em computador ainda estão em sua infância, sendo o PISA 2012 a primeira experiência a nível nacional, o que explica a lacuna existente de pesquisas similares realizadas no país.

Diante desse cenário, esta pesquisa pretende contribuir para a discussão a respeito da implementação dos testes em computador nas avaliações educacionais em larga escala aplicadas no Brasil, fornecendo uma base de sustentação teórica para futuros estudos relacionados à comparabilidade dos dois modos de aplicação no referido contexto.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Comparar as avaliações educacionais em papel e em computador com base na prova de Matemática do PISA 2012.

1.3.2 Objetivos Específicos

1. Verificar a dimensionalidade das provas de Matemática em papel e em computador do PISA 2012 para o conjunto de alunos brasileiros que responderam a ambas as provas;
2. Observar o efeito do formato de aplicação no desempenho dos alunos nas provas de Matemática em papel e em computador do PISA 2012;
3. Comparar o posicionamento dos itens de Matemática em papel e em computador da edição de 2012 na escala do PISA.

1.4 JUSTIFICATIVA

Em 2012, foi aplicada no Brasil, a nível nacional, a primeira avaliação educacional em larga escala respondida em computador, o Programa Internacional de Avaliação de Estudantes (*Programme for International Student Assessment*) – PISA. Os resultados mostraram que os nossos alunos se saíram muito melhor na prova em computador comparada à prova em papel e o país se destacou por apresentar, nessa direção, a maior diferença para a área de Matemática (OECD, 2104a).

De acordo com o relatório publicado pela Organização para Cooperação e Desenvolvimento Econômico – OCDE (em inglês, *Organisation for Economic Co-operation and Development – OECD*) (OECD, 2014a, p. 491), “em geral, houve um elevado grau de consistência no desempenho dos alunos em itens respondidos em papel e em computador”². No entanto, foram observadas exceções importantes em alguns países, seja a favor da avaliação em papel ou da avaliação em computador, tendo a própria OCDE apontado para a necessidade de serem realizadas outras análises a fim de explorar a extensão em que essas diferenças são causadas pela natureza distinta das tarefas, pelos diferentes modos de aplicação ou pela familiaridade dos alunos com os computadores.

Embora existam alguns estudos na literatura sobre o impacto da aplicação de provas em computador no desempenho dos alunos (BENNETT, 2003; CLARIANA & WALLACE, 2002), não foram encontradas publicações relacionadas ao Brasil que permitam aprofundar a compreensão

² Tradução livre. No original: “*In general, there is a high degree of consistency in student performance on items delivered on paper and by computer*” .

dos fatores que podem explicar essas diferenças no contexto da nossa realidade. Dessa forma, os dados do PISA 2012 são uma excelente oportunidade para verificar a comparabilidade desses dois modos de aplicação, importante para identificar aspectos determinantes em uma possível transição.

Aguiar (2010), em referência a Ferrer (2003), destaca que um estudo educacional comparativo não deve se limitar apenas a medir e comparar os resultados brutos, conseguidos pelos alunos, e é nessa perspectiva que essa pesquisa se torna relevante, ao conjugar e articular os paradigmas quantitativo e qualitativo, recorrendo a metodologias que ajudam a explicar as diferenças encontradas. Do ponto de vista teórico, esse estudo ainda evidencia alguns dos desafios, vantagens e desvantagens da avaliação em computador.

A escolha pela Matemática se deu por esse ter sido o principal domínio avaliado no PISA 2012 e, portanto, com mais informações disponíveis. Além disso, a matriz da Avaliação de Leitura Digital é baseada em uma escala adicional, apenas com tarefas de leitura digital (OECD, 2013), impossibilitando algumas das análises pretendidas.

1.5 LIMITES

O embasamento teórico do presente estudo limita-se aos testes em computador não-adaptativos, uma vez que esse é o formato adotado pelo PISA. Assim, não abrange comparações diretas com o CAT, tendo em vista as suas particularidades.

Dentro do contexto das avaliações em computador não-adaptativas, as análises realizadas foram baseadas em apenas um domínio, a Matemática, podendo produzir resultados diferentes quando consideradas outras áreas do conhecimento.

Apesar de estarem relacionadas ao mesmo conteúdo, as provas em papel e em computador do PISA, analisadas neste trabalho, apresentam características distintas que vão além do modo de aplicação, como, por exemplo, alguns itens interativos na prova em computador. Nessa perspectiva podem surgir algumas questões que, no entanto, são minimizadas pelo fato do mesmo aluno responder a ambas as provas.

É importante ressaltar ainda que não se teve aqui a pretensão de aprofundar a discussão a respeito dos méritos ou não das avaliações em computador, não obstante elas sejam o pano de fundo dessa dissertação.

1.6 ESTRUTURA DO TRABALHO

Este trabalho está estruturado em cinco capítulos. No primeiro capítulo é feita uma introdução ao assunto, a contextualização do tema, a discussão do problema de pesquisa, a exposição dos objetivos e dadas a justificativa e as limitações da pesquisa.

No capítulo dois é apresentada a fundamentação teórica sobre os temas que sustentam a pesquisa: o PISA, a Teoria da Resposta ao Item e as avaliações em computador, abordando vantagens e desvantagens e alguns estudos e métodos para comparabilidade.

No capítulo três são detalhados os procedimentos metodológicos referentes ao PISA, a amostragem, o tratamento dos dados, a estrutura dos instrumentos, bem como os modelos utilizados para as análises realizadas neste estudo.

No capítulo quatro são apresentados e discutidos os resultados do estudo da dimensionalidade das provas de Matemática em papel e em computador, bem como a comparação das médias e do posicionamento dos itens dentro da escala.

No capítulo cinco são apresentadas as conclusões e as recomendações para trabalhos futuros, seguidas das referências bibliográficas utilizadas, dos apêndices e do anexo.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo revisa a literatura que engloba o tema desta pesquisa e suas variáveis. Uma vez que se baseia nos dados do PISA, é feita uma breve descrição dessa avaliação e da técnica estatística utilizada para a interpretação dos seus resultados, a Teoria da Resposta ao Item – TRI. Dentro desse contexto são também expressas considerações sobre a dimensionalidade dos testes. No que tange às avaliações em computador, são abordados os principais conceitos e evidenciadas algumas das suas vantagens e desvantagens. Por fim é lançada luz sobre os modelos recomendados para o estudo da comparabilidade dos dois modos de aplicação, i.e., em papel e em computador.

2.1 O PISA

O PISA é uma avaliação internacional em larga escala, desenvolvida e coordenada pela OCDE e aplicada a alunos na faixa dos 15 anos, idade em que se pressupõe o término da escolaridade básica obrigatória na maioria dos países. A cada edição, o estudo tem maior ênfase em um dos principais domínios avaliados: Leitura, Matemática e Ciências.

Organizado em ciclos trienais, a primeira aplicação do PISA ocorreu em 2000, tendo como foco principal a avaliação de Leitura; em 2003, maior ênfase foi dada à avaliação de Matemática e a Resolução de Problemas foi incluída como um domínio opcional a ser avaliado, apesar de não ter feito parte das edições de 2006 e 2009, que tiveram como destaque as avaliações de Ciências e Leitura, respectivamente. A edição de 2009 teve ainda como componente opcional a avaliação de Leitura Digital, oferecida novamente em 2012, quando a Matemática voltou a ser o eixo central e outras avaliações opcionais foram acrescentadas: Letramento Financeiro em papel e Matemática e Resolução de Problemas em computador. Para essa última, uma nova matriz foi elaborada e metodologias adicionais de avaliação foram implementadas, permitindo a captura em tempo real da interação do aluno com o problema (OECD, 2013).

Participaram da edição de 2012 do PISA 65 países e economias, incluindo todos os 34 países membros da OCDE e 31 países e economias parceiras, sendo que 44 optaram por participar da avaliação de Resolução de Problemas e, dentre esses, 32 participaram também das avaliações de Leitura (*Digital Reading Assessment – DRA*) e Matemática (*computer-based assessment of Mathematics – CBAM*), que foram oferecidas em conjunto em uma avaliação de letramento em computador (*computer-based literacies – CBAL*) (OECD, 2014a). O Brasil participa do PISA desde a sua primeira

edição, mas somente em 2012 foram aplicados componentes opcionais, incluindo no estudo definitivo todos os domínios oferecidos em computador (Resolução de Problemas e *CBAL*).

A concepção, implementação e coleta de dados do PISA ficam sob a responsabilidade de um consórcio internacional e em cada país participante há uma coordenação nacional que, no Brasil, está a cargo do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) (OECD, 2014b; INEP, c2011). O escopo e a natureza das avaliações, bem como as informações a serem coletadas são decididos em conjunto por especialistas e governantes dos países participantes, orientados por uma política de interesses comuns. Os rigorosos mecanismos de controle de qualidade utilizados produzem resultados com alto nível de confiança e validade, que permitem, por exemplo, aos gestores em todo o mundo medirem os conhecimentos e as habilidades dos alunos em seus países e em comparação com outros países, estabelecer parâmetros de referência para a melhoria da educação e entender os pontos fortes e fracos de seus sistemas de ensino, aprendendo com políticas e práticas de excelência (OECD, 2013).

A proposta do PISA é avaliar os alunos quanto aos conhecimentos e as habilidades consideradas essenciais para uma plena participação na sociedade moderna, indo além do domínio de um currículo escolar específico. A avaliação tem por base o conceito de letramento, que envolve a capacidade que o indivíduo tem de analisar, raciocinar e comunicar-se com eficiência, na medida em que identifica, interpreta e resolve problemas em diferentes contextos. Assim, examina não apenas a capacidade do aluno de reproduzir o que foi aprendido, mas de extrapolar esse conhecimento, aplicando-o também em situações não-familiares, dentro e fora da escola (OECD, 2014a). Além das provas, são utilizados questionários a fim de obter informações para a elaboração de indicadores contextuais, que permitem relacionar o desempenho dos alunos a variáveis demográficas, socioeconômicas e educacionais (INEP, c2011).

Uma vez que a natureza das competências que contribuem para o sucesso dos indivíduos está em constante transformação, avaliações como o PISA precisam avançar e se adaptar para acompanhá-las. Dessa forma, em 2006 o PISA deu um primeiro passo para a inclusão da tecnologia nos seus instrumentos de avaliação, realizando um estudo na Dinamarca, Islândia e Coréia que comparou os resultados de Ciências em papel com um componente da mesma área aplicado em computador (OECD, 2010). Em 2012, a inclusão de uma avaliação de Matemática em computador no PISA foi justificada por duas razões:

A primeira é que itens em computador podem ser mais interativos, autênticos e atraentes que itens em papel. Eles podem ser apresentados em novos formatos (ex: arrastar e soltar), incluir dados do mundo real (como um conjunto de dados grande e ordenável) e utilizar cores, gráficos e movimento para ajudar na compreensão. Podem ser apresentados aos estudantes estímulos em movimento e representações de objetos tridimensionais que podem ser rotacionados ou permitir um acesso mais flexível a informações relevantes. (...) Segundo, os computadores se tornaram ferramentas essenciais para representar, visualizar, explorar e experimentar todos os tipos de processos, fenômenos e objetos matemáticos, sem mencionar a compreensão de todos os tipos de cálculos – em casa, na escola, no trabalho. (OECD, 2014a, p. 491)³.

Embora os resultados dessa avaliação tenham sido reportados separados da avaliação de Matemática em papel, ambos foram apresentados na escala de Matemática do PISA. O relatório da OCDE (OECD, 2014a) é bem sucinto quanto à análise dos dados da avaliação em computador, mas destaca que alguns países apresentaram diferenças significativas entre os dois modos de aplicação. No Brasil, a aplicação de Matemática em computador se mostrou mais vantajosa, com a maior diferença em relação à avaliação em papel (aproximadamente 30 pontos na escala do PISA). A Tabela 1 evidencia a diferença na média de desempenho dos alunos nesses dois formatos de aplicação para todos os países/economias participantes, tanto para a avaliação de Matemática, quanto para a avaliação de Leitura.

³ Tradução livre. No original: “*First, computer-based items can be more interactive, authentic and engaging than paper-based items. They can be presented in new formats (e.g. drag-and-drop), include real-world data (such as a large, sortable dataset), and use colour, graphics and movement to aid comprehension. Students may be presented with a moving stimulus or representations of three-dimensional objects that can be rotated, or have more flexible access to relevant information. (...) Second, computers have become essential tools for representing, visualising, exploring, and experimenting with all kinds of mathematical objects, phenomena and processes, not to mention for realising all types of computations – at home, at school, and at work*”.

Tabela 1. Médias nas avaliações de Matemática e Leitura em papel (PB) e em computador (CB) e diferença entre elas por país/economia – PISA 2012.

| País/economia | Matemática | | | Leitura | | |
|-------------------|------------|-----|-----------|---------|-----|-----------|
| | CB | PB | Diferença | CB | PB | Diferença |
| Alemanha | 509 | 514 | -5 | 494 | 508 | -14 |
| Austrália | 508 | 504 | 4 | 521 | 512 | 9 |
| Áustria | 507 | 506 | 1 | 480 | 490 | -10 |
| Bélgica | 512 | 515 | -3 | 502 | 509 | -7 |
| Brasil | 421 | 391 | 30 | 436 | 410 | 26 |
| Canadá | 523 | 518 | 5 | 532 | 523 | 9 |
| Chile | 432 | 423 | 9 | 452 | 441 | 11 |
| China – Hong Kong | 550 | 561 | -11 | 550 | 545 | 5 |
| China – Macao | 543 | 538 | 5 | 515 | 509 | 6 |
| China – Shangai | 562 | 613 | -51 | 531 | 570 | -39 |
| China – Taipei | 537 | 560 | -23 | 519 | 523 | -4 |
| Cingapura | 566 | 573 | -7 | 567 | 542 | 25 |
| Colômbia | 397 | 376 | 21 | 396 | 403 | -7 |
| Coréia | 553 | 554 | -1 | 555 | 536 | 19 |
| Dinamarca | 496 | 500 | -4 | 495 | 496 | -1 |
| Emirados Árabes | 434 | 434 | 0 | 407 | 442 | -35 |
| Eslováquia | 497 | 482 | 15 | 474 | 463 | 11 |
| Eslovênia | 487 | 501 | -14 | 471 | 481 | -10 |
| Espanha | 475 | 484 | -9 | 466 | 488 | -22 |
| Estados Unidos | 498 | 481 | 17 | 511 | 498 | 13 |
| Estônia | 516 | 521 | -5 | 523 | 516 | 7 |
| França | 508 | 495 | 13 | 511 | 505 | 6 |
| Hungria | 470 | 477 | -7 | 450 | 488 | -38 |
| Irlanda | 493 | 501 | -8 | 520 | 523 | -3 |
| Israel | 447 | 466 | -19 | 461 | 486 | -25 |
| Itália | 499 | 485 | 14 | 504 | 490 | 14 |
| Japão | 539 | 536 | 3 | 545 | 538 | 7 |
| Noruega | 498 | 489 | 9 | 500 | 504 | -4 |
| Polônia | 489 | 518 | -29 | 477 | 518 | -41 |
| Portugal | 489 | 487 | 2 | 486 | 488 | -2 |
| Rússia | 489 | 482 | 7 | 466 | 475 | -9 |
| Suécia | 490 | 478 | 12 | 498 | 483 | 15 |

Fonte: Elaboração da autora (a partir de dados da OECD, 2014a, p. 19, 495 e 501).

Devido às diferenças observadas, a própria OCDE indicou a necessidade de serem realizadas outras análises a fim de encontrar suas causas, apontando como possíveis fatores a natureza distinta das tarefas, a familiaridade dos alunos com os computadores e os diferentes modos de aplicação (OECD, 2014a). Este estudo teve como propósito avaliar a influência desse último fator nas diferenças encontradas no Brasil.

2.1.1 A avaliação de Matemática no PISA

As informações apresentadas nesta seção foram retiradas da matriz do PISA (OECD, 2013), que descreve para cada domínio, os conteúdos que os alunos precisam adquirir, assim como os processos que precisam ser realizados e os contextos em que as habilidades e conhecimentos são aplicados. A matriz de letramento em Matemática foi desenvolvida em conjunto entre o *Australian Council for Educational Research (ACER)* e *Achieve, Inc.*, uma organização dos Estados Unidos para o desenvolvimento da educação, tendo sido completamente revisada na avaliação de 2012. Além de incorporar a avaliação em computador, ela incluiu os processos em que os alunos se envolvem ao solucionar problemas e as capacidades matemáticas fundamentais que estão na base desses processos.

Para o PISA, a proficiência em Matemática significa a capacidade do indivíduo de formular, aplicar e interpretar a Matemática em uma variedade de contextos. O termo descreve a capacidade dos indivíduos de raciocinar matematicamente e utilizar os conceitos, procedimentos, fatos e ferramentas matemáticas para descrever, explicar e prever fenômenos. O letramento matemático não é um atributo que o indivíduo tem ou não e sim uma habilidade que pode ser desenvolvida ao longo da vida (OECD, 2014a, p. 25)⁴.

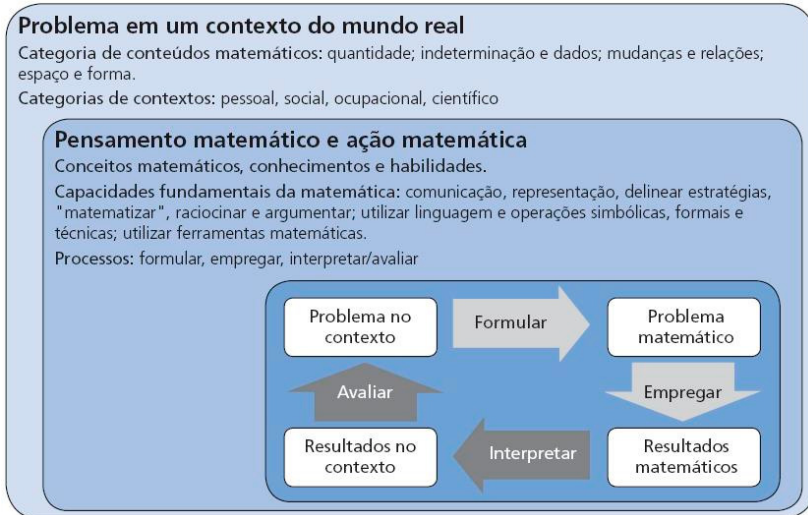
A definição de letramento em Matemática do PISA cita explicitamente o uso de ferramentas matemáticas, descritas na matriz como equipamentos físicos e digitais, *softwares* e dispositivos de cálculo, sendo particularmente apropriada à avaliação em computador, que oferece a oportunidade de incluir uma ampla gama de ferramentas matemáticas nos itens do teste (*softwares* estatísticos, utilitários de visualização e construção geométrica e instrumentos de medida virtuais).

Essa definição tem ainda a intenção de integrar a noção de modelagem matemática representada na Figura 1, aspecto central na concepção que o

⁴ Tradução livre. No original: “*For PISA, mathematics proficiency means the capacity of individuals to formulate, employ and interpret mathematics in a variety of contexts. The term describes the capacities of individuals to reason mathematically and use mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. Mathematics literacy is not an attribute that an individual either has or does not have; rather, it is a skill that can be developed over a lifetime*” .

PISA tem dos alunos como solucionadores de problemas. Para resolver um problema contextualizado, o indivíduo deve aplicar um pensamento matemático e uma ação ao desafio apresentado, sendo que, na medida em que trabalha no problema, realiza processos e ativa sucessiva e simultaneamente suas capacidades fundamentais em Matemática, recorrendo a conteúdos matemáticos para encontrar a solução.

Figura 1. Modelo de letramento em Matemática na prática.



Fonte: Relatório Nacional do PISA 2012 (INEP, 2013).

A parte mais externa da Figura 1 mostra que o letramento em matemática se dá com um desafio ou problema situado no mundo real, organizado na matriz em duas categorias: contextos e conteúdos.

Os contextos identificam em que área da vida os problemas surgem, podendo ser de cunho pessoal, envolvendo desafios enfrentados pelo indivíduo, sua família ou seus pares; social, com foco em sua comunidade, seja ela local, nacional ou global; ocupacional, centrado no mundo do trabalho; ou científico, relacionado à aplicação da Matemática no mundo natural e à questões e tópicos da ciência e tecnologia.

Os conteúdos matemáticos são definidos em quatro grandes classes de fenômenos que a Matemática analisa: quantidade, incerteza e dados, mudanças e relações e espaço e forma. Esses conteúdos estão relacionados à vertentes curriculares familiares como números, álgebra e geometria.

A parte mais interna da Figura 1 representa as etapas do ciclo de modelagem matemática de forma simplificada e idealizada. Ao se deparar

com um “problema no contexto”, o indivíduo com letramento em Matemática tenta reconhecer os aspectos matemáticos e formular a situação matematicamente de acordo com os conceitos e relações identificadas e suposições simples realizadas. Dessa forma, transforma o “problema no contexto” em um “problema matemático”, passível de tratamento matemático. Para solucioná-lo, emprega conceitos, procedimentos, fatos e ferramentas matemáticas para obter “resultados matemáticos”. Esses resultados devem então ser interpretados nos termos do problema original devendo ser inseridos no contexto (“resultados no contexto”). Para tanto, é necessário avaliar os resultados matemáticos e sua razoabilidade dentro do contexto de determinado problema.

Os processos formular, empregar e interpretar são componentes centrais do ciclo de modelagem matemática e da definição de letramento em Matemática do PISA. No entanto, em uma avaliação, geralmente não é necessário que o aluno se envolva em todas as etapas desse ciclo. Nos itens do PISA, em alguns casos, partes significativas são apresentadas para que sejam manipuladas diretamente (por exemplo, representações matemáticas como gráficos e equações), de modo a responder uma questão ou chegar a uma conclusão.

Os resultados do PISA para o processo formular situações matematicamente indicam o quanto os alunos são efetivamente capazes de reconhecer e identificar oportunidades de utilizar a Matemática em situações-problema, estando aptos a fornecer as estruturas e representações matemáticas necessárias para transformar o problema contextualizado em uma forma amena de tratamento matemático.

Para o processo empregar, os resultados indicam o quanto os alunos são capazes de aplicar raciocínio matemático e utilizar conceitos, procedimentos, fatos e ferramentas matemáticas para chegar a soluções matemáticas, estabelecendo regularidades, identificando conexões e criando argumentos matemáticos.

Já para o processo interpretar, aplicar e avaliar resultados matemáticos, os resultados indicam o quanto os alunos são efetivamente capazes de refletir sobre conclusões e soluções matemáticas, interpretando-as no contexto de um problema do mundo real e determinando quando são razoáveis e fazem sentido dentro do contexto.

Cada um desses três processos é baseado nas sete capacidades matemáticas fundamentais: comunicar, representar, delinear estratégias, “matematizar”, raciocinar e argumentar, utilizar linguagem e operações simbólicas, formais e técnicas e utilizar ferramentas matemáticas. Essas capacidades cognitivas são utilizadas pelos indivíduos para compreender e interagir com o mundo de forma matemática ou para solucionar problemas.

À medida que o letramento em Matemática aumenta, o indivíduo desenvolve um nível maior dessas capacidades, que são indicadores da demanda cognitiva, estando associadas ao nível de dificuldade do item. Assim, os estudos do PISA utilizam a necessidade de ativação dessas capacidades como base para a descrição dos diferentes níveis de proficiência da escala de letramento em Matemática, reproduzida no Quadro 1.

Quadro 1. Escala de proficiência em Matemática (continua).

| Nível | Limite inferior de pontos | Características das atividades |
|-------|---------------------------|--|
| 6 | 669,3 | No Nível 6, os estudantes são capazes de conceituar, generalizar e utilizar informações com base em suas investigações e em modelagem de situações-problema complexas. Conseguem estabelecer ligações entre diferentes fontes de informações e representações, e de transitar entre elas com flexibilidade. Os estudantes situados neste nível utilizam pensamento e raciocínio matemáticos avançados. São capazes de associar sua percepção e sua compreensão a um domínio de operações e relações matemáticas simbólicas e formais, de modo a desenvolver novas abordagens e estratégias para enfrentar novas situações. Os estudantes situados neste nível são capazes de formular e comunicar com precisão suas ações e reflexões relacionadas a constatações, interpretações e argumentos, bem como de adequá-los às situações originais. |
| 5 | 607,0 | No Nível 5, os estudantes são capazes de desenvolver modelos para situações complexas e trabalhar com eles, identificando restrições e especificando hipóteses. Conseguem selecionar, comparar e avaliar estratégias adequadas de resolução de problemas para lidar com problemas complexos relacionados a esses modelos. Os estudantes situados neste nível são capazes de trabalhar estrategicamente, utilizando habilidades de pensamento e raciocínio abrangentes e bem desenvolvidas, representações conectadas de maneira adequada, caracterizações simbólicas e formais, e percepção relativa a essas situações. São capazes de refletir sobre suas ações e de formular e comunicar suas interpretações e seu raciocínio. |

Quadro 1. Escala de proficiência em Matemática (conclusão).

| Nível | Limite inferior de pontos | Características das atividades |
|-------------|---------------------------|---|
| 4 | 544,74 | No Nível 4, os estudantes podem trabalhar de maneira eficaz com modelos explícitos para situações concretas complexas, que podem envolver restrições ou exigir formulação de hipóteses. São capazes de selecionar e integrar diferentes representações, inclusive representações simbólicas, relacionando-as diretamente a aspectos de situações da vida real. Nesses contextos, os estudantes situados neste nível são capazes de utilizar habilidades desenvolvidas e raciocínio, com flexibilidade e alguma percepção. São capazes de construir e comunicar explicações e argumentos com base em interpretações, argumentos e ações. |
| 3 | 482,4 | No Nível 3, os estudantes são capazes de executar procedimentos descritos com clareza, inclusive aqueles que exigem decisões sequenciais. Conseguem selecionar e aplicar estratégias simples de resolução de problemas. Os estudantes situados neste nível são capazes de interpretar e utilizar representações baseadas em diferentes fontes de informação e de raciocinar diretamente a partir delas. Conseguem desenvolver comunicações curtas que relatam interpretações, resultados e raciocínio. |
| 2 | 420,1 | No Nível 2, os estudantes são capazes de interpretar e reconhecer situações em contextos que não exigem mais do que inferência direta. São capazes de extrair informações relevantes de uma única fonte e de utilizar um modo simples de representação. Os estudantes situados neste nível conseguem empregar algoritmos, fórmulas, procedimentos ou convenções de nível básico. São capazes de raciocinar diretamente e de fazer interpretações literais dos resultados. |
| 1 | 357,8 | No Nível 1, os estudantes são capazes de responder a questões definidas com clareza, que envolvem contextos conhecidos, nas quais todas as informações relevantes estão presentes. Conseguem identificar informações e executar procedimentos rotineiros de acordo com instruções diretas em situações explícitas. São capazes de executar ações óbvias e dar continuidade imediata ao estímulo dado. |
| Abaixo de 1 | | A OCDE não especifica as habilidades desenvolvidas |

Fonte: Relatório Nacional do PISA 2012 (INEP, 2013).

As escalas do PISA foram estabelecidas na primeira edição em que cada uma das áreas foi o foco da avaliação, quando a matriz de habilidades foi completamente desenvolvida e o domínio exaustivamente avaliado. Os pontos de referência foram definidos de forma que a média e o desvio-padrão igualmente ponderados dos países da OCDE que participaram e tiveram taxas de resposta aceitáveis fossem iguais a 500 e 100, respectivamente (OECD 2014b).

A escala geral de letramento em Matemática foi estabelecida em 2003, primeira vez em que esse foi o principal domínio avaliado. Para o PISA 2012, a definição dos seis níveis de proficiência em Matemática foi feita da mesma forma que os níveis correspondentes na escala do PISA 2003, variando do “Nível 1” (mais baixo) ao “Nível 6” (mais alto), embora suas descrições tenham sido atualizadas para refletir as novas categorias de processos matemáticos da matriz e o grande número de novos itens desenvolvidos (OECD 2014a).

Além da escala geral de letramento em Matemática, o PISA fornece ainda subescalas para os três processos e as quatro categorias de conteúdo já citadas, cujas descrições podem ser consultadas em OECD, 2014a.

É desejável que as avaliações de Matemática do PISA, tanto em papel quanto em computador, possuam um equilíbrio adequado de itens que reflitam os componentes da matriz de letramento descrita nesta seção.

Em relação aos processos, o objetivo é que a avaliação tenha peso igual entre os dois processos que envolvem fazer uma conexão entre o mundo real e o mundo da matemática (formular e interpretar – aproximadamente 25% da pontuação para cada) e o processo que requer que os alunos trabalhem em um problema formulado matematicamente (empregar – aproximadamente 50% da pontuação).

Em relação ao conteúdo, o objetivo é que haja uma distribuição balanceada entre as quatro categorias (aproximadamente 25% da pontuação para cada), uma vez que todas são importantes para formar cidadãos construtivos, engajados e reflexivos.

O mesmo é indicado em relação ao contexto, em que cada uma das quatro categorias deve contribuir com aproximadamente 25% da pontuação, abrangendo uma ampla gama de interesses individuais e situações que se espera encontrar no mundo real.

Os itens em cada categoria de processo, conteúdo e contexto devem possuir também diversos níveis de dificuldade e de demanda matemática.

Uma vez que a teoria de medida utilizada tem implicações na interpretação dos resultados das pesquisas (ALLEN & YEN, 2002), a próxima seção apresenta alguns princípios da TRI, metodologia utilizada para a análise dos dados do PISA (OECD, 2014b).

2.2 A TEORIA DA RESPOSTA AO ITEM (TRI)

Diferentemente da Teoria Clássica dos Testes – TCT, que se preocupa em explicar o escore total (soma das respostas corretas dadas a uma série de itens) (PASQUALI, 2008), a TRI considera como unidade básica de análise o item (ANDRADE, LAROS & GOUVEIA, 2010). Assim, é usada de forma complementar à TCT, superando algumas das suas limitações, como por exemplo, os parâmetros dos itens serem dependentes da amostra e a proficiência do indivíduo variar de acordo com o teste aplicado (HAMBLETON, SWAMINATHAN & ROGERS, 1991).

A TRI vem sendo aplicada em avaliações educacionais brasileiras desde 1995, quando foi utilizada no Sistema de Avaliação da Educação Básica (SAEB), sendo posteriormente implementada também no Exame Nacional para Certificação de Competências de Jovens e Adultos (ENCCEJA), Prova Brasil e Exame Nacional do Ensino Médio (ENEM), entre outras. No âmbito internacional vem sendo utilizada largamente por diversos países, a exemplo dos Estados Unidos, França, Holanda, Coreia do Sul, China, além dos países participantes do PISA (INEP, 2012).

O uso da TRI permite equalizar escores de diferentes testes ou formas alternativas de um mesmo teste e descrever e interpretar escores de diferentes testes em uma mesma escala (HAMBLETON, SWAMINATHAN & ROGERS, 1991), tornando-os comparáveis. Os valores obtidos são interpretados qualitativamente pela construção de escalas de proficiência, ferramenta básica para relatar os resultados das avaliações que utilizam essa metodologia. Andrade, Tavares e Valle (2000) definem a TRI como um conjunto de modelos matemáticos que representa a probabilidade de um examinando dar certa resposta a um item em função dos parâmetros do item e do(s) traço(s) latente(s) do respondente.

O traço latente representado na TRI pela letra grega θ é uma característica do indivíduo que não pode ser observada diretamente e que determina a forma como ele responde ao teste (ANDRADE, LAROS & GOUVEIA, 2010). Medir o nível do θ significa estimar a habilidade do indivíduo, no estudo em questão, em Matemática. Adicionalmente à estimação dos parâmetros dos itens (conhecida como calibração), a estimação da habilidade dos respondentes é uma das etapas mais importantes da TRI (ARAÚJO, ANDRADE & BORTOLOTTI, 2009). Geralmente, os itens de múltipla escolha em avaliação educacional são analisados utilizando modelos de um, dois ou três parâmetros, que envolvem apenas a dificuldade do item (parâmetro b), a dificuldade e a discriminação (parâmetro a) ou a dificuldade,

a discriminação e o acerto ao acaso (parâmetro c) (CONDÉ & LAROS, 2007).

O parâmetro de dificuldade do item (b) e o \square são mensurados na mesma escala, em que cada item está associado a um intervalo no qual tem maior poder de discriminação definido em torno do valor de b e cada examinando é posicionado de acordo com a sua habilidade (ANDRADE, TAVARES & VALLE, 2000). Por conveniência computacional, essa métrica possui média igual a 0 (zero) e desvio-padrão igual a 1 (um) e posteriormente é feita uma transformação linear das estimativas para a escala de habilidade desejada (ANDRADE, LAROS & GOUVEIA, 2010).

Andrade, Tavares e Valle (2000) destacam que a probabilidade de um indivíduo responder corretamente a um certo item é a mesma, independentemente da escala utilizada para medir a sua habilidade e, assim, não faz qualquer sentido analisar itens a partir dos valores de seus parâmetros sem conhecer a escala na qual eles foram determinados.

Uma vez especificada a escala, ela é interpretada a partir do posicionamento dos itens, permitindo uma análise qualitativa dos resultados encontrados. Dessa forma é possível saber, por exemplo, qual o conhecimento de um aluno que obteve a estimativa do traço latente 1, numa escala (0,1), quando o traço latente representa a proficiência em Matemática (ARAÚJO, ANDRADE & BORTOLOTTI, 2009).

Segundo Valle (2000) os vários modelos da TRI propostos na literatura dependem fundamentalmente (i) da natureza do item – dicotômicos ou não dicotômicos; (ii) do número de populações envolvidas; e (iii) da quantidade de traços latentes que está sendo medida. Nesse sentido, cabe observar que o PISA utiliza o modelo unidimensional de um parâmetro, conhecido como modelo de Rasch, que será discutido com mais detalhes na seção 3.2.3.

Para que se possa fazer uso dos modelos unidimensionais, diversos autores (HAMBLETON, SWAMINATHAN & ROGERS, 1991; ANDRADE, TAVARES & VALLE, 2000; PASQUALI, 2003; PASQUALI & PRIMI, 2003; CONDÉ & LAROS, 2007; ARAÚJO, ANDRADE & BORTOLOTTI, 2009) apontam dois pressupostos relevantes que devem ser observados: a independência local dos itens, que tem como princípio que o desempenho do indivíduo em um item não pode afetar o desempenho em outro item e a avaliação de um único traço latente, ou seja, deve haver apenas uma habilidade responsável pela realização do conjunto de itens. Essa última condição, a unidimensionalidade, será aprofundada a seguir.

2.2.1 A unidimensionalidade dos testes

Vitória, Almeida e Primi (2006), em uma revisão das principais definições e métodos de avaliação da unidimensionalidade evidenciaram a dificuldade de estabelecer critérios empíricos consensuais para esse conceito. Pasquali (2003) afirma que para satisfazer a condição de unidimensionalidade é suficiente admitir a existência de um fator dominante, que se supõe estar sendo medido pelo conjunto de itens. O autor observa que uma vez que o desempenho humano é sempre multideterminado ou multimotivado, mais de um traço latente estará presente na execução de qualquer tarefa.

Seguindo nessa linha, Reckase (2009) salienta que os itens de um teste podem ser sensíveis a diversos tipos de fatores, mas as dimensões de sensibilidade devem estar relacionadas com o que se pretende medir. Laros, Pasquali e Rodrigues (2000) chamam a atenção ainda para o fato de que “a ausência de unidimensionalidade nos itens de uma prova conduz à uma diminuição da validade de construto dificultando a interpretação de seus escores”.

Segundo Green (1968), espera-se que a transição de um teste em papel para computador não afete a sua validade. Ele pondera que se o modo de aplicação alterar apenas a escala (tornando o teste mais ou menos difícil) ou a distribuição das proficiências, os efeitos são mantidos essencialmente constantes e podem ser facilmente controlados para permitir uma interpretação válida dos resultados. Entretanto, se a tarefa for modificada, diferentes construtos serão avaliados, tornando baixa a correlação entre as proficiências e impactando na validade do construto.

Neste estudo, os termos comparabilidade e equivalência serão utilizados indistintamente, conforme Al-Amri (2009), que compreende ainda que, se um teste mede o mesmo construto, independente do seu modo de aplicação, em papel ou em computador, as duas versões podem ser consideradas equivalentes e, portanto, comparáveis.

Dado que a comparação entre os testes em papel e em computador, tanto dos parâmetros dos itens, quanto dos resultados encontrados só faz sentido quando se assume que ambos os testes estão avaliando o mesmo traço latente, considerou-se de extrema importância determinar em que medida o modo de aplicação poderia estar atuando concomitantemente com o fator dominante a que os itens estão relacionados.

Ademais, como afirma Reckase (2009), além da sensibilidade dos itens associada a uma habilidade cognitiva ou tipo de conhecimento, a unidimensionalidade também depende dos respondentes. Uma vez que o recorte dessa pesquisa se limitou a uma subamostra dos alunos brasileiros

participantes do PISA, a verificação desse pressuposto serviu como alicerce para discussão de todas as análises subsequentes.

De acordo com Pasquali (2003), atualmente o método que parece mais recomendável no contexto da análise da unidimensionalidade de uma série de itens é o chamado análise fatorial de informação completa (*Full Information Factor Analysis – FIFA*), proposto por Bock e Aitkin (1981). Vitória, Almeida e Primi (2006) explicam que esse método está baseado na TRI e não requer o cálculo das intercorrelações entre os itens, trabalhando ao invés disso com padrões distintos de resposta ao item, utilizando o modelo multifatorial de Thurstone (1947) baseado em estimativas de máxima verossimilhança marginal e no algoritmo EM (*Expectation-Maximization*) de Dempster, Laird e Rubin (1977).

As vantagens da utilização desse método são enumeradas por Laros, Pasquali e Rodrigues:

A análise fatorial *full information* supera muitos dos problemas que acompanham a análise fatorial de coeficientes de correlações tetracóricas: (1) ela evita o problema de coeficientes tetracóricos indeterminados de itens extremamente fáceis ou difíceis; (2) ela acomoda facilmente efeitos do acerto ao acaso e de itens não apresentados; (3) ela proporciona um teste de significância estatística para testar a propriedade dos fatores adicionados sucessivamente ao modelo (LAROS, PASQUALI & RODRIGUES, 2000, p. 30).

Por conseguinte, diante das recomendações encontradas na literatura, a análise de dimensionalidade proposta neste estudo será realizada por esse método.

2.3 AS AVALIAÇÕES EM COMPUTADOR

Segundo Choi et. al (2003), as avaliações em computador e os testes adaptativos foram foco de muitas pesquisas no campo da psicometria desde o começo dos anos 70. De lá para cá, com o avanço da tecnologia, essas avaliações passaram por diversas mudanças e tem se tornado uma alternativa cada vez mais viável às avaliações tradicionais em papel (LUECHT & SIRECI, 2011).

O sistema educacional dos Estados Unidos foi um dos primeiros, se não o primeiro, a conduzir avaliações *online* em larga escala com o *National Assessment of Educational Progress (NAEP)* e atualmente os componentes

digitais vêm sendo implementados em vários sistemas de avaliações nacionais em larga escala ao redor do mundo como parte de um processo de integração da tecnologia ao ensino e aprendizagem, a exemplo de países como Islândia, Dinamarca, Hungria, Israel e Austrália (BELLER, 2013).

No Brasil, além da aplicação dos opcionais do PISA 2012, algumas iniciativas indicam que o país já começou a dar os primeiros passos nessa direção. Por volta de 2005, o Centro de Seleção e Promoção de Eventos da Universidade de Brasília – CESPE/UnB iniciou alguns estudos na área de avaliação adaptativa computadorizada (FERNANDES, 2009). Desde então, a instituição vem desenvolvendo uma plataforma tecnológica para permitir a aplicação de testes nesse formato. O projeto evoluiu a partir das provas de proficiência em Inglês Instrumental realizadas para alunos da UnB, que foram respondidas pela primeira vez em julho de 2010 utilizando um protótipo dessa ferramenta. A intenção é adequar a plataforma para permitir a execução de avaliações *online* em todo o território nacional (CASTRO, 2013).

A Geekie, uma *startup* de tecnologia aplicada à educação, inaugurada em 2011, se baseia no conceito de aprendizado adaptativo, utilizando a tecnologia para personalizar o estudo de acordo com as características individuais de cada um. Para tanto, realiza diagnósticos através da aplicação de testes *online*, incluindo até mesmo simulados do ENEM, que permitem identificar quais são as dificuldades e níveis de proficiência dos alunos em diferentes assuntos (GEEKIE, 2015).

Esse cenário condiz com o que foi apresentado em 1998 por Bennett (1998), que descreve três gerações de avaliações em computador, distintas pelo propósito da avaliação, formato e conteúdo do teste, local de aplicação e extensão em que capitaliza a nova tecnologia:

A primeira geração estabelece a infraestrutura básica para a testagem eletrônica. Na segunda geração, as avaliações em larga escala passam por uma mudança qualitativa, mas os seus propósitos e mecanismos de aplicação continuam essencialmente os mesmos. A última geração traz um repensar sobre os propósitos e mecanismos das avaliações em larga escala (BENNETT, 1998, p. 2).⁵

⁵ Tradução livre. No original: “*The first generation lays the basic infrastructure for electronic testing. In the second generation, large-scale tests undergo qualitative change, but their purposes and delivery mechanisms remain essentially the same. The last generation brings a rethinking of the purposes and mechanisms of large-scale assessment.*”

Para Bennett (1998), as avaliações em computador da primeira geração são semelhantes às aquelas aplicadas em papel, combinando basicamente avanços da psicometria à tecnologia para desenvolver o CAT. De fato, a definição do método em que o teste será implementado é uma das inúmeras questões a serem consideradas para o desenvolvimento de um sistema de avaliação em computador. Hoje os testes em computador (*computer-based tests – CBTs*) são uma ampla indústria que engloba uma variedade de tipos de testes com desenhos de aplicação e tipos de itens apropriados para os mais diferentes propósitos e não há um modelo único adequado para todas as avaliações educacionais (LUECHT & SIRECI, 2011).

Nesse sentido, Luecht e Sireci (2011) fazem uma classificação extensiva dos métodos de aplicação, diferenciando os CBTs em oito grupos, de acordo basicamente com três aspectos: (1) uso de algoritmos adaptativos (2) tamanho das unidades de aplicação do teste e (3) natureza e extensão em que o teste é montado automaticamente. Já Parshall et. al (2002) fazem uma classificação mais simples, dividindo-os em apenas quatro grupos: “testes fixos computadorizados” (*computerized fixed test*), “testes de montagem automática para aplicação online” (*automated test assembly for online delivery*); “testes adaptativos computadorizados” (*computerized adaptive tests*) e “testes computadorizados para classificação” (*computerized classification tests*).

A afirmação de Luecht e Sireci (2011) de que “a tecnologia fundamental que distingue vários modelos de CBTs é o grau em que o teste é adaptativo”⁶ e o fato dos dois primeiros métodos descritos por Parshall et. al (2002) serem lineares e os dois últimos adaptativos, evidenciam o uso primário dessa característica para classificação dos CBTs.

Os testes lineares ou testes fixos computadorizados são similares aos testes tradicionais em papel, com tamanho e formato definidos, o que permite que sejam desenvolvidos em um prazo relativamente curto. Tipicamente, o que ocorre é que os testes em papel são apenas transferidos para um ambiente computadorizado, podendo ser utilizadas múltiplas formas fixas, contanto que sejam desenhadas com propriedades estatísticas ou escores comparáveis (PARSHALL et. al, 2002).

Já o CAT se baseia na TRI para adaptar o teste à habilidade de cada respondente, o que requer um banco robusto de itens calibrados por nível de dificuldade. Tipicamente, o teste tem início com um item de nível de dificuldade intermediário. Caso o aluno responda corretamente, uma questão

⁶ Tradução livre. No original: “A fundamental technology that distinguishes among many CBT models is the degree to which the test is made adaptive.”

mais difícil lhe é apresentada ou vice-versa, permitindo gerar uma estimativa mais precisa do seu nível de proficiência. Ao contrário dos testes convencionais em papel, no CAT os itens apresentados, a ordem de apresentação e, usualmente, a quantidade de itens variam para cada respondente, tornando esse tipo de teste mais eficiente tanto na confiabilidade dos scores, quanto na economia de tempo (PARSHALL et. al, 2002; LUECHT & SIRECI, 2011).

2.3.1 Vantagens e desvantagens dos CBTs

Um estudo de De-Siqueira et. al (2009) realizado com professores e alunos da Universidade Politécnica de Valencia, na Espanha, mostrou que a maioria deles preferem escrever nos CBTs se comparado aos testes em papel (*paper-based tests – PBTs*), o que lhes permite economizar um tempo razoável. Também consideraram ser esse o modo de aplicação mais apropriado, apesar de se sentirem inseguros quanto à falhas no computador e à realização de testes pela internet.

Outra grande vantagem dos CBTs é a rapidez com que os resultados podem ser divulgados e os diferentes tipos de *feedback* que podem ser fornecidos tanto para os alunos quanto para os aplicadores. Tais testes permitem registrar uma série de informações que ajudam a identificar as estratégias e interações dos examinandos (ordem em que as questões são respondidas, respostas modificadas, tempo de resposta por item, movimento entre questões, etc.) e analisar esses dados proporciona uma compreensão mais rica do aprendizado (ALDERSON, 2000; CONOLE & WARBURTON, 2005).

Alderson (2000) se aprofunda nessa questão e afirma que o uso dos CBTs também tem grande impacto pedagógico, podendo encorajar a incorporação e integração das avaliações diretamente aos processos de ensino e aprendizagem. Ele argumenta que os CBTs oferecem ao desenvolvedor do teste a possibilidade de proporcionar um retorno mais significativo para os examinandos, uma vez que pode ser apresentado a qualquer momento, seja depois de decorrido certo tempo, ao término de todo o teste, ao final de uma parte específica do teste ou após cada item ser respondido, podendo permitir até mesmo uma segunda tentativa, com ou sem prejuízo da pontuação.

Ainda segundo Alderson (2000), os computadores também podem propiciar uma forma mais amigável de teste ao incluir algumas facilidades, como, por exemplo, dicionários e ajuda *online* para esclarecer instruções e dar exemplos e/ou dicas do que deve ser feito. O pesquisador pontua que o monitoramento do uso de tais suportes pode ser levado em consideração no resultado ou pontuação do teste, bem como o nível de confiança do

examinando, medido através do questionamento sobre o quanto ele se sente confiante com sua resposta (associação que pode evidenciar o chute).

Vários são os artigos que citam como uma das vantagens dos CBTs a flexibilidade no agendamento do teste (PEARSON, 2009a; LOTTRIDGE et. al, 2008). Ao contrário dos PBTs que normalmente exigem a definição de locais e datas fixas, os CBTs, a depender das suas características, possibilitam ao examinando escolher quando e onde realizar o teste, o que pode acarretar também na redução de custos com a logística de aplicação.

A propósito, fazer a transição dos PBTs para os CBTs implica em diferenças significativas em termos de custos. Parshall et. al (2002) menciona esse como um dos aspectos positivos dos CBTs, uma vez que não há a necessidade de imprimir as provas, distribuí-las e retorná-las para correção. No entanto, Bennet (1998) observa que como muitas implementações iniciais, a primeira geração de CBTs, que consiste na simples transposição dos itens em papel para o computador, pode elevar os custos das avaliações de alto impacto, em que o banco de itens deve ser continuamente alimentado para apoiar a administração contínua. Al-Amri (2009) pondera ainda a carga extra, imposta aos locais que não possuem os requisitos tecnológicos necessários para a aplicação. Adicionalmente, as despesas com *softwares* comerciais e suas licenças foram apontadas por Conole e Warburton (2005) como uma das maiores barreiras institucionais para a adoção dos CBTs. Como devidamente concluído por Farcot e Latour (2009, p. 115), “a diversidade dos CBTs é tão grande que é ilusório tentar encontrar uma resposta única, geral e transponível relativa à eficiência dos custos dos CBTs em oposição aos PBTs”⁷.

A possibilidade de desenvolver o CAT é outra vantagem dos CBTs comumente citada na literatura (AL-AMRI, 2009; PEARSON 2010). Esse formato torna o teste mais eficiente e reduz o tempo necessário para sua aplicação, posto que o examinando responde somente itens próximos ao seu nível de proficiência. Assim reforça também a segurança, já que cada um realiza um teste diferente. Entretanto, um dos problemas do CAT consiste no fato de não permitir que as respostas sejam alteradas após ser exibido o próximo item (ALDERSON, 2000).

Apesar de possibilitarem o desenvolvimento de itens inovadores, capazes de avaliar habilidades difíceis ou impossíveis de serem aferidas em papel, para que os resultados sejam rapidamente reportados, os testes geralmente são automaticamente corrigidos, limitando os tipos de itens que

⁷ Tradução livre. No original: “the diversity of CBA is so large that searching for a unique, general and transposable answer concerning cost efficiency of CBA as opposed to P&P is misleading”.

podem ser utilizados. Itens de múltipla escolha são quase onipresentes nos CBTs, embora dependendo do que se pretenda avaliar outras técnicas sejam mais apropriadas, sendo, no entanto, mais difíceis de serem implementadas (ALDERSON, 2000; CONOLE & WARBURTON, 2005).

Moe (2009) ainda destaca que para o desenvolvimento de CBTs é necessário trabalhar de perto com diferentes grupos: professores, os que constroem a plataforma técnica, os estatísticos, pessoas com conhecimento de aprendizagem eletrônica e o artista que desenha as figuras, se tornando um grande desafio a integração de profissionais de meios tão distintos.

Ainda que o uso das TICs seja cada vez mais comum, os CBTs requerem habilidades mínimas, como o uso do teclado e mouse, que podem torná-los desfavoráveis em relação aos PBTs. Ler um texto no computador, rolar a página ou mover entre telas não é o mesmo que passar páginas impressas em papel (ALDERSON, 2000) e é preciso considerar o impacto dessas diferenças no desempenho do examinando.

2.3.2 Métodos e modelos para comparabilidade

A Agência Educacional do Texas (*Texas Education Agency – TEA*) destaca em seu relatório técnico (TEA, 2008) que o primeiro e mais crítico passo em um estudo de comparabilidade é coletar bons dados para conduzir as análises e embasar as conclusões, de maneira que a comparação entre os modos de aplicação seja válida. O documento cita ainda três desenhos experimentais mais comumente utilizados para esse fim: com o mesmo indivíduo, com grupos aleatoriamente equivalentes e *quasi-experimental*. Analogamente, Lottridge et. al (2008) faz uma classificação mais geral, distinguindo esses desenhos apenas por dentro-indivíduos e entre-indivíduos.

O desenho dentro-indivíduos é aquele em que o mesmo aluno responde ao teste nos dois formatos, em papel e em computador, compostos ou não pelos mesmos itens. Segundo Lottridge et. al (2008), nesse desenho a ordem de aplicação dos testes normalmente é contrabalanceada para minimizar efeitos como fadiga, prática ou motivação. Embora a sua aplicação seja um tanto quanto limitada por problemas administrativos decorrentes da dupla testagem, esse desenho requerer uma amostra menor e é a fonte mais rica de informações sobre comparabilidade, capaz de detectar pequenas diferenças entre os modos de aplicação, uma vez que elimina a influência de fatores adicionais relacionados às características do aluno. Contudo, cabe observar que, caso sejam utilizadas provas diferentes, deve-se garantir a equivalência entre elas, e, do contrário, caso as provas utilizadas sejam iguais, deve-se considerar a influência da memória, ponderando também uma

possível aprendizagem no intervalo de tempo entre as aplicações (TEA, 2008; LOTTRIDGE et. al, 2008; PEARSON, 2009b).

O desenho entre-indivíduos é aquele em que os alunos são divididos em dois ou mais grupos e cada grupo responde aos mesmos itens nos dois formatos, em papel e em computador, incluindo variações nas técnicas utilizadas para composição dos grupos (LOTTRIDGE et. al, 2008). Nessa definição podemos encaixar tanto os desenhos denominados de grupos aleatoriamente equivalentes quanto os desenhos *quasi*-experimentais, mencionados no Relatório Técnico da TEA (2008) e em Pearson (2009b), com a diferença que, nesse último desenho, os grupos não são necessariamente equivalentes no momento da coleta dos dados.

Lottridge et. al (2008), faz uma revisão dos métodos utilizados na literatura para investigar a comparabilidade dos PBTs e CBTs e afirma que os pesquisadores têm usado uma série de desenhos experimentais e análises para verificar os diferentes aspectos relacionados ao tema. Os autores descrevem ainda as principais questões metodológicas que devem ser consideradas no desenho desse tipo de estudo, dividindo-as em quatro grandes dimensões: amostragem, instrumentos, aplicação e correção, organizadas por fatores que influenciam a validade externa (generalização de resultados) e interna (interpretação dos resultados). Nesse sentido, dado que esta pesquisa se baseia em dados secundários, o desenho experimental utilizado é o empregado pelo PISA, que será caracterizado no capítulo 3.

Os tipos de desenhos e análises que podem ser pensados, em termos das hipóteses a serem testadas sobre a equivalência dos escores e do construto também foram discutidos por Lottridge et. al (2008) e estão resumidos na Tabela 2.

Tabela 2. Desenho e análise do estudo por tipo de questão de comparabilidade (continua).

| Foco | Desenho | Análise sugerida |
|---|----------------------------|---|
| Comparabilidade de escores | | |
| Os escores devem ter a mesma precisão de medida | Entre ou dentre indivíduos | * Valores gerais de confiabilidade * Erros padrão de medida (SEM) * SEM Condicional / Informação do teste |
| A distribuição dos escores deve ser diferente apenas em dificuldade e, consequentemente, ser equivalentes | Entre ou dentre indivíduos | * Distribuição de frequência dos escores ou histograma * Medidas de dispersão e tendência central (média, desvio padrão) * Tabelas de escore bruto – escore escalar * Distribuição de examinandos por níveis de desempenho |

Tabela 2. Desenho e análise do estudo por tipo de questão de comparabilidade (conclusão).

| Foco | Desenho | Análise sugerida |
|---|---|--|
| Comparabilidade de construto | | |
| O conteúdo do teste e as especificações de conteúdo devem ser comparáveis | Nenhum, entre ou dentre indivíduos | * Diagrama de comparações * Revisão de modificações de item CBT que requer habilidades adicionais, realizada por especialistas * Comparações de diferentes formatos de teste |
| Os escores devem ter a mesma estrutura fatorial | Entre ou dentre indivíduos | * Teste de dimensionalidade a nível de item * Teste de dimensionalidade a nível de grupos de itens * Comparações dos parâmetros CBT e PBT a nível de item (TRI, TCT) |
| Os escores devem ser altamente correlacionados uns aos outros | Dentre indivíduos | * Correlação (corrigida e não corrigida para desconfiança) * Acordo de atribuição dentro dos níveis de desempenho |
| Os escores devem ter a mesma relação com outras medidas relacionadas | Entre ou dentre disciplinas, a critério | * Correlação (corrigida e não corrigida para confiabilidade) * A teoria deve conduzir à inclusão de medidas (i.e., ansiedade do teste influenciando as pontuações PBT) * Outras medidas podem incluir testes de habilidade, medidas de atitude, características dos examinandos (como sexo e raça) |

Fonte: Lottridge et. al, 2008, p. 29 (tradução livre).

Esses dois focos de análise, comparabilidade dos escores e do construto, são também comentados por Mc Donald (2002), que, no entanto, pondera que as investigações estatísticas tem ignorado amplamente o fato do teste em computador proporcionar uma experiência de teste qualitativamente diferente do teste em papel. Ele apresenta um modelo de como essa experiência pode impactar a equivalência estatística dos testes em papel e em computador:

“Neste modelo, o desempenho no teste resulta de uma interação entre o examinando e o teste em si, aqui está o aspecto chave se o teste é aplicado em papel ou em computador. Fora desta interação existe também a experiência do indivíduo na situação de teste, que pode variar de acordo com uma série de fatores, incluindo o formato do teste e suas reações a ele. Se essas reações individuais diferirem como resultado da variação no formato do teste, irão afetar o constructo medido por cada teste e, assim sua equivalência estatística” (p. 302).⁸

Em seu estudo, Al-Amri (2009) acrescentou ainda uma nova dimensão a esse modelo: a equivalência dos processos cognitivos, ou seja, das estratégias usadas pelos examinandos para a resolução do teste em cada modo de aplicação. Mc Donald (2002) reflete também sobre algumas diferenças físicas existentes entre PBTs e CBTs que podem afetar o desempenho do indivíduo (apresentação do texto, habilidade de revisar e pular questões) e faz uma revisão da literatura a respeito das principais características individuais que têm sido apontadas como fatores que contribuem para a experiência do teste, como a familiaridade e experiência com computadores, “ansiedade com computadores” (*computer anxiety*) e “atitudes em relação a testes em computador” (*computer attitudes*).

A discussão dos modelos utilizados nos estudos de comparabilidade se torna relevante na medida em que a presente pesquisa se propõe a verificar a comparabilidade dos testes em papel e em computador no contexto das avaliações educacionais em larga escala, tendo como base as provas de Matemática do PISA 2012 aplicadas no Brasil. Entretanto, diante da amplitude do tema, torna-se inviável aprofundar todos os pontos aqui indicados. Ainda assim, no intuito de abranger os principais aspectos de validade e confiança, o delineamento proposto tem como foco a comparabilidade de escores e de construto, no que se refere à distribuição dos examinandos por níveis de desempenho e à estrutura fatorial dos testes, sem a

⁸ Tradução livre. No original: “*In this model, test performance results from an interaction between the test taker and test itself, here the key aspect of the test being whether it is paper- or computer-based. Out of this interaction also comes the individual's experience of the testing situation which can vary according to a range of factors including the format of the test and their reactions to this. If these individual reactions differ as a result of the varying test format, they will affect the constructs measured by each test and so their statistical equivalence*” .

pretensão de esgotar as potenciais fontes de desigualdade que podem ser investigadas.

Vale destacar que a metodologia adotada tanto para elaboração dos instrumentos e sua aplicação, quanto para a análise dos resultados limitam algumas das abordagens sugeridas pelos autores citados, refletindo as particularidades dessa avaliação, que serão discutidas mais adiante. Lottridge et. al (2008) observam que, enquanto limitações como essas influenciarem a habilidade de identificar o impacto do modo de aplicação e a generalização dos resultados, será importante continuar conduzindo tais estudos, devendo o pesquisador esclarecer qualquer limitação no desenho experimental e realizar as análises levando-as em consideração.

3 METODOLOGIA

Este capítulo apresenta a estrutura do material cognitivo do PISA 2012, bem como os procedimentos metodológicos adotados para sua aplicação que são relevantes para este estudo, incluindo a amostragem e o modelo da TRI utilizado para estimação dos parâmetros dos itens e para o cálculo dos valores plausíveis. Na sequência é descrita a técnica utilizada para o estudo da dimensionalidade das provas em papel e em computador considerando o conjunto de respostas dos alunos brasileiros e as etapas compreendidas na comparação das médias e dos itens dentro da escala.

3.1 CARACTERIZAÇÃO DA PESQUISA

Uma das questões básicas no desenvolvimento de uma pesquisa é a definição dos procedimentos necessários para atingir seus objetivos, o que permite caracterizá-la utilizando diferentes critérios.

Demo (1995) afirma que podemos reconhecer, pelo menos, quatro gêneros mais delineáveis de pesquisa: teórica, metodológica, empírica e prática. O presente estudo pode ser classificado como teórico e prático, uma vez que busca aprimorar os fundamentos teóricos sobre a comparabilidade dos testes educacionais em papel e em computador com o intuito de contribuir para a fundamentação da aplicação de provas semelhantes nos diferentes formatos.

Diante desse contexto, para melhor descrever os processos envolvidos, a forma de abordagem do problema foi definida como quali-quantitativa. Segundo Silva e Menezes (2005), a pesquisa quantitativa classifica e analisa informações e opiniões traduzidas em números, enquanto a pesquisa qualitativa tem como focos principais a interpretação de fenômenos e a atribuição de significados. Neste estudo, a construção das interpretações pedagógicas está associada ao emprego de técnicas estatísticas para o tratamento dos dados.

Gil (2002) classifica as pesquisas quanto aos seus objetivos gerais e aos procedimentos técnicos utilizados. Por ter como objetivo primordial estabelecer a relação entre o formato da prova e o desempenho dos alunos com base em modelos padronizados e, por utilizar os referenciais teóricos do PISA, este estudo pode ser considerado como descritivo, compreendendo os procedimentos técnicos de pesquisa bibliográfica e experimental.

3.2 CARACTERIZAÇÃO DO PISA 2012

O PISA é uma avaliação metodologicamente complexa e, por essa razão, a primeira etapa desta pesquisa consistiu em fazer uma revisão detalhada dos seus principais referenciais teóricos, especialmente os de 2012, com o intuito de entender melhor não apenas sua estrutura e os procedimentos adotados para sua aplicação, como também as técnicas utilizadas para composição do banco de dados. As informações apresentadas nesta seção foram retiradas dos documentos da OECD (2009, 2012, 2013, 2014a, 2014b) e no Relatório Nacional do PISA 2012 (INEP, 2013) contendo os resultados do Brasil.

3.2.1 Escopo e amostragem

O PISA 2012 teve Matemática como foco principal e Leitura e Ciências como áreas menores de avaliação. Isso significa dizer que esse domínio foi avaliado com mais detalhes, permitindo um estudo mais aprofundado de suas subáreas. Esse ciclo incluiu também, como opcionais, uma avaliação em papel de Letramento Financeiro e avaliações em computador de Matemática, Leitura e Resolução de Problemas, tendo o Brasil optado por participar apenas dos opcionais em computador no estudo definitivo.

O público-alvo do PISA são os estudantes na faixa dos 15 anos de idade – mais especificamente que possuem entre 15 anos e 3 meses e 16 anos e 2 meses completos no mês de aplicação, com no máximo um mês de variação – que tenham completado pelo menos seis anos de escolarização, independente do tipo de instituição em que estejam matriculados, seja em tempo integral ou meio período. De acordo com o último critério, no Brasil foram elegíveis para participar do PISA 2012 todos os alunos matriculados do 7º ano do Ensino Fundamental em diante⁹.

Em 2012, participaram do PISA aproximadamente 510.000 alunos, representando cerca de 28 milhões de estudantes nos 65 países/economias participantes. Em geral, em cada país/economia, entre 4.500 e 10.000 alunos foram selecionados para a avaliação. No Brasil, a amostra foi ampliada para permitir a obtenção de resultados regionais, tendo sido selecionados para a

⁹ Devido à ampliação do ensino fundamental brasileiro de 8 para 9 anos, os alunos matriculados no 7º ano, apesar de terem feito parte da amostra, não foram incluídos no banco de dados, uma vez que poderiam ter iniciado seus estudos no sistema anterior, e, portanto, possuir escolaridade correspondente à 6ª série, estando fora do critério de elegibilidade estabelecido para o PISA.

avaliação em papel 24.954 alunos de 902 escolas, sendo que, dessas, uma subamostra de 247 escolas foram selecionadas também para a avaliação em computador.

A amostra do PISA é estratificada em dois estágios. O primeiro estágio consiste na seleção sistemática das escolas que possuem alunos elegíveis, realizada pelo consórcio internacional a partir de uma lista geral fornecida pelos países participantes. Para tanto, é utilizado o método de probabilidade proporcional ao tamanho considerando a estimativa do número de alunos elegíveis matriculados em cada escola.

A estratificação das escolas é feita de acordo com a meta de alunos por escola (*Target Cluster Size – TCS*), que em 2012 foi de 35 alunos. Assim, as escolas são classificadas como “grandes” (com quantidade de alunos elegíveis maior que o *TCS*), “moderadamente pequenas” (com quantidade de alunos elegíveis no intervalo entre metade do *TCS* e o *TCS*) e “muito pequenas” (com quantidade de alunos elegíveis menor que metade do *TCS*). Em 2012, as escolas “muito pequenas” foram ainda divididas em escolas com zero, um ou dois alunos elegíveis e escolas com mais de dois e menos de 18 alunos elegíveis (metade do *TCS*), tornando possível a adoção de procedimentos com o objetivo de minimizar problemas causados por uma possível amostra com muitas escolas pequenas.

Antes do sorteio as escolas são organizadas em grupos mutuamente exclusivos, de acordo com as características estabelecidas nos estratos explícitos, o que permite aprimorar a precisão das estimativas. O desenho da amostra brasileira em 2012 considerou como estratos explícitos as 27 Unidades da Federação e a dependência administrativa (pública federal/estadual, pública municipal ou privada) e, como substratos, a dependência administrativa (pública federal ou estadual), a localização (urbana ou rural), o quintil do Índice de Desenvolvimento Humano (IDH) do Município e os níveis de ensino ofertados (Fundamental e/ou Médio ou Profissionalizante). Por questões operacionais de logística ou de adequação dos instrumentos, foram excluídas as escolas indígenas, as escolas rurais da região Norte¹⁰ e as escolas exclusivamente de educação especial¹¹. A taxa combinada de exclusão de escolas e alunos manteve-se abaixo do critério de 5% da população-alvo, que evita distorções significativas nas médias de desempenho nacional.

¹⁰ Para a divulgação de alguns resultados, posteriormente também foram excluídas as escolas rurais dos demais estados.

¹¹ Os alunos com necessidades especiais matriculados em escolas regulares foram incluídos no universo para sorteio e o apoio necessário para a aplicação foi provido pela escola.

Em princípio, todas as escolas da amostra participaram também da avaliação em computador nos países que fizeram essa opção, tendo sido facultado àqueles com uma amostra grande como o Brasil, a aplicação em apenas uma subamostra. Nesse caso, foi retirada uma subamostra das escolas amostradas em cada estrato explícito.

O segundo estágio da amostragem consiste no sorteio dos alunos elegíveis que estão matriculados nas escolas selecionadas. Uma vez definidas, as escolas são contatadas para envio de uma relação contendo os nomes de todos os alunos que atendem aos critérios de elegibilidade. Esses dados são inseridos no *KeyQuest*, *software* fornecido pelo consórcio internacional que faz o sorteio e atribui ao aluno o modelo de prova e questionário que ele deverá responder. Na edição de 2012, tipicamente foram selecionados 35 alunos por escola (ou todos, quando a lista era menor) para a avaliação em papel, usando a amostragem aleatória simples. E, ainda, nas escolas selecionadas para a avaliação em computador, dessa amostra de alunos, foi retirada uma subamostra de 18 alunos para responder à prova também nesse formato.

Sendo a amostra do PISA desenhada em dois estágios, se torna necessário atribuir pesos diferentes para cada aluno e para cada escola, uma vez que: (i) eles não tem necessariamente a mesma probabilidade de seleção, (ii) dependendo do tipo de escola e característica dos alunos, as taxas de participação requerem diferentes ajustes de não-resposta e (iii) alguns estratos explícitos podem ter suas amostras ampliadas para permitir análises nacionais. O peso final de cada aluno, considerando o plano amostral descrito, é calculado pelo consórcio internacional e aparece como uma das variáveis da base de dados, devendo ser utilizado nas análises a fim de evitar viés nas estimativas.

A fim de minimizar o viés causado pela não-resposta, ou seja, pela ausência, os padrões de qualidade dos dados do PISA estabelecem taxas de participação mínimas tanto para escolas quanto para alunos. A taxa de participação mínima das escolas é de 85% da amostra original, sendo aceitável uma taxa entre 65% e 85%, se utilizadas escolas substitutas. A taxa de participação mínima dos alunos é de 80% a nível nacional e 50% dentro de cada escola.

No Brasil, ao final da aplicação, os dados consistidos mostraram que a prova em papel foi aplicada em 19.204 alunos de 840 escolas e a prova em computador em 5.506 alunos de 242 escolas. Para efeitos desta pesquisa, foram considerados apenas os alunos brasileiros das escolas urbanas e rurais efetivamente participantes e que responderam a prova de Matemática em papel e em computador, totalizando 1.390 alunos em 239 escolas. As subseções a seguir levam em consideração esse recorte, sendo muitas vezes

omitidas informações sobre o PISA não relevantes para o alcance dos objetivos propostos.

3.2.2 Aplicação e estrutura do material cognitivo

Os procedimentos para aplicação do PISA são descritos detalhadamente no manual do aplicador, elaborado pelo consórcio internacional de modo a uniformizar a aplicação em todos os países/economias participantes. As sessões de teste são conduzidas por um aplicador treinado, sendo a sessão de avaliação em papel dividida em duas partes: 2 horas de teste para avaliar conhecimentos e habilidades e 30 minutos para responder ao questionário do aluno. A sessão de avaliação em computador, com duração de 40 minutos, é aplicada posteriormente, não necessariamente no mesmo dia (no Brasil, essa aplicação ocorreu no dia seguinte). Em ambas as aplicações (em papel e em computador), os alunos recebem instruções padronizadas e são apresentados à questões de prática para se familiarizarem com o formato da prova e tipos de itens antes de iniciar o teste.

A seleção dos itens do PISA tem como base sua adequação à matriz e aos parâmetros de medida calculados pela aplicação de pré-testes realizados antes do estudo definitivo. Os itens são organizados em unidades com um estímulo comum, que pode ser de diferentes tipos: fragmentos de textos, tabelas, gráficos, diagramas ou uma combinação deles. Responder a uma série de itens relacionados permite que o aluno se envolva mais com o contexto ou problema apresentado, ao mesmo tempo em que favorece o uso mais eficiente do tempo do teste. A nomenclatura dos itens representa essa organização, sendo definida da seguinte forma:

- A primeira letra indica o formato (*P = paper; C = computer*);
- A segunda letra indica o domínio (*R = Reading; M = Math; S = Science; P = Problem Solving*);
- Os três números subsequentes indicam a unidade;
- A letra Q sucedida de dois algarismos indica o número da questão.

Cada unidade possui de um a cinco itens e é agrupada em blocos do mesmo domínio para composição dos cadernos (prova em papel) ou formulários de prova (prova em computador). A distribuição das unidades nos blocos leva em consideração a dificuldade dos itens e o tempo esperado para resposta.

Quando vários cadernos/formulários diferentes são utilizados, os dados coletados devem ser interligados de alguma forma para que os

resultados possam ser relatados em uma mesma escala. Para tanto, o PISA utiliza itens de ligação respondidos por diferentes alunos, de forma balanceada incompleta e independente para cada um dos dois modos de aplicação. Esse desenho garante que a resposta não seja influenciada pela respectiva localização do item nos diferentes cadernos/formulários. Apesar de não existirem itens comuns de Matemática em papel e em computador aplicados em 2012, uma vez que a população é a mesma, as duas escalas foram equalizadas para permitir a comparação desses resultados (OECD, 2014b).

Cada caderno/formulário é respondido por um número suficiente de alunos para que possam ser obtidas estimativas confiáveis para os níveis de proficiência dos alunos de cada país e em subgrupos relevantes dentro do país (meninos e meninas, alunos de diferentes contextos sociais e econômicos).

Tanto a prova em papel quanto a prova em computador incluem uma variedade de tipos de questões, que podem ser de resposta selecionada (múltipla escolha) ou de resposta construída.

As respostas selecionadas incluem itens de múltipla escolha simples, em que os alunos devem julgar a melhor resposta dentre as apresentadas (geralmente quatro ou em um número menor de casos, cinco) e múltipla escolha complexa, em que são apresentadas sentenças em que os alunos devem escolher entre duas ou mais respostas possíveis (exemplo: sim/não, verdadeiro/falso). Para os itens em computador, existem ainda as “**variações de respostas selecionadas**”, como a **seleção de opções** em um menu suspenso em que existe mais de uma caixa ou mais de uma resposta correta ou em que é possível selecionar mais de uma opção.

As respostas construídas incluem aquelas que podem ser codificadas manualmente pela comparação com uma única resposta correta (por exemplo, um único número ou frase simples) e aquelas que necessitam de codificação realizada por especialistas (por exemplo, respostas que incluem uma explicação ou um cálculo longo). Essas últimas permitem uma ampla gama de respostas aceitáveis e marcações mais complexas que podem incluir respostas parcialmente corretas (crédito parcial). As respostas construídas incluem ainda as auto-codificadas, quando a resposta não é baseada em textos, podendo ser dada, por exemplo, destacando-se segmentos de mapas para mostrar a melhor rota ou arrastando e deixando cair um objeto de um ponto a outro.

Cada questão possui um guia de codificação que consiste em uma lista de categorias de respostas (total, parcial e sem crédito), cada uma com seu próprio código de pontuação, descrição do tipo de respostas que deverão ser atribuídas a cada código e exemplos para cada categoria de resposta.

Uma das recomendações para a aplicação do PISA foi de que fosse fornecido aos alunos lápis, borracha, régua e até mesmo calculadora, que fica a critério dos países de acordo com o praticado rotineiramente nas salas de aula. Apesar de não ser exigida para resolução do teste, diferentemente dos anos anteriores, o uso da calculadora na edição de 2012 poderia favorecer os alunos em alguns itens de Matemática. Na prova em papel, o Brasil optou por distribuir estojos contendo esse material, sendo que, na prova de Matemática em computador esse recurso estava disponível para os alunos na própria tela, sendo também permitido o uso do dispositivo portátil. Em ambas as provas, as fórmulas necessárias foram fornecidas no início do caderno de questões ou a um clique, em um ícone acessível durante todo o teste.

3.2.2.1 Prova em papel

Em 2012, os itens em papel do estudo definitivo foram alocados em 15 blocos, compostos por 12 a 13 itens: nove blocos de Matemática (PM1 a PM5, PM6A, PM6B, PM7A e PM7B), três blocos de Ciências (PS1 a PS3) e três blocos de Leitura (PR1 a PR3), cada bloco representando 30 minutos de teste. A Tabela 3 apresenta o desenho utilizado para rotação dos blocos na composição dos cadernos, bem como o número de respondentes considerados para efeitos deste estudo em cada um dos cadernos aplicados no Brasil.

Tabela 3. Rotação dos blocos para composição dos cadernos da prova em papel do PISA 2012 e número de respondentes considerados para efeitos deste estudo em cada um dos cadernos aplicados no Brasil.

| Caderno | Bloco | | | | Número de respondentes |
|---------|-------|------|------|------|------------------------|
| | 1 | 2 | 3 | 4 | |
| B8 | PS2 | PR2 | PM4 | PS1 | 96 |
| B9 | PR2 | PM3 | PM5 | PR1 | 118 |
| B10 | PM3 | PM4 | PS3 | PM1 | 100 |
| B11 | PM4 | PM5 | PR3 | PM2 | 99 |
| B12 | PS1 | PR1 | PM2 | PS3 | 115 |
| B13 | PR1 | PM1 | PS2 | PR3 | 117 |
| B21 | PM5 | PS3 | PM6B | PS2 | 103 |
| B22 | PS3 | PR3 | PM7B | PR2 | 105 |
| B23 | PR3 | PM6B | PS1 | PM3 | 112 |
| B24 | PM6B | PM7B | PR1 | PM4 | 108 |
| B25 | PM7B | PS1 | PM1 | PM5 | 109 |
| B26 | PM1 | PM2 | PR2 | PM6B | 95 |
| B27 | PM2 | PS2 | PM3 | PM7B | 113 |

Fonte: Elaboração da autora (a partir de dados da OECD, 2014b, p. 32).

Dos nove blocos de Matemática, três foram compostos por itens de ligação utilizados em edições anteriores do PISA (PM1, PM2 e PM3), quatro foram denominados “blocos padrão” (PM4, PM5, PM6A e PM7A), contendo itens novos de diferentes níveis de dificuldade e dois blocos foram compostos por itens mais fáceis (PM6B e PM7B), calibrados pela TRI. Cada país/economia participante utilizou apenas sete desses blocos: três “blocos de ligação” e quatro “blocos padrão” ou três “blocos de ligação”, dois “blocos padrão” e dois “blocos fáceis”. Essa estratégia tem como objetivo obter informações descritivas melhores sobre o que os estudantes nos níveis mais baixos de proficiência sabem, bem como tornar a experiência do teste mais satisfatória para cada um deles individualmente, uma vez que respondem itens mais próximos ao seu nível de proficiência. Assim, países com médias abaixo de 450 são orientados a aplicar os “blocos fáceis” e o Brasil, seguindo o já estabelecido nas edições anteriores, fez essa opção, substituindo dois “blocos padrão” de Matemática (PM6A e PM7A) por outros dois “mais fáceis” (PM6B e PM7B), porém equivalentes em termos de aspecto e formato dos itens.

Em cada país foram aplicados 13 cadernos de prova diferentes, com quatro blocos cada, abrangendo uma ou mais áreas avaliadas em um total de duas horas de teste. Os blocos foram rotacionados de forma balanceada incompleta, com cada bloco aparecendo somente uma vez em cada uma das quatro posições possíveis e cada dupla de blocos aparecendo em apenas um caderno. Com 13 diferentes cadernos em um grupo de 35 alunos, não mais do que três alunos receberam o mesmo caderno em cada escola. Considerando o desenho utilizado por todos os países, para a avaliação de Matemática em papel foram selecionados 110 itens, em um total de 56 unidades, sendo 36 itens de ligação relacionados à edição de 2003 (utilizados também para interligar as edições de 2006 e 2009) e 74 itens novos, selecionados de um conjunto de 172 itens pré-testados um ano antes do estudo definitivo.

Sendo Matemática o foco do PISA 2012, os cadernos foram montados de forma a conter pelo menos um e no máximo três blocos desse domínio, ou seja, todos os alunos que participaram dessa edição da avaliação responderam de 12 a 37 itens de Matemática em papel.

3.2.2.2 Prova em computador

A avaliação em computador do PISA 2012 ofereceu um teste de Resolução de Problemas e um teste de letramento em computador (*CBAL*) composto por provas de Matemática e Leitura Digital. Os países participantes aplicaram uma das duas opções: apenas a avaliação de Resolução de Problemas ou a avaliação de Resolução de Problemas e o *CBAL*. Ao

contrário das edições anteriores do PISA, não foi exigida a participação do aluno no teste em papel para validar sua participação no teste em computador.

A Tabela 4 apresenta o desenho utilizado para rotação dos blocos na composição dos formulários, bem como o número de respondentes considerados para efeitos deste estudo em cada um dos formulários aplicados no Brasil.

Tabela 4. Rotação dos blocos para a composição dos formulários da prova em computador do PISA 2012 e número de respondentes considerados para efeitos deste estudo em cada um dos formulários aplicados no Brasil.

| Formulário | Bloco | | Número de respondentes |
|------------|-------|-----|------------------------|
| | 1 | 2 | |
| 41 | CP1 | CP2 | - |
| 42 | CR1 | CR2 | - |
| 43 | CM3 | CM4 | 115 |
| 44 | CP3 | CR1 | - |
| 45 | CR2 | CM2 | 110 |
| 46 | CM1 | CP4 | 116 |
| 47 | CR2 | CR1 | - |
| 48 | CM2 | CM1 | 125 |
| 49 | CP3 | CP4 | - |
| 50 | CM4 | CR2 | 122 |
| 51 | CP1 | CM3 | 114 |
| 52 | CR1 | CP2 | - |
| 53 | CM1 | CM3 | 111 |
| 54 | CP4 | CP1 | - |
| 55 | CR1 | CR2 | - |
| 56 | CP2 | CM4 | 110 |
| 57 | CR2 | CP3 | - |
| 58 | CM2 | CR1 | 112 |
| 59 | CP2 | CP3 | - |
| 60 | CM4 | CM2 | 122 |
| 61 | CR2 | CR1 | - |
| 62 | CM3 | CP1 | 118 |
| 63 | CR1 | CM1 | 115 |
| 64 | CP4 | CR2 | - |

Fonte: Elaboração da autora (a partir de dados da OECD, 2014b, p. 33).

Os itens em computador do estudo definitivo foram alocados em 10 blocos: quatro blocos de Resolução de Problemas (CP1 a CP4), quatro blocos de Matemática (CM1 a CM4) e dois blocos de Leitura Digital (CR1 a CR2),

cada bloco representando 20 minutos de teste. Foram utilizados 24 formulários diferentes, com dois blocos cada, abrangendo uma ou mais áreas avaliadas em um total de 40 minutos de teste. Os blocos foram rotacionados de tal forma que os alunos dos países que optaram por não participar do *CBAL* responderam apenas formulários que continham dois blocos de Resolução de Problemas, enquanto aqueles que fizeram essa opção responderam formulários contendo dois, um ou nenhum bloco desse domínio.

As unidades e os itens dentro unidades foram alocados em uma ordem fixa nos blocos e não foi permitido aos alunos retornar após terem passado para o próximo item/unidade. Cada vez que o aluno clicava no botão de navegação para o próximo item, uma mensagem era exibida para confirmar sua intenção, alertando-o de que caso prosseguisse não seria possível retornar, dando-lhe a possibilidade de cancelar a ação. Os itens de Matemática em computador foram apresentados em uma única tela de estímulo. Embora a prova fosse toda realizada no computador, os alunos puderam contar com “papel e caneta” para auxiliá-los em seu raciocínio.

A prova em computador do PISA foi realizada *offline* com o auxílio de *pen drives*, que armazenaram também os resultados da aplicação. Para tanto, foi utilizada uma plataforma de testes em computador de código aberto, TAO™ (*Testing Assisté par Ordinateur*) para facilitar a navegação dos alunos e fornecer todos os elementos de interface da prova (como a barra indicativa do progresso no teste). Dentre os requisitos mínimos de configuração necessários para rodar a prova, os computadores deveriam ter uma tela colorida com resolução de no mínimo 1024 x 768 pixels.

Além das respostas que contribuíram diretamente para estimar as proficiências dos alunos, o sistema coletou também informações comportamentais como a sequência de páginas visitadas e o uso dos elementos de estímulo como os menus suspensos. Em alguns casos, como nos itens em que a resposta era a própria interação, esses dados fizeram parte da própria pontuação.

A avaliação de Matemática em computador do PISA foi desenhada levando-se em consideração a necessidade de se distinguir entre as demandas matemáticas e aquelas relacionadas às TICs e ao novo formato de apresentação, de forma a assegurar que o raciocínio matemático tivesse precedência sobre a habilidade do aluno em utilizar o computador como ferramenta. Para estabelecer um controle sobre a ampla gama de características de um teste em computador, as demandas dos itens foram descritas envolvendo três aspectos:

- Demandas matemáticas (como nos itens em papel), compreendendo competências do letramento em Matemática que poderiam ser testadas em qualquer formato.

- Demandas relacionadas à interação da Matemática com as TICs, compreendendo conhecimentos matemáticos aplicados com o auxílio de computadores ou dispositivos portáteis. Tais competências foram testadas em alguns, mas não todos, os itens em computador e incluíram, por exemplo: a) a construção de gráficos (de setores circulares, de barra, de linha) a partir de dados (incluindo aqueles fornecidos em tabelas) utilizando um assistente simples; b) a construção de gráficos de funções e seu uso para responder questões a esse respeito; c) o planejamento e implementação de estratégias eficientes para classificação, localização e coleta de informações desejadas em uma planilha; d) o uso de calculadoras portáteis ou fornecidas na tela do computador; e) o uso de instrumentos virtuais, como réguas e transferidores; e f) a transformação de imagens utilizando uma caixa de diálogo ou *mouse* para girar, refletir ou traduzir a imagem.
- Competências em TIC compreendendo habilidades fundamentais necessárias para o uso do computador, incluindo conhecimentos básicos de *hardware* (por exemplo, *mouse* e teclado) e algumas convenções (por exemplo, setas para avançar e voltar e botões específicos para executar comandos). Essas habilidades foram intencionalmente mantidas a um nível mínimo em cada item.

Com o intuito de ajudar a monitorar o impacto (positivo ou negativo) do uso do computador no desempenho dos alunos, nem todos os itens foram interativos.

3.2.3 O PISA e a TRI

São necessários muitos itens para avaliar amplamente um domínio, como, por exemplo, o letramento em Matemática. Considerando que o PISA avalia vários domínios por ciclo, seria inviável aplicar um teste muito longo, razão pela qual apenas um conjunto de alunos responde a cada item, conforme descrito na seção anterior. Apesar de ser de fácil compreensão, a comparação de resultados de alunos submetidos a diferentes testes baseada na porcentagem de respostas corretas apresenta certa fragilidade, pois a diferença observada depende da dificuldade de cada teste.

Para superar esse desafio, o PISA se baseia na TRI para a análise dos resultados. Em particular, utiliza o modelo Rasch para itens dicotômicos, que estima a proficiência do aluno associando o número de respostas corretas à dificuldade do item, tornando os resultados totalmente comparáveis. Esse é

um dos modelos mais simples da TRI, com apenas um parâmetro (b) para o item, que representa o seu grau de dificuldade, expresso na mesma métrica/escala da proficiência (θ) do aluno. De acordo com a premissa desse modelo, para que o aluno tenha uma probabilidade de acerto de 0,5, sua posição na escala (θ) deve ser igual ao parâmetro (b) do item.

A probabilidade de um aluno j , com proficiência θ_j responder corretamente ao item i de dificuldade b_i é representada por (OECD, 2009):

$$P(u_{ij} = 1) = \frac{e^{\theta_j - b_i}}{1 + e^{\theta_j - b_i}}$$

onde u_{ij} é a resposta do aluno j ao item i (0 ou 1; erro ou acerto).

No processo de estimação dos parâmetros (b_i) dos itens desse modelo, também chamado de calibração, é utilizado o método da máxima verossimilhança marginal e a proficiência do aluno (θ_j) é estimada pelo método da máxima verossimilhança ponderada, que gera estimativas mais precisas da proficiência.

Os itens politômicos do PISA são calibrados pelo modelo de crédito parcial, uma extensão do modelo de Rasch para itens dicotômicos. De acordo com Andrade, Tavares e Valle (2000), supondo que o item i possui ($m_i + 1$) categorias de resposta ordenáveis, o modelo de crédito parcial é dado por:

$$P_{i,k}(\theta_j) = \frac{e^{\theta_j - b_{i,k}}}{\sum_{k=0}^{m_i} e^{\theta_j - b_{i,k}}}$$

com $i = 1, 2, \dots, I; j = 1, 2, \dots, n; k = 0, 1, \dots, m_i$ e $b_{i,0} = 0$, onde:

$P_{i,k}(\theta_j)$ é a probabilidade de um indivíduo com proficiência θ_j escolher a categoria k , dentre as ($m_i + 1$) categorias do item i ;

$b_{i,k}$ é o parâmetro do item que regula a probabilidade de escolher a categoria k em vez da categoria adjacente ($k - 1$) no item i .

Assim, para os itens do PISA com três categorias de resposta (incorreta, parcialmente correta e correta, ordenadas como 0, 1 e 2), são estimados 2 parâmetros de dificuldade para cada item.

Os parâmetros de dificuldade dos itens do PISA 2012 foram calculados pelo consórcio internacional usando um conjunto de dados para calibração de 31.500 alunos para a avaliação de Matemática em papel e

29.568 alunos para a avaliação de Matemática em computador, que representa uma seleção aleatória de 500 alunos de cada um dos países/economias participantes, exceto Chipre e Liechtenstein.

Para monitorar a qualidade dos dados e ajudar a decidir o tratamento que deve ser dado ao item nos diferentes países, o PISA realiza também calibrações nacionais separadas para cada país, usando dados não ponderados. Durante o processo de estimação, as médias dos países são fixadas em zero. Nos países que, como o Brasil, utilizam itens mais fáceis, os níveis de dificuldade nacionais sofrem uma transformação linear de forma que os itens padrão tenham a mesma média dos itens padrão obtida pela calibração da amostra dos países da OCDE. Algumas análises são utilizadas para verificar a necessidade de exclusão de um item da avaliação ou de um país específico, durante o processo de equalização.

O cálculo da proficiência dos alunos é baseado em valores plausíveis, que representam o intervalo de habilidades em que se localiza a verdadeira proficiência do aluno. Ao invés de calcular diretamente a proficiência, são estimados cinco valores plausíveis para a habilidade do aluno em cada escala de proficiência, associado à probabilidade de ocorrência de cada um desses valores. Os resultados são agregados para se obter as estimativas finais e seus respectivos erros padrão, podendo, no entanto, ser realizadas análises estatísticas independentes para cada um.

Um dos pressupostos da TRI e, portanto, do modelo de Rasch, se refere a independência local dos itens, ou seja, a probabilidade de um aluno responder corretamente a um item não pode ser influenciada pela resposta dada aos demais itens. No PISA, esse pressuposto é observado sempre que uma unidade compreendida por mais de um item é utilizada. Na prova em computador, duas características contribuem para garantir a independência dos itens: as unidades e os itens dentro das unidades são apresentados em uma ordem fixa, não permitindo que os alunos retornem depois de terem avançado para o próximo item/unidade e os alunos veem a mesma página no início de cada item da unidade, independente de onde tenham parado no item anterior.

3.3 METODOLOGIA PARA COMPARABILIDADE DOS TESTES

Para análise dos dados, foram considerados um total de 125 itens de Matemática em papel (84 itens) e em computador (41 itens), utilizados na aplicação do estudo definitivo do PISA 2012 no Brasil. Um item (PM603Q02) foi retirado da base de dados internacional devido à problemas com o nível de consistência exigido nas regras de codificação aplicadas em todos os países e, portanto, apesar de ter sido aplicado, não foi posicionado na escala do PISA, nem utilizado para o cálculo dos valores plausíveis (OECD, 2014b).

Para proceder as análises, os códigos representativos dos itens que foram excluídos após a avaliação devido à erros de impressão ou de tradução e das respostas ausentes e consecutivas no final da sessão de teste foram tratados como não apresentados, seguindo o mesmo procedimento adotado pelo PISA em edições anteriores (OECD, 2014b).

A comparabilidade das avaliações em papel e em computador foi explorada de três formas:

- verificação da equivalência dos construtos pela análise da dimensionalidade do conjunto de itens de Matemática em papel e em computador;
- identificação de possíveis fatores com impacto no desempenho dos alunos pelo cálculo das médias dos alunos considerando diferentes agrupamentos (por acesso à computadores, sexo, ano de ensino, localização da escola e região/estado);
- comparação do posicionamento dos itens em papel e em computador na escala de proficiência do PISA pela classificação dos itens quanto às categorias de conteúdo, contexto e processo e pela verificação dos parâmetros de dificuldade de itens semelhantes nos dois formatos.

3.3.1 Análise da dimensionalidade

A avaliação do número de dimensões do conjunto de dados foi feita pelo método de análise fatorial de informação completa, utilizando o pacote *mirt* (CHALMERS, 2012) implementado no *software R* (R CORE TEAM, 2012). Neste método, as probabilidades de acertar os itens são especificadas pelos modelos da TRI unidimensional e multidimensional compensatório. Para melhor interpretação dos fatores, foram realizadas rotações oblíquas (não-ortogonais) no espaço multidimensional, uma vez que todos os itens são

de Matemática, sendo mais natural que eventuais traços latentes estivessem correlacionados.

A fim de determinar o número de dimensões aferido pelos itens em computador e em papel, utilizou-se para comparação o teste da razão de máxima verossimilhança, também chamado de estatística qui-quadrado (X^2), que é dada por (REGAZZI & SILVA, 2004):

$$X^2 = \sum_{i=1}^n \sum_{r=1}^R \frac{(O_{ir} - E_{ir})^2}{E_{ir}}$$

onde:

- n é o número de avaliados;
- O_{ir} é a proporção de indivíduos que marcou o padrão r de resposta; e
- E_{ir} é a estimativa da probabilidade do avaliado marcar o padrão r de resposta, sob o modelo adotado.

O número de dimensões também foi avaliado com base nos critérios de informação bayesiano (*Bayesian Information Criterion – BIC*) (SCHWARZ, 1978) e no critério de informação de Akaike (*Akaike Information Criterion – AIC*) (AKAIKE, 1973), definidos por:

$$BIC = -2 \ln L(\hat{\theta}) + k \ln n$$

$$AIC = -2 \ln L(\hat{\theta}) + 2k$$

onde:

- k é o número de parâmetros a serem estimados;
- n é o tamanho da amostra; e
- $L(\hat{\theta})$ é o valor máximo da função de verossimilhança.

Tezza (2012) explica que estes critérios comparam o possível aumento de informação em um modelo com número de dimensão m em relação a um modelo com número de dimensões $m+1$, com base no valor de máxima verossimilhança e nos graus de liberdade, de tal forma que o melhor modelo é aquele que possui o menor valor dessas estatísticas.

Conforme descrito anteriormente, para efeitos desta pesquisa, foram considerados apenas os alunos brasileiros das escolas urbanas e rurais efetivamente participantes e que responderam a prova de Matemática em papel e em computador, num total de 1.390 alunos de 239 escolas.

Uma vez que o número de respostas por item foi muito pequeno, nesta análise foram utilizados apenas os 115 itens que possuíam percentual de respostas corretas em relação ao total de respostas válidas maior que 3%,

sendo eliminados os seguintes itens: PM00GQ01 (1,5%), PM406Q02 (2,4%), PM446Q02 (1,4%), PM462Q01 (1,1%), PM985Q03 (0%), PM992Q03 (2,4%), CM008Q02 (2,7%), CM015Q02 (3,0%), CM035Q04 (2,9%) e CM036Q02 (0,5%). A frequência de respostas por item na amostra considerada nesse estudo pode ser visualizada no Apêndice A.

Após a verificação da calibração dos itens, foram analisadas as cargas para cada fator. Itens com carga fatorial superior a 0,3 foram considerados adequados para aquela dimensão (PASQUALI, 2003). Cada grupo de itens característicos de uma dada dimensão foi avaliado quanto ao seu formato e demais características, considerando o tipo de prova e as classificações definidas na matriz de Matemática do PISA.

3.3.2 Cálculo das médias

É importante destacar que os valores plausíveis são números aleatórios desenhados a partir da distribuição de escores que podem ser razoavelmente atribuídos para cada indivíduo, não sendo notas e, portanto, não devendo ser tratados como tal (OECD, 2014b). Os valores plausíveis não são estatísticas adequadas para estimar parâmetros individuais e, assim, o cálculo das médias em papel e em computador realizado neste estudo foi limitado à agrupamentos populacionais com características que supostamente poderiam ter algum impacto no desempenho do aluno, como acesso à computadores, sexo, ano de ensino, localização da escola e região/estado.

O manual de análise dos dados do PISA (OECD, 2009) orienta que estimativas com esses valores devem ser realizadas cinco vezes, uma para cada valor plausível relevante para a análise, para que então possa ser feita uma média dos resultados e computadas suas variações. Para todos os alunos das escolas amostradas, a base do PISA continha cinco valores plausíveis para a avaliação de Matemática em papel e cinco valores plausíveis para a avaliação de Matemática em computador. As médias foram calculadas no SPSS, com o auxílio de macros disponibilizadas pelo consórcio internacional e destinadas a esse fim (disponíveis em: <http://www.oecd.org/pisa/pisaproducts/pisadataanalysismanualspssandsassecondedition.htm>).

3.3.3 Comparação do posicionamento dos itens

O posicionamento dos itens na escala de proficiência do PISA foi feito a partir dos parâmetros fornecidos pelo consórcio internacional. Com os parâmetros de dificuldade (b) já calculados e a métrica já definida, procedeu-se a uma interpretação pedagógica dos itens da categoria de conteúdo com

maior representação em 2012, a categoria *Espaço e Forma*, tendo como base a sua posição, formato e classificação nas categorias de processo e contexto, seguindo também a descrição das subescalas. A distribuição dos itens aplicados no Brasil de acordo com as categorias da matriz de letramento em Matemática e tipo de resposta pode ser visualizada na Tabela 5.

Tabela 5. Distribuição dos itens de Matemática em papel e em computador do PISA 2012 aplicados no Brasil, de acordo com as categorias da matriz de letramento em Matemática e do tipo de resposta.

| Categoria | Papel | Computador | Total |
|--|-------|------------|-------|
| Conteúdo | | | |
| Espaço e forma | 22 | 12 | 34 |
| Incerteza e dados | 19 | 9 | 28 |
| Mudanças e relações | 22 | 11 | 33 |
| Quantidade | 21 | 9 | 30 |
| Processo | | | |
| Empregar | 40 | 22 | 62 |
| Formular | 22 | 9 | 31 |
| Interpretar | 22 | 10 | 32 |
| Contexto | | | |
| Científico | 16 | 8 | 24 |
| Ocupacional | 18 | 9 | 27 |
| Pessoal | 17 | 13 | 30 |
| Social | 33 | 11 | 44 |
| Tipo de resposta | | | |
| Múltipla escolha complexa | 12 | 4 | 16 |
| Múltipla escolha simples | 25 | 8 | 33 |
| Resposta construída auto-codificada | 3 | 22 | 25 |
| Resposta construída codificada manualmente | 23 | 0 | 23 |
| Resposta construída codificada por especialistas | 21 | 4 | 25 |
| Variações de resposta de seleção | 0 | 3 | 3 |

Fonte: Elaboração da autora (com base nos dados da OECD, 2014b, Anexo A).

Para auxiliar nessa discussão, também considerou-se importante classificar os itens em computador de acordo com seu nível de interatividade (nenhuma, acessória e obrigatória), baseado no tipo de atividade que o aluno

deveria realizar para acessar os recursos disponíveis e que seria necessária para responder ao item.

A classificação dos itens de Matemática em computador nesses três níveis foi realizada com o auxílio de um especialista da área e seguiu os seguintes critérios:

- Itens não-interativos: itens estáticos que podem ser facilmente reproduzidos em papel e que não demandam do aluno nenhum tipo de interação (exemplo: item CM038Q05, vide Anexo);
- Itens com interação acessória: itens que demandam do aluno algum tipo de interação, podendo ou não ser necessária para sua resolução (exemplo: item CM038Q03, vide Anexo);
- Itens com interação obrigatória: itens cuja interação é a própria resposta (exemplo: item CM020Q03, vide Figura 5.3).

Os itens que ilustram as discussões deste estudo foram liberados pelo PISA para exemplificar os tipos de questões da avaliação e encontram-se disponíveis, na versão em Português, no site do INEP: <http://portal.inep.gov.br/internacional-novo-pisa-itens>. No Anexo foram reproduzidos todos os itens divulgados do estudo definitivo de 2012 aplicados no Brasil (exceto aqueles já destacados ao longo do texto), sendo um bloco inteiro de Matemática em papel (contendo 13 itens de cinco unidades diferentes) e três unidades de Matemática em computador (contendo dez itens).

4 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados e discutidos os resultados das análises descritas nos capítulos 1 e 3.

4.1 ANÁLISE DA DIMENSIONALIDADE DAS PROVAS

Nesta seção serão apresentados os resultados da análise da dimensionalidade dos itens das provas em papel e em computador do PISA 2012. Apesar dos resultados terem sido colocados em uma mesma escala, uma vez que este estudo se baseou em uma subamostra de 1.390 alunos brasileiros, considerou-se necessário avaliar o cumprimento desse pressuposto. O ponto de partida se deu pelo fato da unidimensionalidade ser dependente tanto da sensibilidade dos itens a apenas uma habilidade cognitiva ou tipo de conhecimento quanto dos respondentes (Reckase, 2009). Antes da execução das análises fatoriais, entretanto, foi realizada uma análise exploratória.

4.1.1 Análise exploratória

Como já relatado anteriormente, diferentemente da prova em papel, a prova em computador do PISA não permitiu que os alunos retornassem depois de terem avançado para o próximo item/unidade, tendo sido apresentada ainda uma barra para controle do tempo. Essas características podem ter efeito na administração do tempo pelos alunos e, conseqüentemente, influenciar no seu desempenho.

Para determinar a equivalência das provas nesse quesito, foi realizado o cálculo do percentual de alunos que responderam ao último item de cada caderno/formulário quando o bloco em questão era de Matemática. Assim, para essa análise foram considerados apenas os cadernos 10, 11, 23, 24, 25, 26 e 27 e os formulários 43, 45, 48, 51, 53, 56, 60 e 63. A Tabela 6 apresenta esses resultados.

Ao examinar os dados, verifica-se que a menor porcentagem de respostas ao último item do bloco da prova em papel foi de 63% no caderno B24, enquanto a menor porcentagem de respostas ao último item do bloco da prova em computador foi de 47% no formulário F63, sugerindo que o tempo fornecido para as respostas foi, em geral, insuficiente para o grau de dificuldade dos itens, com impacto ainda maior na prova em computador.

No entanto, é importante observar que o tempo limite de aplicação da prova em papel foi de duas horas para quatro blocos e o da prova em computador foi de quarenta minutos para dois blocos. Apesar dos blocos

dessas duas avaliações serem compostos por um número diferente de questões, o tempo médio disponível para responder a cada item nos dois formatos foi de dois minutos e meio para papel e dois minutos para computador, o que talvez justifique esse resultado.

Tabela 6. Percentual de respostas ao último item de Matemática dos cadernos em papel e formulários em computador aplicados no Brasil no PISA 2012.

| Papel | | | Computador | | |
|---------|-------------|-----------------------|------------|-------------|-----------------------|
| Caderno | Último item | Respostas válidas (%) | Formulário | Último item | Respostas válidas (%) |
| B24 | PM00KQ02 | 63,0 | F63 | CM002Q02 | 47,0 |
| B27 | PM967Q03 | 67,3 | F48 | CM002Q02 | 48,0 |
| B25 | PM998Q04 | 68,8 | F60 | CM038Q06 | 50,8 |
| B10 | PM034Q01 | 69,0 | F43 | CM014Q03 | 55,7 |
| B26 | PM934Q02 | 78,9 | F53 | CM008Q02 | 59,5 |
| B23 | PM800Q01 | 81,3 | F51 | CM008Q02 | 61,4 |
| B11 | PM564Q02 | 83,8 | F56 | CM014Q03 | 65,5 |
| | | | F45 | CM038Q06 | 76,4 |

Fonte: Elaboração da autora.

Ainda que um dos pressupostos implícitos da TRI seja de que os testes não podem ser aplicados sob condições de rapidez (HAMBLETON & SWAMINATHAN, 1985), não é possível afirmar que a falta de tempo influenciou no desempenho dos alunos em qualquer uma das provas. Tendo em vista o que afirma Oshima (2005) sobre o efeito desse fator ser maior na estimação dos parâmetros dos itens que na estimação das proficiências e, considerando ainda que o PISA é aplicado em diversos países e que esse estudo foi realizado com uma subamostra dos alunos brasileiros participantes da avaliação, qualquer conclusão nesse sentido seria precipitada. Dessa forma, decidiu-se proceder à análise de dimensionalidade admitindo-se que não houve violação desse pressuposto.

Uma vez que o número de respostas válidas por item foi muito pequeno, variando de 331 a 467, na análise de dimensionalidade foram utilizados apenas os 115 itens que possuíam percentual de respostas corretas em relação ao total de respostas válidas maior que 3%, por observar que os itens que não atendiam à esse critério apresentaram problemas de convergência na análise fatorial (Apêndice A). Assim, foram eliminados os seguintes itens: PM00GQ01 (1,5%), PM406Q02 (2,4%), PM446Q02 (1,4%),

PM462Q01 (1,1%), PM985Q03 (0%), PM992Q03 (2,4%), CM008Q02 (2,7%), CM015Q02 (3,0%), CM035Q04 (2,9%) e CM036Q02 (0,5%). Vale ressaltar que, dos seis itens em papel eliminados, cinco estavam entre os 13 com maior valor para o parâmetro b , enquanto os quatro itens em computador eliminados estavam entre os nove itens mais difíceis, o que talvez justifique o baixo percentual de respostas corretas.

Para obtenção das estimativas dos itens necessária para a análise fatorial, foi criada uma sintaxe no *software R*, usando o pacote *mirt*. Essa verificação visou tão somente identificar se algum item apresentava problema de convergência, condição em que deveria ser retirado da análise. Nenhum dos 115 itens (78 em papel e 37 em computador) apresentou problemas de convergência, sendo todos utilizados na análise fatorial (Apêndice B).

4.1.2 Determinação do modelo

Para a análise de dimensionalidade, as provas em papel e em computador foram consideradas como uma única prova, dado que o mesmo aluno respondeu itens nos dois formatos. O número de dimensões dessa prova única foi determinado por uma análise de variância (ANOVA), que retorna a razão de verossimilhança do teste qui-quadrado (X^2), bem como a diferença nos valores de AIC e BIC (CHALMERS, 2012). Foi feita a comparação de dois modelos, o primeiro assumindo uma dimensão (mod1) e o segundo, duas dimensões (mod2), cuja hipótese era estar relacionada ao formato dos itens (papel ou computador). A Tabela 7 apresenta esses resultados.

Tabela 7. Comparação dos modelos de uma e duas dimensões.

| Modelo | X^2 | Graus de liberdade | p-valor | AIC | BIC |
|--------|--------|--------------------|---------|-----------------|-----------------|
| Mod1 | | | | 44808,18 | 46080,78 |
| Mod2 | 285,66 | 114 | < 0,01 | 44750,51 | 46620,14 |

Fonte: Elaboração da autora.

Verifica-se pela Tabela 7 que, pelo teste da razão de verossimilhança, a diferença entre o modelo 1 e o modelo 2 é estatisticamente significativa ao nível de 5%, apontando para a escolha do modelo de duas dimensões, também indicada pelo AIC. Tais aspectos demonstram que a adição do segundo fator agrega ganho significativo de informação, sugerindo a melhor adequação do modelo bidimensional. No entanto, o mesmo não ocorre quando se observa o BIC, que sugere ser o

modelo unidimensional o mais adequado ao conjunto de itens, reforçando o que aponta Kang e Cohen (2007) sobre a tendência do BIC em selecionar modelos mais simples em relação ao AIC.

Dessa forma, considerando os objetivos e o recorte desta pesquisa, foi admitido *a priori* o modelo bidimensional, optando-se por realizar uma análise empírica que pudesse contribuir para a definição do número exato de dimensões. Assim, na próxima seção serão descritas e avaliadas não apenas as cargas fatoriais de cada item nos referidos modelos, mas também algumas tentativas de atribuir um significado claro para essas dimensões, de forma a enriquecer a discussão final dos resultados ora apresentados.

4.1.3 Análise fatorial

Tendo em vista que, independentemente do formato, todos os itens analisados são de Matemática, seria natural esperar que eventuais traços latentes existentes estivessem correlacionados, razão pela qual foi escolhido o método não-ortogonal de rotação *oblimin*, que indicou a correlação de 0,538 dos fatores no modelo bidimensional.

Com isso, pode-se afirmar que existe alta correlação entre os fatores (LAROS, 2004), o que corrobora a premissa assumida. As cargas fatoriais irão possibilitar a avaliação da contribuição individual de cada item em cada fator. Para que um item possa ser considerado na interpretação de uma dimensão, pesquisadores como Pasquali (2003) e Laros (2004) consideram o valor mínimo de 0,3, independente do sinal. No modelo unidimensional, itens que não possuem carga fatorial alta na dimensão não estão medindo o mesmo traço latente dos demais. No modelo bidimensional, essa análise deve ser feita para cada fator.

A Tabela 8 apresenta as cargas fatoriais dos itens em papel e em computador, considerando os modelos unidimensional e bidimensional (valores marcados em negrito indicam uma carga fatorial acima de 0,3), permitindo uma comparação do comportamento dos itens em cada um dos modelos.

No modelo bidimensional, dos 115 itens analisados, quinze itens não apresentaram carga acima de 0,3 em nenhum fator, sendo oito itens em papel (PM305Q01, PM411Q02, PM564Q01, PM800Q01, PM915Q01, PM934Q02, PM982Q04 e PM998Q04) e sete itens em computador (CM004Q01, CM005Q03, CM015Q01, CM020Q02, CM020Q04, CM036Q01, CM038Q03), o que representa 13% dos itens. No entanto, é importante observar que treze desses itens possuem carga fatorial acima de 0,22 em pelo menos um dos fatores, valor já considerado com alguma relevância, conforme critério adotado por Laros, Pasquali e Rodrigues (2000).

Tabela 8. Análise fatorial dos itens de Matemática do PISA 2012 (continua).

| Item | Modelo | | Item | Modelo | | F2 |
|----------|----------------|---------------|----------|----------------|---------------|---------------|
| | unidimensional | bidimensional | | unidimensional | bidimensional | |
| | F1 | F1 | | F1 | F1 | |
| PM00KQ02 | 0,605 | 0,750 | PM915Q02 | 0,742 | 0,176 | 0,740 |
| PM033Q01 | 0,516 | 0,383 | PM934Q01 | 0,721 | 0,858 | -0,189 |
| PM034Q01 | 0,631 | 0,373 | PM934Q02 | 0,319 | 0,275 | 0,066 |
| PM155Q01 | 0,658 | 0,801 | PM936Q01 | 0,757 | 0,587 | 0,257 |
| PM155Q02 | 0,779 | 0,670 | PM936Q02 | 0,797 | 0,593 | 0,302 |
| PM155Q03 | 0,887 | 0,919 | PM939Q01 | 0,577 | 0,466 | 0,179 |
| PM155Q04 | 0,457 | 0,228 | PM939Q02 | 0,725 | 0,574 | 0,233 |
| PM192Q01 | 0,647 | 0,656 | PM942Q01 | 0,356 | 0,432 | -0,088 |
| PM273Q01 | 0,518 | 0,330 | PM942Q02 | 0,784 | 0,758 | 0,043 |
| PM305Q01 | 0,447 | 0,297 | PM942Q03 | 0,832 | 0,738 | 0,148 |
| PM406Q01 | 0,822 | 0,632 | PM948Q01 | 0,576 | 0,065 | 0,665 |
| PM408Q01 | 0,573 | 0,415 | PM948Q02 | 0,593 | 0,311 | 0,384 |
| PM411Q01 | 0,838 | 0,642 | PM948Q03 | 0,795 | 0,660 | 0,209 |
| PM411Q02 | 0,248 | 0,140 | PM949Q01 | 0,572 | 0,626 | -0,030 |
| PM420Q01 | 0,588 | 0,393 | PM949Q02 | 0,501 | 0,383 | 0,172 |
| PM423Q01 | 0,563 | 0,545 | PM949Q03 | 0,746 | 0,609 | 0,198 |
| PM442Q02 | 0,703 | 0,371 | PM955Q01 | 0,519 | 0,511 | 0,037 |
| PM446Q01 | 0,747 | 0,572 | PM955Q02 | 0,629 | 0,617 | 0,026 |
| PM447Q01 | 0,523 | 0,592 | PM955Q03 | 0,703 | 0,765 | -0,118 |
| PM464Q01 | 0,856 | 0,712 | PM957Q01 | 0,421 | 0,560 | -0,164 |
| PM474Q01 | 0,541 | 0,429 | PM957Q02 | 0,430 | 0,673 | -0,309 |
| PM496Q01 | 0,679 | 0,598 | PM957Q03 | 0,771 | 0,789 | -0,007 |
| PM496Q02 | 0,595 | 0,578 | PM961Q02 | 0,802 | 0,813 | -0,001 |
| PM559Q01 | 0,421 | 0,178 | PM961Q03 | 0,599 | 0,288 | 0,435 |
| PM564Q01 | 0,304 | 0,282 | PM961Q05 | 0,510 | 0,358 | 0,220 |
| PM564Q02 | 0,438 | 0,393 | PM967Q01 | 0,726 | 0,541 | 0,274 |
| PM571Q01 | 0,509 | 0,583 | PM967Q03 | 0,576 | 0,601 | -0,027 |
| PM603Q01 | 0,610 | 0,522 | PM982Q01 | 0,526 | -0,237 | 0,913 |
| PM800Q01 | 0,402 | 0,237 | PM982Q02 | 0,624 | 0,250 | 0,503 |
| PM803Q01 | 0,809 | 0,549 | PM982Q03 | 0,466 | 0,178 | 0,382 |
| PM828Q01 | 0,701 | 0,054 | PM982Q04 | 0,380 | 0,271 | 0,161 |

Tabela 8. Análise fatorial dos itens de Matemática do PISA 2012 (conclusão).

| Item | Modelo unidimensional | | Modelo bidimensional | | Item | Modelo unidimensional | | Modelo bidimensional | |
|----------|-----------------------|--------------|----------------------|----------|--------------|-----------------------|---------------|----------------------|--|
| | F1 | F1 | F2 | F1 | | F1 | F2 | | |
| PM828Q02 | 0,614 | 0,076 | 0,698 | PM985Q01 | 0,607 | 0,392 | 0,315 | | |
| PM828Q03 | 0,605 | 0,098 | 0,664 | PM985Q02 | 0,662 | 0,476 | 0,278 | | |
| PM906Q01 | 0,530 | 0,249 | 0,370 | PM991Q01 | 0,543 | 0,449 | 0,141 | | |
| PM906Q02 | 0,730 | 0,367 | 0,490 | PM991Q02 | 0,770 | 0,800 | -0,039 | | |
| PM909Q01 | 0,681 | 0,685 | 0,030 | PM992Q01 | 0,475 | 0,216 | 0,354 | | |
| PM909Q02 | 0,538 | 0,404 | 0,200 | PM992Q02 | 0,624 | 0,200 | 0,555 | | |
| PM909Q03 | 0,874 | 0,536 | 0,488 | PM998Q02 | 0,615 | 0,818 | -0,204 | | |
| PM915Q01 | 0,117 | 0,288 | -0,182 | PM998Q04 | 0,166 | 0,132 | 0,050 | | |
| CM002Q01 | 0,832 | 0,787 | 0,096 | CM015Q01 | 0,434 | 0,254 | 0,258 | | |
| CM002Q02 | 0,544 | 0,571 | 0,004 | CM015Q03 | 0,827 | 0,496 | 0,470 | | |
| CM004Q01 | 0,358 | 0,151 | 0,295 | CM016Q01 | 0,753 | 0,760 | 0,041 | | |
| CM004Q02 | 0,590 | 0,614 | -0,023 | CM020Q01 | 0,495 | 0,301 | 0,279 | | |
| CM004Q03 | 0,664 | 0,780 | -0,135 | CM020Q02 | 0,042 | -0,132 | 0,220 | | |
| CM005Q01 | 0,415 | 0,346 | 0,121 | CM020Q03 | 0,634 | 0,458 | 0,259 | | |
| CM005Q02 | 0,339 | 0,333 | 0,036 | CM020Q04 | 0,297 | 0,292 | 0,023 | | |
| CM005Q03 | 0,203 | 0,004 | 0,258 | CM025Q01 | 0,685 | 0,289 | 0,542 | | |
| CM005Q04 | 0,708 | 0,338 | 0,507 | CM025Q02 | 0,692 | 0,410 | 0,396 | | |
| CM006Q01 | 0,711 | 0,609 | 0,185 | CM028Q03 | 0,580 | 0,498 | 0,131 | | |
| CM006Q02 | 0,663 | 0,484 | 0,267 | CM035Q01 | 0,502 | 0,375 | 0,196 | | |
| CM006Q03 | 0,699 | 0,964 | -0,306 | CM035Q02 | 0,506 | 0,569 | -0,049 | | |
| CM008Q01 | 0,547 | 0,223 | 0,441 | CM035Q03 | 0,338 | 0,302 | 0,069 | | |
| CM011Q01 | 0,535 | 0,384 | 0,217 | CM036Q01 | 0,200 | -0,012 | 0,271 | | |
| CM011Q02 | 0,784 | 0,719 | 0,115 | CM036Q03 | 0,145 | 0,405 | -0,311 | | |
| CM011Q03 | 0,737 | 0,644 | 0,153 | CM038Q03 | 0,299 | 0,111 | 0,252 | | |
| CM014Q01 | 0,542 | 0,644 | -0,119 | CM038Q05 | 0,518 | 0,167 | 0,483 | | |
| CM014Q02 | 0,690 | 0,544 | 0,230 | CM038Q06 | 0,738 | 0,633 | 0,171 | | |
| CM014Q03 | 0,622 | 0,714 | -0,127 | | | | | | |

Fonte: Elaboração da autora.

Do ponto de vista estatístico, verifica-se pelas cargas fatoriais dos itens em cada uma das dimensões que 83 itens (72,2%) se ajustaram ao fator 1 e 31 itens (27%) se ajustaram ao fator 2, sendo que desses, 14 itens (12,2%) se ajustaram aos dois fatores. O formato dos itens não contribuiu para a definição das dimensões, uma vez que a distribuição dos itens em papel e em computador em cada um dos fatores foi muito semelhante. Dos 78 itens em papel, 56 itens (71,8%) se ajustaram ao fator 1, 23 itens (29,5%) se ajustaram ao fator 2 e, desses, nove itens (11,5%) se ajustaram aos dois fatores. Dos 37 itens em computador, 27 itens (73%) se ajustaram ao fator 1, oito itens (21,6%) se ajustaram ao fator 2 e, desses, cinco itens (13,5%) se ajustaram aos dois fatores.

Levando-se em consideração que quanto mais alto o valor da carga fatorial, melhor o item representa o fator (LAROS, 2004), foi feita ainda a análise dos itens que mais contribuíram para a mensuração de cada fator. Em relação ao fator 1, o item em papel que mais contribuiu para a mensuração do fator foi o item PM155Q03, com carga fatorial de 0,919, enquanto em computador foi o item CM006Q03, com carga fatorial de 0,964. Em relação ao fator 2, o item em papel que mais contribuiu para a mensuração do fator foi o item PM982Q01, com carga fatorial de 0,913, enquanto em computador foi o item CM025Q01, com carga fatorial de 0,542.

Observa-se que as cargas fatoriais dos itens que mais contribuíram para a mensuração do fator 1 tanto em papel quanto em computador apresentaram valores muito próximos, indicando haver uma alta correlação entre o que está sendo medido em ambos os itens, independente do seu formato. Os valores das cargas fatoriais dos itens que mais contribuíram para a mensuração do fator 2, apesar de serem significativamente diferentes em papel e em computador, são representantes úteis do fator, corroborando a ideia de que o modo de aplicação não teve impacto no traço latente medido.

Quando considerado o modelo unidimensional, dos 115 itens analisados, apenas nove itens não apresentaram carga acima de 0,3, sendo três itens em papel (PM411Q02, PM915Q01, PM998Q04) e seis itens em computador (CM005Q03, CM020Q02, CM020Q04, CM036Q01, CM036Q03, CM038Q03), o que representa 7,8% dos itens. Observa-se que oito dos nove itens que não se ajustaram ao modelo unidimensional também não haviam se ajustado a nenhum fator do modelo bidimensional, a exceção do item CM036Q03. Assim, fica evidente que o não ajuste desses itens independe do modelo escolhido, estando provavelmente relacionado à proficiência dos respondentes da amostra. Dessa forma, se as provas tivessem sido aplicadas apenas no Brasil, seria sugerida a exclusão desses itens no intuito de contribuir para um aumento na validade do construto das provas, o que melhora as estimativas das proficiências dos alunos e dos parâmetros dos

itens, aperfeiçoando o processo de equalização e diminuindo o número de itens com viés (LAROS, PASQUALI & RODRIGUES, 2000).

Como todos os itens analisados estão relacionados ao mesmo domínio (Matemática), também foi feita a verificação da distribuição dos itens nos dois fatores de acordo com características como tipo de resposta, conteúdo, contexto e processo, ainda buscando validar as dimensões e auxiliar na sua interpretação. A Tabela 9 apresenta esse resultado, contabilizando duas vezes os itens que possuem carga fatorial alta nos dois fatores, de forma que a porcentagem total ultrapassa 100% .

Tabela 9. Distribuição dos itens nas duas dimensões de acordo com o tipo de resposta, conteúdo, contexto e processo.

| | N | Fator 1 (%) | | Fator 2 (%) | | Nenhum (%) | |
|--|----|-------------|---------|-------------|--------|------------|--------|
| Tipo de resposta | | | | | | | |
| Múltipla escolha complexa | 16 | 10 | (62,5) | 2 | (12,5) | 4 | (25,0) |
| Múltipla escolha simples | 33 | 18 | (54,5) | 6 | (18,2) | 11 | (33,3) |
| Resposta construída auto-codificada | 21 | 19 | (90,5) | 7 | (33,3) | 0 | (0,0) |
| Resposta construída codificada manualmente | 21 | 14 | (66,7) | 9 | (42,9) | 0 | (0,0) |
| Resposta construída codificada por especialistas | 21 | 19 | (90,5) | 6 | (28,6) | 0 | (0,0) |
| Variações de resposta de seleção | 3 | 3 | (100,0) | 1 | (33,3) | 0 | (0,0) |
| Conteúdo | | | | | | | |
| Espaço e forma | 31 | 23 | (74,2) | 4 | (12,9) | 5 | (16,1) |
| Incerteza e dados | 28 | 17 | (60,7) | 9 | (32,1) | 5 | (17,9) |
| Mudanças e relações | 29 | 23 | (79,3) | 10 | (34,5) | 2 | (6,9) |
| Quantidade | 27 | 20 | (74,1) | 8 | (29,6) | 3 | (11,1) |
| Contexto | | | | | | | |
| Científico | 22 | 13 | (59,1) | 8 | (36,4) | 3 | (13,6) |
| Ocupacional | 25 | 20 | (80,0) | 10 | (40,0) | 1 | (4,0) |
| Pessoal | 27 | 24 | (88,9) | 3 | (11,1) | 3 | (11,1) |
| Social | 41 | 26 | (63,4) | 10 | (24,4) | 8 | (19,5) |
| Processo | | | | | | | |
| Empregar | 59 | 42 | (71,2) | 13 | (22,0) | 9 | (15,3) |
| Formular | 24 | 17 | (70,8) | 8 | (33,3) | 3 | (12,5) |
| Interpretar | 32 | 24 | (75,0) | 10 | (31,3) | 3 | (9,4) |

Fonte: Elaboração da autora.

A partir dos dados constata-se que, para todas as características estudadas, os itens estão bem distribuídos entre os fatores, de forma condizente com o quantitativo em cada categoria. Assim, na impossibilidade de atribuir um significado claro para as duas dimensões, considerou-se que os oito itens cujas cargas fatoriais não contribuíram para mensuração dos fatores em nenhum dos dois modelos poderiam estar originando possíveis dimensões espúrias.

Dado que a classificação do conteúdo de metade desses itens pertence à subárea Incerteza e Dados, apontada por Nunes (2013) como pouco explorada no ensino das escolas brasileiras de nível fundamental, a heterogeneidade da população em relação aos níveis das dimensões pode ainda ter contribuído para que as respostas dos alunos fossem influenciadas por fatores alheios ao conhecimento de Matemática, sendo esse um fator importante na definição do número de traços latentes a serem considerados.

Fragoso (2010) afirma que “escolher um número que superestime a quantidade de traços latentes não é recomendável devido à dificuldade na interpretação de cada traço”. Tendo em vista que o equacionamento das divergências de resultados do teste qui-quadrado e do critério AIC em relação ao critério BIC foi feito com base nas características dos itens associados a cada dimensão, ponderando-se o escopo desta pesquisa, podemos supor que o número de dimensões subjacente ao conjunto de itens analisados resultou em uma estrutura predominantemente unidimensional.

4.2 COMPARAÇÃO DAS MÉDIAS

Assim como na análise de dimensionalidade descrita na seção 4.1, a comparação das médias de desempenho dos alunos nas provas em papel e em computador do PISA 2012 teve como base apenas uma subamostra de 1.390 alunos brasileiros que efetivamente responderam à prova de Matemática nos dois formatos. Em decorrência disso, a média de desempenho do Brasil nessas duas avaliações foi recalculada, resultando em um aumento da média nacional na prova em papel, de 391 pontos para 393 pontos e uma diminuição da média nacional na prova em computador, de 421 pontos para 418 pontos, o que representa uma atenuação da diferença entre elas de 30 pontos para aproximadamente 25 pontos na escala do PISA.

Como explicado anteriormente, uma vez que a proficiência no PISA é expressa em termos de valores plausíveis, a comparação das médias de desempenho dos alunos foi feita apenas para parâmetros populacionais e não para escores individuais. De acordo com Al-Amri (2009), para verificar a comparabilidade dos escores em papel e em computador é essencial garantir que o mesmo construto esteja sendo medido nos dois formatos, sendo a familiaridade dos alunos com computadores uma das variáveis cruciais, ainda que irrelevante para o propósito da avaliação.

Embora um questionário opcional sobre TICs tenha feito parte do PISA 2012 (OECD, 2014b), o Brasil optou por não aplicá-lo, o que não permitiu traçar um perfil mais detalhado dos alunos brasileiros nesse quesito. De todo modo, o questionário contextual do aluno, obrigatório para todos os países, continha duas questões referentes à posse e quantidade de computadores que o aluno possuía em casa, visando o cálculo do seu nível socioeconômico. Considerando que o Panorama Setorial da Internet (CETIC.br, 2013) revelou que as atividades mais realizadas com o apoio do computador e da Internet pelos alunos brasileiros do ensino fundamental e médio são desenvolvidas principalmente nesse ambiente, esse parece ser um bom indicativo da habilidade dos alunos com essas tecnologias. Por essa razão, as respostas dos alunos às referidas questões foram relacionadas às suas médias de desempenho, no intuito de verificar a influência dessa variável. Tais questões, assim como apareceram no questionário, foram reproduzidas na Figura 2, omitindo-se os subitens não utilizados neste estudo.

Figura 2. Questões do questionário do aluno utilizadas para verificar o impacto da posse e uso de computadores nas médias de desempenho do Brasil nas provas de Matemática em papel e em computador do PISA 2012.

| Qual dos seguintes itens você possui em casa? | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|
| <i>(Marque apenas uma opção em cada linha)</i> | | | | |
| | | <i>Sim</i> | <i>Não</i> | |
| Um computador que você pode utilizar para os trabalhos escolares | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| Dos itens relacionados abaixo, <u>quantos</u> existem na sua casa? | | | | |
| <i>(Marque apenas uma opção em cada linha)</i> | | | | |
| | <i>Nenhum</i> | <i>Um</i> | <i>Dois</i> | <i>Três ou mais</i> |
| Computadores | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Fonte: INEP, c2011 (adaptado pela autora).

Dos 1.390 alunos da amostra, 1.322 forneceram respostas válidas - respostas em branco não foram consideradas como respostas válidas - para a primeira questão, sendo que 956 alunos declararam possuir computador em casa disponível para realizar trabalhos escolares, o que representa 72,3%. Na segunda questão, 1.334 alunos forneceram respostas válidas, sendo que 340 alunos (25,5%) declararam não possuir nenhum computador em casa; 544 alunos (40,8%) declararam possuir um; 262 alunos (19,6%) declararam possuir dois e 188 alunos (14,1%) declararam possuir três ou mais computadores em casa disponíveis para realizar trabalhos escolares.

Esses dados refletem o que fora apresentado na Pesquisa TIC Domicílios e Empresas 2013 (CETIC.br, 2014a), que identificou tendência de crescimento na proporção de domicílios brasileiros com computador, seja de mesa, portátil ou *tablet*, atingindo um percentual de 49% em 2013. No entanto, nos domicílios com crianças em idade escolar, esse número se mostrou superior à média nacional, com o computador presente em 70% dos domicílios de alunos das escolas públicas localizadas em áreas urbanas e 97% dos domicílios de alunos das escolas particulares (CETIC.br, 2014b).

Embora a proporção de alunos com acesso a um computador em casa mostre-se significativa, a Pesquisa TIC Educação 2013 (CETIC.br, 2014b) também apontou que, mesmo tendo incorporado o uso das TICs às suas atividades cotidianas, muitos alunos “ainda apresentam dificuldades na realização de atividades de ensino-aprendizagem utilizando computador e Internet” (p. 161). O impacto da posse e quantidade de computadores no

desempenho dos alunos nas provas de Matemática em papel e em computador do PISA 2012 é apresentado na Tabela 10.

Tabela 10. Médias de desempenho dos alunos brasileiros nas provas de Matemática em papel e em computador do PISA 2012 em função das respostas a duas perguntas do questionário do aluno.

| | N | % | Papel | | Computador | | Dif. |
|---|-------|------|-------|------|------------|------|-------|
| | | | Média | E.P. | Média | E.P. | |
| Brasil | 1.390 | 100 | 393,3 | 4,1 | 418,0 | 4,3 | -24,7 |
| Posse de um computador em casa para realizar os trabalhos escolares | | | | | | | |
| Não | 366 | 27,7 | 348,4 | 5,5 | 371,5 | 5,5 | -23,1 |
| Sim | 956 | 72,3 | 417,1 | 4,8 | 441,7 | 4,9 | -24,6 |
| Quantidade de computadores em casa | | | | | | | |
| Nenhum | 340 | 25,5 | 349,6 | 6,3 | 370,4 | 5,6 | -20,8 |
| Um | 544 | 40,8 | 395,0 | 5,1 | 418,5 | 6,1 | -23,5 |
| Dois | 262 | 19,6 | 431,4 | 8,4 | 458,7 | 10,9 | -27,3 |
| Três ou mais | 188 | 14,1 | 465,0 | 10,9 | 491,7 | 10,4 | -26,7 |

Fonte: Elaboração da autora.

De maneira geral, observa-se que a posse de um ou mais computadores teve um impacto positivo no desempenho dos alunos em Matemática, contribuindo para um aumento das médias tanto em papel quanto em computador à medida que uma quantidade maior desse equipamento estava disponível no domicílio do aluno, embora não possa ser feita uma relação causal direta. Esses resultados são condizentes com os achados de Del Porto e Ferreira (2007), que demonstraram que essa variável teve o segundo maior poder explicativo para o desempenho dos alunos em Matemática no PISA 2003.

Seguindo o desempenho nacional dos alunos, o desempenho na prova em computador foi superior ao desempenho na prova em papel em todas as médias calculadas. Em relação à primeira pergunta do questionário, as diferenças de desempenho nas provas em papel e em computador para cada uma das categorias de resposta apresentaram valores semelhantes entre si e muito próximos à média nacional, independente do aluno possuir ou não um computador em casa para realizar os trabalhos escolares. Todavia, quando considerada a quantidade de computadores em seu domicílio, essa diferença aumentou de forma gradual até o máximo de dois computadores, ponto a partir do qual houve uma pequena diminuição.

Cabe ressaltar que embora aparentemente divergentes, os resultados das duas questões podem estar relacionados à própria forma como as perguntas foram elaboradas. Possuir um computador em casa não significa que ele esteja disponível para o aluno realizar os trabalhos escolares. Ademais, mesmo essa condição sendo satisfeita, o aluno pode não utilizá-lo ou não o fazê-lo para tal finalidade. Ou seja, responder “não” à primeira questão não implica necessariamente que o aluno não tenha acesso a essa tecnologia em casa. Contudo, por ser menos específica nesse sentido, a segunda questão evidencia mais efetivamente as diferenças existentes. Verifica-se, portanto, que os alunos que tem ao menos um computador em casa têm médias de desempenho nas provas em papel e em computador mais distantes uma da outra quando comparada com as médias de desempenho nas provas em papel e em computador dos alunos que não possuem nenhum computador em casa.

De todo modo, não podemos afirmar apenas com essa análise que a familiaridade com computadores seja a responsável pelos resultados observados, que possivelmente estão fortemente relacionados a fatores socioeconômicos que impactam o desempenho. Além disso, ter ou não acesso a um computador pode ter ainda outras consequências como, por exemplo, a própria ansiedade do aluno diante da experiência de responder ao teste em um novo formato.

Nesse sentido, Cheek e Agruso (1995) indicam que alunos que possuem computadores em casa demonstram atitudes mais positivas em relação ao seu uso, destacando que os meninos têm mais interesse em utilizar o computador como um fim em si mesmo, enquanto as meninas preferem utilizá-lo como um meio para realizar determinada tarefa. De fato, a diferença entre os sexos também é uma das características individuais do aluno frequentemente exploradas na literatura sobre comparabilidade. Assim, buscou-se fazer a comparação das médias de desempenho dos alunos da amostra considerando também essa variável. A amostra foi composta por 661 meninos (47,6%) e 729 meninas (52,4%) e os resultados são apresentados na Tabela 11.

Tabela 11. Médias de desempenho dos alunos nas provas de Matemática em papel e em computador do PISA 2012 em função do sexo.

| Sexo | N | Papel | | Computador | | Dif. |
|-----------|-------|-------|------|------------|------|-------|
| | | Média | E.P. | Média | E.P. | |
| Brasil | 1.390 | 393,3 | 4,1 | 418,0 | 4,3 | -24,7 |
| Feminino | 729 | 383,4 | 5,0 | 408,2 | 4,9 | -24,8 |
| Masculino | 661 | 403,9 | 4,9 | 428,3 | 5,5 | -24,4 |

Fonte: Elaboração da autora.

A partir dos dados verifica-se que os meninos tiveram médias de desempenho melhores que as meninas tanto na avaliação em papel quanto na avaliação em computador. Essa desigualdade dos sexos no desempenho em Matemática foi apontada no Relatório da OCDE (OECD, 2014a), que revela ainda diferenças preocupantes na atitude das meninas no que se refere à perseverança e motivação para aprender Matemática, já que mesmo aquelas com alto desempenho tem menos autoconfiança nas suas habilidades e um nível maior de ansiedade em relação a essa disciplina quando comparadas aos meninos com desempenho semelhante.

Em contrapartida, as diferenças das médias de desempenho das meninas e dos meninos nas provas em papel e em computador apresentaram valores semelhantes entre si e muito próximos à média nacional, refutando a ideia de que as meninas poderiam experimentar um efeito negativo ainda maior no seu desempenho ao realizarem avaliações em computador, conforme sugere Gallagher, Bridgeman e Cahalan (2000).

A despeito de uma possível afinidade maior dos meninos com as TICs, os computadores tem se tornado cada vez mais acessíveis aos alunos de ambos os sexos, seja em casa ou na escola, o que talvez esteja contribuindo para a minimização dessas diferenças. Mesmo que a frequência e tipo de uso que fazem dos computadores encontrem especificidades decorrentes do sexo, responder a um teste em computador exige apenas o conhecimento de tarefas básicas e rotineiras que a maioria das meninas e meninos parece capaz de desenvolver. Não obstante, em alguns estudos (SINGLETON, 2001; CLARIANA & WALLACE, 2002) as questões relacionadas ao sexo também não se mostraram associadas ao formato do teste.

No que se refere ao resultado favorável dos meninos em Matemática encontrado também em Andrade, Franco e Carvalho (2003), os pesquisadores argumentam que, ao menos em parte, a desigualdade entre os sexos pode ser explicada pelo perfil da população estudantil brasileira, que conforme avança para os anos finais da educação básica vai sendo composta por um percentual cada vez maior de meninas com nível socioeconômico mais baixo em relação ao percentual de meninos na mesma situação, por estes abandonarem mais precocemente a escola. Afirmam ainda que “as condições de escolarização dessa parcela da população são, tipicamente, mais precárias do que a dos alunos de maior nível socioeconômico” (p. 87).

Com o objetivo de aprofundar essa discussão, a comparação das médias dos alunos nas provas em papel e em computador foi realizada considerando-se também os diferentes anos de ensino, sendo os resultados apresentados na Tabela 12.

Tabela 12. Médias de desempenho dos alunos brasileiros nas provas de Matemática em papel e em computador do PISA 2012 em função do ano de ensino.

| Ano | N | Papel | | Computador | | Dif. |
|--------|-------|-------|------|------------|------|-------|
| | | Média | E.P. | Média | E.P. | |
| Brasil | 1.390 | 393,3 | 4,1 | 418,0 | 4,3 | -24,7 |
| 8 | 139 | 311,4 | 6,7 | 349,1 | 6,7 | -37,7 |
| 9 | 215 | 333,7 | 7,1 | 364,8 | 8,1 | -31,1 |
| 10 | 425 | 389,9 | 5,9 | 410,4 | 5,9 | -20,5 |
| 11 | 569 | 437,8 | 5,3 | 459,9 | 5,8 | -22,1 |
| 12 | 42 | 431,5 | 14,0 | 461,0 | 13,7 | -29,5 |

Fonte: Elaboração da autora.

De acordo com a idade correta de entrada no sistema de ensino do Brasil, com 15 anos os alunos devem estar cursando o 1º ano do Ensino Médio, representado no PISA pelo décimo ano de ensino. Embora o ano modal da amostra tenha sido o 2º ano do Ensino Médio (11º ano de ensino), tal condição confirma apenas a tendência de melhoria na composição do alunado já observada por Klein (2011). Assim, esses são os anos de ensino que podem revelar de forma mais fidedigna a influência do modo de aplicação no desempenho dos alunos.

Ao analisar a Tabela 12, podemos perceber que é exatamente nessa faixa, entre o 10º e o 11º ano de ensino que as diferenças de desempenho entre as avaliações em papel e em computador são menores. Apesar da diferença entre essas duas avaliações aumentar novamente no 12º ano de ensino (3º ano do Ensino Médio), é importante destacar que as médias associadas a ela possuem um erro padrão muito maior, em decorrência do quantitativo de alunos dessa etapa de ensino presentes na amostra.

O que se nota é que embora os alunos com defasagem idade-série apresentem pior desempenho quando comparados com os demais, eles demonstram um desempenho muito melhor em computador quando comparados entre si, enquanto os demais possuem médias cerca de 5% mais próximas nas duas avaliações. Todavia, a pesquisa TIC Educação 2013 (CETIC.br, 2014b, p. 166) apontou que “os alunos mais velhos apresentam menor dificuldade em realizar atividades no computador e na Internet”.

Um possível fator motivacional talvez justifique esse resultado. Considerando que entre as principais causas da distorção idade-série estão a evasão e o abandono escolar, intimamente relacionadas à condição socioeconômica do aluno, apesar da quase universalização do acesso aos computadores, tal inovação em uma avaliação pode representar uma

novidade muito maior para a realidade desses alunos, fazendo com que se sintam muito mais motivados para responder ao teste.

Além do nível de escolarização, outra variável que poderia contribuir para a diferença de desempenho nas provas em papel e em computador supostamente seria a localização da escola, já que segundo a Pesquisa TIC Domicílios e Empresas 2013 (CETIC.br, 2014a) o crescimento no número de domicílios com acesso a computadores na área rural não acompanha o da área urbana. Os resultados dessa análise são apresentados na Tabela 13.

Tabela 13. Médias de desempenho dos alunos nas provas de Matemática em papel e em computador do PISA 2012 em função da localização das escolas.

| Localização | N | Papel | | Computador | | Dif. |
|-------------|-------|-------|------|------------|------|-------|
| | | Média | E.P. | Média | E.P. | |
| Brasil | 1.390 | 393,3 | 4,1 | 418,0 | 4,3 | -24,7 |
| Rural | 65 | 317,1 | 10,3 | 358,6 | 10,5 | -41,5 |
| Urbana | 1.325 | 397,8 | 4,4 | 421,5 | 4,6 | -23,7 |

Fonte: Elaboração da autora.

Observa-se que as médias de desempenho dos alunos das escolas rurais nas avaliações em papel e em computador foram menores que as médias de desempenho dos alunos das escolas urbanas, ficando inclusive abaixo das médias nacionais. No entanto, quando observadas as diferenças das médias de desempenho nas duas avaliações de acordo com a localização das escolas, os alunos das escolas rurais tiveram um desempenho na avaliação em computador 13,1% superior em relação à avaliação em papel, enquanto para os alunos das escolas urbanas essa diferença foi de apenas 5,9%. Mesmo considerando o erro padrão elevado devido ao reduzido tamanho da amostra de alunos de escolas rurais, essa diferença mostrou-se muito acentuada, indicando a importância dessa variável.

Da mesma forma que os alunos com defasagem idade-série, podemos pressupor que os alunos das escolas rurais se sentiram muito mais motivados que os alunos das escolas urbanas para realizar um teste em computador, o que fez com que apresentassem diferenças ainda mais marcantes de desempenho quando comparadas as duas avaliações. Embora essa pareça ser uma explicação razoável para os resultados encontrados, para que se possa fazer qualquer afirmação nesse sentido é necessário explorar melhor essa variável e considerar outros possíveis aspectos envolvidos.

Com base nos dados do SAEB 2003, Bezerra e Kassouf (2006) fizeram um estudo dos fatores que afetam o desempenho dos alunos em Português e Matemática e concluíram que a região em que a escola se

localiza é uma das variáveis mais significativas para explicar as diferenças nas escolas rurais. Mais especificamente, os alunos avaliados em Matemática nas escolas urbanas e rurais da região Nordeste apresentaram um desempenho menor em comparação com os alunos das demais regiões, principalmente das mais desenvolvidas como as regiões Sudeste e Sul. Por conseguinte, essa variável também foi considerada para a comparação das médias de desempenho nas avaliações em papel. O resultado dessa análise é apresentado na Tabela 14.

Tabela 14. Médias de desempenho dos alunos nas provas de Matemática em papel e em computador do PISA 2012 em função da região/estado do Brasil (continua).

| Região/Estado | N | Papel | | Computador | | Dif. |
|----------------|-------|-------|------|------------|------|-------|
| | | Média | E.P | Média | E.P. | |
| Brasil | 1.390 | 393,3 | 4,1 | 418,0 | 4,3 | -24,7 |
| Norte | 232 | 363,8 | 10,6 | 404,4 | 8,9 | -40,6 |
| Acre | 40 | 410,4 | 6,9 | 443,1 | 5,4 | -32,7 |
| Amapá | 32 | 385,9 | 9,9 | 411,1 | 14,1 | -25,2 |
| Amazonas | 28 | 351,8 | 27,2 | 409,2 | 18,7 | -57,4 |
| Pará | 39 | 343,2 | 11,9 | 387,3 | 20,4 | -44,1 |
| Rondônia | 25 | 366,6 | 11,6 | 396,1 | 11,2 | -29,5 |
| Roraima | 30 | 381,5 | 35,1 | 403,6 | 13,2 | -22,1 |
| Tocantins | 38 | 397,1 | 46,5 | 431,7 | 22,3 | -34,6 |
| Nordeste | 339 | 360,7 | 7,7 | 388,9 | 7,7 | -28,2 |
| Alagoas | 34 | 338,1 | 22,5 | 389,5 | 28,7 | -51,4 |
| Bahia | 45 | 337,9 | 15,5 | 363,5 | 14,3 | -25,6 |
| Ceará | 68 | 370,4 | 25,8 | 402,5 | 25,5 | -32,1 |
| Maranhão | 43 | 333,8 | 18,1 | 358,4 | 26,1 | -24,6 |
| Paraíba | 28 | 395,7 | 15,4 | 415,1 | 26,3 | -19,4 |
| Pernambuco | 38 | 361,9 | 13,7 | 404,3 | 12,3 | -42,4 |
| Piauí | 26 | 407,7 | 17,2 | 394,0 | 12,5 | 13,7 |
| R. G. do Norte | 29 | 343,7 | 12,1 | 381,1 | 17,6 | -37,4 |
| Sergipe | 28 | 393,2 | 16,2 | 429,9 | 14,6 | -36,7 |
| Sudeste | 495 | 411,1 | 6,5 | 434,5 | 7,1 | -23,4 |
| Espírito Santo | 35 | 431,1 | 12,2 | 440,3 | 13,9 | -9,2 |
| Minas Gerais | 119 | 418,9 | 14,8 | 437,4 | 21,0 | -18,5 |
| Rio de Janeiro | 62 | 406,0 | 11,3 | 439,4 | 12,3 | -33,4 |
| São Paulo | 279 | 408,0 | 8,2 | 430,8 | 7,8 | -22,8 |

Tabela 14. Médias de desempenho dos alunos nas provas de Matemática em papel e em computador do PISA 2012 em função da região/estado do Brasil (conclusão).

| Região/Estado | N | Papel | | Computador | | Dif. |
|------------------|-----|-------|------|------------|------|-------|
| | | Média | E.P | Média | E.P. | |
| Sul | 179 | 410,2 | 8,1 | 420,8 | 11,3 | -10,6 |
| Paraná | 83 | 407,7 | 12,1 | 421,5 | 9,4 | -13,8 |
| R. G. do Sul | 60 | 411,2 | 16,2 | 434,5 | 8,0 | -23,3 |
| Santa Catarina | 36 | 413,4 | 12,9 | 400,9 | 43,9 | 12,5 |
| Centro-oeste | 145 | 377,2 | 19,4 | 409,8 | 15,2 | -32,6 |
| Distrito Federal | 24 | 432,2 | 36,5 | 452,9 | 43,4 | -20,7 |
| Goiás | 48 | 365,9 | 11,8 | 382,6 | 21,1 | -16,7 |
| Mato Grosso | 34 | 321,0 | 45,3 | 389,8 | 25,8 | -68,8 |
| M. G. do Sul | 39 | 428,4 | 21,2 | 453,5 | 20,2 | -25,1 |

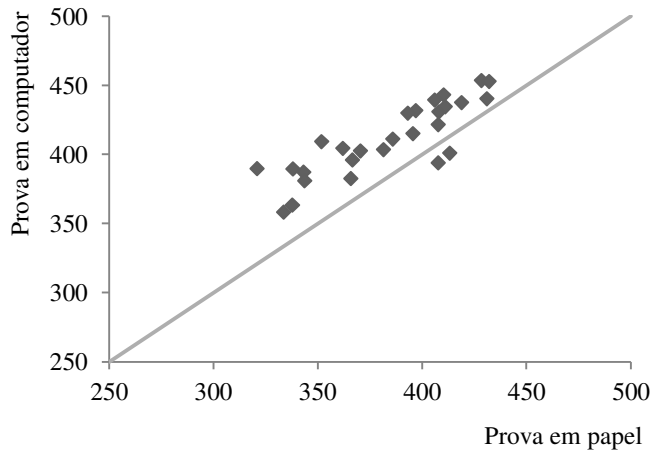
Fonte: Elaboração da autora.

Antes de discorrer sobre esses resultados é importante fazer uma ponderação quanto ao tamanho reduzido da amostra utilizada nesta pesquisa. Não se teve aqui a pretensão de fornecer médias de desempenho dos alunos representativas de cada estado. Embora essas estatísticas tenham sido calculadas, os valores de erro padrão encontrados são tão elevados que impedem que sejam utilizadas para se fazer afirmações precisas. Assim, considerou-se prudente analisá-las apenas agrupadas por regiões, que oferecem dados mais confiáveis, embora também devam ser interpretadas com certa cautela.

Refletindo as médias de desempenho nacionais, os alunos das cinco regiões apresentaram desempenho na prova em computador superior ao desempenho na prova em papel. Contudo, quando consideradas as diferenças entre elas, os alunos da região Norte se destacaram por terem desempenho 11,2% melhor na avaliação em computador comparada à avaliação em papel, a maior diferença entre todas as regiões. Inversamente, os alunos da região Sul apresentaram a menor diferença de desempenho entre as duas avaliações, de apenas 2,6%.

Considerando que esta pesquisa não teve como foco analisar as médias de desempenho dos alunos e sim a diferença entre elas e não sendo a dispersão entre as médias dos estados afetada pela medida do erro padrão, elas foram representadas no Gráfico 1.

Gráfico 1. Diagrama de dispersão entre as médias dos estados brasileiros nas provas de Matemática em papel e em computador do PISA 2012.



Fonte: Elaboração da autora.

Comparando-se as médias de desempenho dos alunos de cada estado brasileiro nas avaliações em papel e em computador, encontramos uma correlação positiva, que demonstra que quando maior o desempenho em papel, maior é o desempenho em computador e vice-versa. Entretanto, o Gráfico 1 sugere um deslocamento na escala do PISA entre as provas de papel e de computador, com resultados favoráveis à prova no computador, exceto para o Piauí e Santa Catarina.

Resumidamente, buscou-se nesta seção analisar algumas das características individuais do aluno e da escola apontados na literatura como associados ao desempenho escolar. Dentre elas, o acesso a computadores teve um pequeno impacto positivo no desempenho dos alunos na avaliação em computador, enquanto o sexo não se mostrou relacionado às diferenças observadas. Para as variáveis ano de ensino e localização (rural/urbana e região/estado) da escola, a aplicação do teste em computador contribuiu para diminuir as diferenças entre os grupos. Ou seja, para todos os grupos em que se esperava um desempenho menor, apesar desse resultado ter se confirmado, a diferença de desempenho nas provas em papel e em computador foi superior quando comparada à dos grupos conhecidamente com desempenho maior, representando uma diminuição na desigualdade de desempenho entre eles quando a prova foi aplicada em computador.

4.3 ANÁLISE DAS PROVAS

Como citado na seção 2.1.1, a escala de Matemática do PISA 2012 é dividida em seis níveis de proficiência. A porcentagem de alunos brasileiros em cada um desses níveis é apresentada na tabela 15¹².

Tabela 15. Porcentagem de alunos brasileiros em cada nível de proficiência na escala de Matemática do PISA 2012 de acordo com o formato de aplicação.

| Formato | Nível | | | | | | | | | | | | | |
|------------|-------------|------|-------------|------|------|------|------|------|-----|------|-----|------|-----|------|
| | Abaixo de 1 | | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
| | % | E.P. | % | E.P. | % | E.P. | % | E.P. | % | E.P. | % | E.P. | % | E.P. |
| Papel | 36,9 | 0,9 | 31,4 | 0,7 | 19,7 | 0,7 | 8,5 | 0,4 | 2,7 | 0,3 | 0,7 | 0,2 | 0,0 | 0,0 |
| Computador | 23,6 | 1,8 | 28,8 | 1,2 | 26,8 | 1,6 | 13,4 | 0,9 | 5,7 | 1,0 | 1,6 | 0,5 | 0,2 | 0,1 |

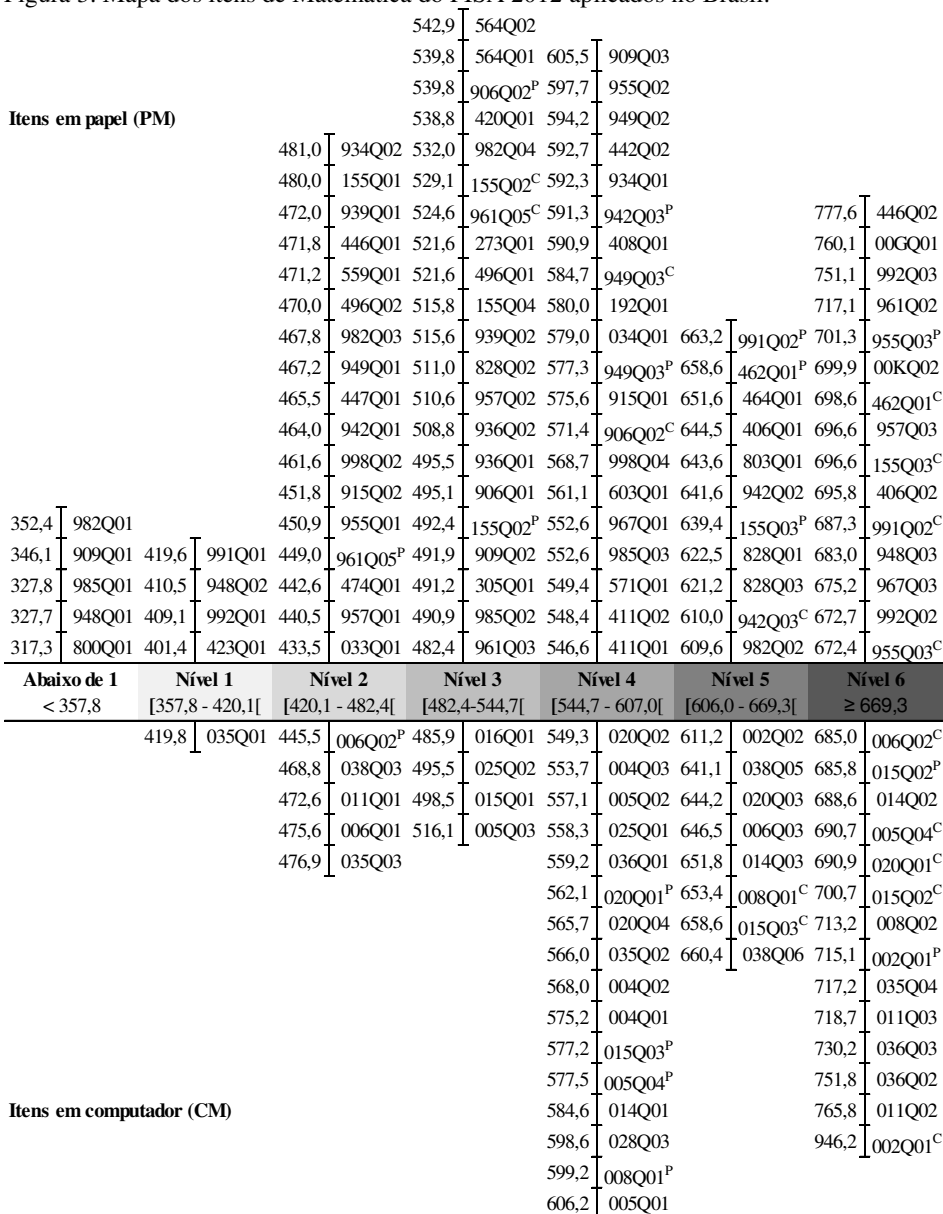
Fonte: OECD, 2014a, p. 295 (adaptado pela autora).

De maneira semelhante ao total dos países da OCDE, em que a maioria dos alunos se encontra no nível 2 (de 420,1 a menos de 482,4) na avaliação em papel e no nível 3 (de 482,4 a menos de 544,7) na avaliação em computador (OECD, 2014a), no Brasil a maioria dos alunos se encontra um nível acima na avaliação em computador em relação a avaliação em papel, estando posicionados abaixo do nível 1 (abaixo de 357,8) na avaliação em papel e no nível 1 (de 357,8 a menos de 420,1) na avaliação em computador.

Tendo em vista que o PISA é uma avaliação internacional e que as médias de desempenho do Brasil são inferiores à média de desempenho dos países da OCDE e, ainda, que a maioria dos alunos brasileiros estão posicionados nas escalas de Matemática em papel e em computador abaixo do nível 2, considerado pela OCDE como o nível básico de proficiência em Matemática necessário para uma plena participação na sociedade (OECD, 2014a), as provas em papel e em computador foram analisadas mais detalhadamente, a fim de tornar mais robusta a discussão proposta nesta pesquisa. A distribuição dos itens de Matemática em papel e em computador nos seis níveis de proficiência da escala do PISA 2012 é apresentada na Figura 3.

¹² Os resultados do Brasil foram divulgados pela OCDE incluindo e excluindo as escolas rurais. No entanto, nesta seção são reproduzidos apenas os resultados brasileiros incluindo as escolas rurais, a fim de manter a coerência com a subamostra utilizada nas análises apresentadas anteriormente nesta pesquisa.

Figura 3. Mapa dos itens de Matemática do PISA 2012 aplicados no Brasil.



Fonte: Elaboração da autora (a partir de dados da OECD, 2014b, p. 406-409; 415-416).

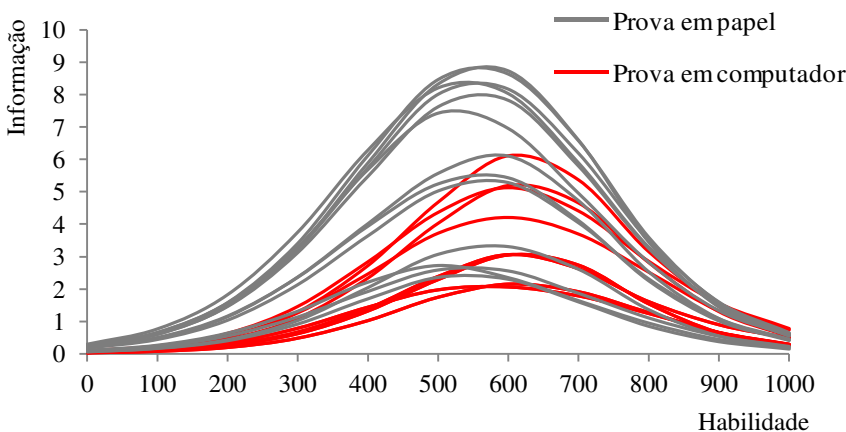
^C = crédito completo; ^P = crédito parcial

Enquanto a maioria dos itens em papel estão posicionados no nível 3 (de 482,4 a menos de 544,7) da escala de proficiência em Matemática do PISA, a maioria dos itens em computador estão posicionados no nível 4 (de 544,7 a menos de 607,0), ou seja, em geral, os itens de Matemática em computador foram mais difíceis que os itens de Matemática em papel.

Quando comparados o posicionamento dos alunos brasileiros *versus* o posicionamento dos itens na escala verifica-se que poucos itens possuem parâmetro de dificuldade condizente com o nível 1 ou abaixo dele na escala, níveis que, no entanto, representam juntos mais da metade dos alunos brasileiros (68,3% dos alunos na avaliação em papel e 52,4% dos alunos na avaliação em computador) (OECD, 2014a). Isso se torna ainda mais evidente para os itens em computador, em que apenas um item (CM035Q01) ficou posicionado abaixo do nível 2.

Como o PISA possui vários modelos de prova em papel e formulários em computador, no intuito de explorar melhor a distribuição desses itens, foram analisadas também as curvas de informação de cada uma das provas. A curva de informação representa o somatório das informações de todos os itens, indicando em que ponto da escala o teste foi mais informativo (ANDRADE, LAROS & GOUVEIA, 2010). O Gráfico 2 mostra as curvas de informação de todas as provas de Matemática em papel e em computador aplicadas no Brasil no PISA 2012.

Gráfico 2. Curvas de informação das provas de Matemática em papel e em computador aplicadas no Brasil no PISA 2012.



Fonte: Elaboração da autora.

Em decorrência do quantitativo de blocos de Matemática em cada um dos cadernos/formulários, verifica-se que algumas provas em papel foram mais informativas que as provas em computador. As curvas de informação dos formulários da prova em computador ficaram deslocadas para a direita em relação às curvas de informação dos cadernos da prova em papel, refletindo o nível de dificuldade geralmente mais elevado desses itens.

Embora à primeira vista pareça contraditório que sendo os formulários da prova em computador mais difíceis que os cadernos da prova em papel os alunos tenham se saído, em média, melhor na avaliação em computador, esse é possivelmente um dos fatores que ajudam a explicar a grande diferença observada entre as médias de desempenho nas duas avaliações no Brasil, bem como os resultados encontrados na seção 4.2.

Para estimar de forma precisa a proficiência de um aluno é necessário que lhe sejam apresentados itens na região da escala próxima à sua proficiência. Quando as proficiências dos alunos estão muito acima do limite superior ou muito abaixo do limite inferior que as provas alcançam, não é possível conhecer sua verdadeira proficiência (ALAVARSE & MELO, 2012) e o seu desempenho será subestimado ou superestimado dependendo do nível de dificuldade dos itens respondidos.

Nas provas de Matemática em computador analisadas, o que se observa é que o ponto mais informativo das curvas (que representa a região da escala em que os alunos são avaliados de forma mais precisa) está muito acima do ponto em que a maioria dos alunos brasileiros se encontra, de tal maneira que a proficiência dos alunos nesse formato de aplicação pode estar superestimada, ficando ainda mais discrepante nos grupos de menor desempenho.

4.3.1 Comparação do posicionamento dos itens

Como parte qualitativa dos resultados, foi realizada uma análise comparativa do posicionamento dos itens em papel e em computador na escala de Matemática do PISA 2012. O ponto da escala em que cada item se encontra é definido pela sua dificuldade (parâmetro b), que por sua vez está diretamente associada a um número de características do item que demandam dos alunos variados níveis de ativação de cada uma das capacidades fundamentais descritas na matriz de Matemática (OECD, 2013).

Dado que os itens das duas provas são diferentes, para que essa comparação pudesse ser feita, foram consideradas em conjunto as três categorias de classificação utilizadas pelo PISA: conteúdo, processo e contexto. Além disso, em virtude do quantitativo elevado e do caráter sigiloso da maioria dos itens, que impossibilita sua descrição mais detalhada,

essa análise se restringiu à categoria de conteúdo com maior número de itens nessa edição, a categoria *Espaço e Forma*.

No PISA 2012 a categoria de conteúdo *Espaço e Forma* foi avaliada por 34 itens, 22 em papel e 12 em computador, posicionados na escala geral de letramento em Matemática nos seis níveis (nível 1 ao nível 6) com parâmetro de dificuldade variando entre 409,1 e 946,2. Assim como na escala geral de Matemática, no que se refere a subescala *Espaço e Forma*, a maioria dos alunos brasileiros, 71,8% também se encontra no nível 1 ou abaixo dele (OECD, 2014a), embora esses níveis tenham sido representados por apenas três itens dessa categoria de conteúdo. A Figura 4 apresenta a distribuição dos itens de *Espaço e Forma* nas categorias de processo e contexto, bem como o tipo de resposta e parâmetro de dificuldade para cada um deles (itens em negrito indicam itens liberados para divulgação).

Figura 4. Mapa dos itens de Matemática do PISA 2012 classificados na categoria de conteúdo Espaço e Forma (continua).

| | | Espaço e Forma | | | | |
|-------|---------------------|----------------|---------------------------|---------------------------|---------------------------|------------|
| Nível | Itens em Papel (PM) | | Posição na escala | Itens em Computador (CM) | | TR |
| | Interpretar | Formular | | Empregar | Empregar | |
| 6 | | | 946,2 | 002Q01^C | | RCA |
| | | 00GQ01 | 760,1 | | | RCM |
| | | | 715,1 | 002Q01^P | | RCA |
| | | 00KQ02 | 699,9 | | | RCE |
| | | | 698,6 | 462Q01^C | | RCE |
| | | 406Q02 | 695,8 | | | RCE |
| | | | 690,9 | 020Q01^C | | RCA |
| | | | 991Q02^C | 687,3 | | RCE |
| | | 967Q03 | 675,2 | | | MEC |
| | | 992Q02 | 672,7 | | | RCM |
| 5 | | | 991Q02^P | 663,2 | | RCE |
| | | | 462Q01^P | 658,6 | | RCE |
| | | | | 653,4 | 008Q01^C | RCA |
| | | 464Q01 | 651,6 | | | RCA |
| | | | 406Q01 | 644,5 | | RCE |
| | | | | 644,2 | 020Q03 | RCA |
| | | | | 611,2 | 002Q02 | RCA |

Figura 4. Mapa dos itens de Matemática do PISA 2012 classificados na categoria de conteúdo Espaço e Forma (conclusão).

| Espaço e Forma | | | | | | |
|----------------|---------------------|---------------------|-------------------|---------------------------|----------------------|------------|
| Nível | Itens em Papel (PM) | | Posição na escala | Itens em Computador (CM) | | TR |
| | Interpretar | Formular Empregar | | Empregar | Formular Interpretar | |
| 4 | | | 599,2 | 008Q01 ^P | | RCA |
| | | 949Q02 | 594,2 | | | MEC |
| | | 934Q01 | 592,3 | | | RCM |
| | | 949Q03 ^C | 584,7 | | | RCE |
| | | 034Q01 | 579,0 | | | RCA |
| | | 949Q03 ^P | 577,3 | | | RCE |
| | | | 575,2 | 004Q01 | | MEC |
| | | | 568,0 | 004Q02 | | MEC |
| | | | 565,7 | 020Q04 | | MES |
| | | | 562,1 | 020Q01^P | | RCA |
| | | | 553,7 | | 004Q03 | RCA |
| | | 967Q01 | 552,6 | | | RCM |
| | | | 549,3 | 020Q02 | | MES |
| | 3 | | 273Q01 | 521,6 | | |
| | | 305Q01 | 491,2 | | | MES |
| 2 | | 934Q02 | 481,0 | | | MES |
| | | | 476,9 | | 035Q03 | MES |
| | | 949Q01 | 467,2 | | | MEC |
| | | 447Q01 | 465,5 | | | MES |
| | 033Q01 | | 433,5 | | | MES |
| 1 | | | 419,8 | | 035Q01 | MES |
| | 991Q01 | | 419,6 | | | MES |
| | | 992Q01 | 409,1 | | | RCM |

Fonte: Elaboração da autora (a partir de dados da OECD, 2014b, p. 406-409; 415-416).

Contexto: Científico Pessoal Social Ocupacional
 TR=Tipo de resposta; MES=Múltipla escolha simples; MEC=Múltipla escolha complexa; RCA=Resposta construída autocodificada; RCM=Resposta construída codificada manualmente; RCE=Resposta construída codificada por especialistas; VRS=Variações de resposta de seleção.

^C=crédito completo; ^P=crédito parcial.

Apesar de alguns itens classificados nessa categoria de conteúdo terem sido liberados, eles pertencem a categorias de processos e/ou contextos diferentes, o que impossibilita sua comparação direta, ainda que alguns deles sejam utilizados para ilustrar as discussões. Dessa forma, dos 34 itens, apenas 21 itens puderam ser comparados por terem sido classificados também na mesma categoria de processo e contexto.

Na categoria de processo *Empregar*, contexto *Científico*, foram comparados cinco itens: o item em papel PM462Q01 – O TERCEIRO LADO e os itens em computador da unidade CM020 – PONTO ESTRELA, questões 01, 02, 03 e 04, reproduzidos nas Figuras 5.1, 5.2, 5.3 e 5.4. Embora pertençam à mesma categoria de conteúdo, processo e contexto, as duas unidades (em papel e em computador) enfatizam distintas propriedades geométricas dos objetos, o que pode justificar seu posicionamento em níveis diferentes, ainda que não muito distantes da escala. Nesse sentido, o item mais difícil foi o item em papel PM462Q01 (crédito completo), seguido do item em computador CM020Q01 (crédito completo), ambos posicionados no nível 6 da escala. No entanto, apesar de apenas 0,1% dos alunos brasileiros atingirem esse nível nas subescalas de conteúdo *Espaço e Forma* e de processo *Empregar* (OECD, 2014a), analisando o número de alunos da amostra desta pesquisa que responderam corretamente a esses itens, observa-se que o item em papel teve um percentual muito menor de acerto, de apenas 0,5%, enquanto para o item em computador esse percentual foi de 4,2%. Uma possível explicação para essa diferença pode estar na forma como o item em computador foi elaborado, já que para que o aluno recebesse o crédito completo deveria atender a duas solicitações: criar um ponto estrela e um ponto que não fosse estrela, de forma que a necessidade de raciocínio inverso pode ter contribuído para aumentar ilusoriamente o nível de dificuldade do item, visto que seu crédito parcial teve 26,3% de acerto e ficou posicionado no nível 4 da escala, juntamente com as questões 02 e 04 dessa mesma unidade. Apesar dos quatro itens em computador serem interativos, inclusive de forma obrigatória nas questões 01 e 03, o que poderia ter impacto na sua resolução, seja de forma negativa pela necessidade de compreensão da dinâmica do item, seja de forma positiva, por propiciar uma melhor visualização espacial, os níveis de proficiência exigidos por cada um se mostraram razoavelmente condizentes com os níveis descritos na escala geral de proficiência em Matemática do PISA e nas subescalas de processo *Formular* e de conteúdo *Espaço e Forma*¹³, sugerindo que o formato do item não interferiu no seu nível de dificuldade.

¹³ A descrição dos níveis de todas as subescalas de processo e de conteúdo de Matemática do PISA 2012 podem ser consultadas em OECD, 2014a.

Figura 5.1. Item CM020Q01.

pt-PT Programme for International Student Assessment 2012

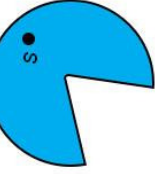
PONTOS ESTRELA

Em qualquer figura, um ponto, S , chama-se um ponto estrela se, para qualquer outro ponto, P , no interior da figura, o segmento de reta SP , ficar totalmente no interior da figura.

Os botões PUNTO (S) e SEGMENTO (SP) usam-se do seguinte modo:

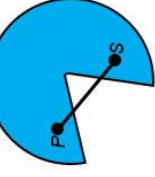
- Clica no botão PUNTO (S) e depois clica numa figura para criares um único ponto.
- Clica no botão SEGMENTO (SP) e depois clica numa figura para criares um segmento de reta entre os pontos S e P .
- Para moveres um ponto ou um segmento clica sobre ele e arrasta-o.
- Para eliminares um ponto ou um segmento clica sobre ele.

Figura 1




S é um ponto estrela

Figura 2



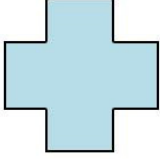
S não é um ponto estrela

Figura 3



PUNTO (S) SEGMENTO (SP) REINICIAR

Figura 4



Questão 1: PONTOS ESTRELA CM020Q01

Em cima, mostram-se quatro figuras. Na Figura 1, o ponto S é um ponto estrela porque, onde quer que coloques P , o segmento SP fica sempre dentro da figura. Mas na Figura 2, o ponto S não é um ponto estrela, porque há alguns segmentos SP , como o do exemplo apresentado, que saem para fora da figura. Cria um ponto estrela para a Figura 3 e um ponto que **não** seja um ponto estrela para a Figura 4.

? ↩

Fonte: INEP, c2011.

Figura 5.2. Item CM020Q02.



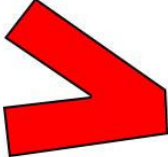

pt-PT Programme for International Student Assessment 2012

PONTOS ESTRELA

Em qualquer figura, um ponto, *S*, chama-se um ponto estrela se, para qualquer outro ponto, *P*, no interior da figura, o segmento de reta *SP* ficar totalmente no interior da figura.

Os botões PUNTO (S) e SEGMENTO (SP) usam-se do seguinte modo:

- Clica no botão PUNTO (S) e depois clica numa figura para criares um único ponto.
- Clica no botão SEGMENTO (SP) e depois clica numa figura para criares um segmento de reta entre os pontos *S* e *P*.
- Para moveres um ponto ou um segmento clica sobre ele e arrasta-o.
- Para eliminares um ponto ou um segmento clica sobre ele.

Figura 1 
 Figura 2 
 Figura 3 
 Figura 4 

Questão 2: PONTOS ESTRELA CM020Q02

Algumas figuras têm muitos pontos estrela e outras não têm ponto estrela. Para uma das figuras acima, é impossível encontrar um ponto estrela. Qual das figuras não tem ponto estrela?

Figura 1
 Figura 2
 Figura 3
 Figura 4

1 2 3

?

↑

INEP

Fonte: INEP, c2011.

Figura 5.3. Item CM020Q03.

pt-PT Programme for International Student Assessment 2012

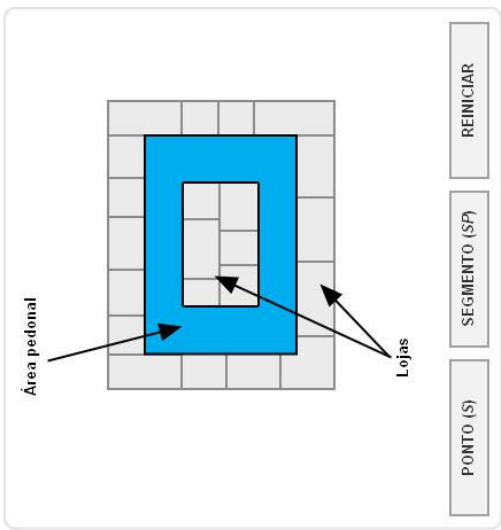
1 2 3

PONTOS ESTRELA

Em qualquer figura, um ponto, *S*, chama-se um ponto estrela se, para qualquer outro ponto, *P*, no interior da figura, o segmento de reta *SP* ficar totalmente no interior da figura.

Os botões **PONTO (S)** e **SEGMENTO (SP)** usam-se do seguinte modo:

- Clica no botão **PONTO (S)** e depois clica numa figura para criares um único ponto.
- Clica no botão **SEGMENTO (SP)** e depois clica numa figura para criares um segmento de reta entre os pontos *S* e *P*.
- Para moveres um ponto ou um segmento clica sobre ele e arrasta-o.
- Para eliminares um ponto ou um segmento clica sobre ele.



Área pedonal

Lojas

PONTO (S) SEGMENTO (SP) REINICIAR

Questão 3: PONTOS ESTRELA CM020Q03

Em cima está representada uma planta de um novo centro comercial com um único piso. A zona colorida representa áreas onde as pessoas podem caminhar. Vão ser instaladas câmaras de vigilância para observar toda a área pedonal. As câmaras de vigilância serão montadas no teto da área pedonal, onde podem ter uma visão de 360° e flimar a área pedonal visível.

Coloca pontos na planta acima, para mostrar onde devem ser instaladas as câmaras de vigilância, de forma a respeitarem as seguintes condições:

- Qualquer parte da área pedonal pode ser observada através de, pelo menos, uma câmara.
- E usado o **interior** número de câmaras.

? ↑

INEP

Fonte: INEP, c2011.

Figura 5.4. Item CM020Q04.

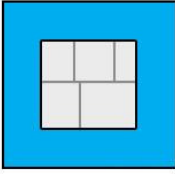
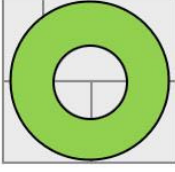
pt-PT Programme for International Student Assessment 2012

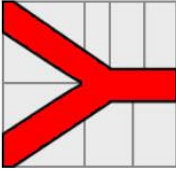
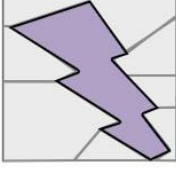
PONTOS ESTRELA

Em qualquer figura, um ponto, *S*, chama-se um ponto estrela se, para qualquer outro ponto, *P*, no interior da figura, o segmento de reta *SP* ficar totalmente no interior da figura.

Os botões **PONTO (S)** e **SEGMENTO (SP)** usam-se do seguinte modo.

- Clica no botão **PONTO (S)** e depois clica numa figura para criares um único ponto.
- Clica no botão **SEGMENTO (SP)** e depois clica numa figura para criares um segmento de reta entre os pontos *S* e *P*.
- Para moveres um ponto ou um segmento clica sobre ele e arrastá-o.
- Para eliminares um ponto ou um segmento clica sobre ele.

Planta 1  **Planta 2** 

Planta 3  **Planta 4** 

PONTO (S) **SEGMENTO (SP)** **REINICIAR**

Questão 4: PONTOS ESTRELA CM020Q04

Em cima, estão quatro plantas diferentes para áreas pedonais de um edifício com um único piso. As áreas pedonais estão coloridas. Vão ser instaladas câmaras de vigilância para observar estas áreas. As câmaras de vigilância serão montadas no teto da área pedonal, onde podem ter uma visão de 360° e filmar a área pedonal visível.

Qual das quatro plantas acima, requer o **menor** número de câmaras de vigilância?

Planta 1 Planta 2 Planta 3 Planta 4

? ↑

Fonte: INEP, c2011.

Na categoria de processo *Empregar*, contexto *Pessoal*, foram comparados outros cinco itens: o item em papel PM967Q01 – TRENZINHO DE MADEIRA e os itens das unidades em computador CM002 - AGRUPAMENTO DE CUBOS, questões 01 e 02 e CM004 – TORRADA, questões 01 e 02. Infelizmente nenhum desses itens foi liberado para divulgação, impossibilitando uma discussão mais profunda dessa análise comparativa. De maneira geral, os itens em computador tiveram um nível de dificuldade mais alto, sendo que os dois itens da unidade CM002 ficaram posicionados nos níveis 5 e 6 da escala, enquanto os itens da unidade CM004 ficaram posicionados no nível 4, juntamente com o item em papel PM967Q01. O item CM002Q01 (crédito completo) foi o item mais difícil de todos os itens de Matemática aplicados no PISA 2012, de modo que nenhum aluno da amostra desta pesquisa obteve crédito completo e apenas 4,2% dos alunos conseguiu respondê-lo de forma parcialmente correta. Embora os quatro itens em computador fossem interativos, os itens da unidade CM004 exigiam resposta de múltipla escolha complexa, do tipo sim/não, de forma que a interatividade era apenas acessória, ainda que auxiliasse o aluno na resolução do item, ao contrário dos itens da unidade CM002, em que a resposta do aluno deveria ser dada com a própria interação. No que se refere especificamente ao item CM002Q01, a necessidade de atender a duas condições para receber o crédito completo pode ter colaborado para elevar ainda mais o nível de dificuldade do item, fazendo com que os valores dos parâmetros b para os créditos parcial e completo ficassem tão distantes. Apesar de não estarem classificados na mesma categoria de processo e contexto, o item CM002Q01 se assemelha ao item em papel PM034Q01 – TIJOLOS (processo *Formular*, contexto *Ocupacional*), já que ambos trabalham com a ideia de perspectiva. O que se verifica, entretanto, é que possivelmente pela impossibilidade de rotacionar o objeto, o item em papel foi elaborado com uma imagem muito mais clara que o item em computador, tornando sua resolução mais fácil, o que é evidenciado pelo seu posicionamento em um nível mais baixo da escala, o nível 4, descrito na subescala *Espaço e Forma* exatamente como o ponto a partir do qual o aluno é capaz de “analisar a estrutura de um objeto tridimensional baseado em duas perspectivas diferentes dele” (OECD, 2014a, p. 103). De todo modo, por não serem os mesmos itens, que além de tudo possuem características inerentes ao seu próprio formato, não podemos afirmar que o modo de aplicação foi o responsável pelos diferentes níveis de dificuldade.

Na categoria de processo *Formular*, contexto *Pessoal*, foram comparados quatro itens: os itens em papel PM00GQ01 – UMA COLUNA PUBLICITÁRIA, PM00KQ02 – BASQUETEBOL SOBRE RODAS e PM967Q03 – TRENZINHO DE MADEIRA e o item em computador

CM004Q03 – TORRADA. Da mesma forma que no grupo anterior, nenhum desses itens foi liberado para divulgação e, assim, foram analisados de modo mais geral. Os três itens em papel ficaram posicionados no mesmo nível da escala geral de letramento em Matemática, o nível 6, cuja descrição na subescala *Espaço e Forma* abrange claramente as habilidades exigidas nesses itens. No entanto, no item em computador, a descrição do nível 5 da subescala se encaixa um pouco melhor com as habilidades exigidas do aluno do que a descrição do nível 4 em que ele ficou posicionado, indicando que a interatividade do item, classificada nesta pesquisa como acessória, pode ter facilitado sua resolução.

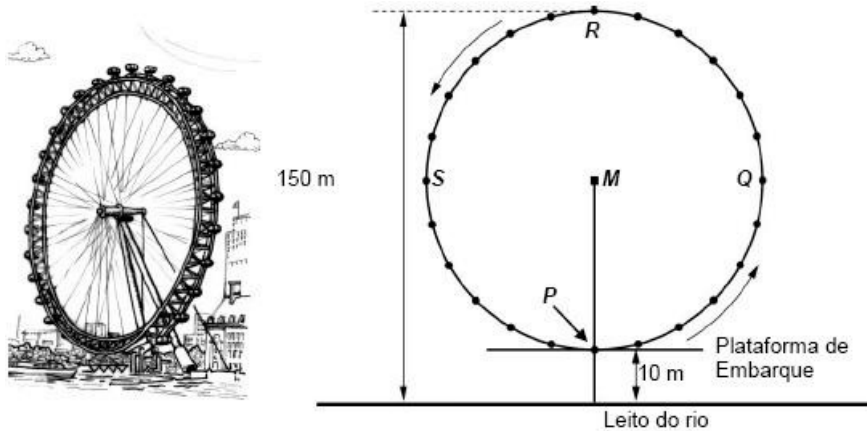
Ainda na categoria de processo *Formular*, mas no contexto *Social*, foram comparados outros quatro itens: os itens em papel PM406Q02 – PISTAS DE ATLETISMO, PM464Q01 – A CERCA e PM934Q02 – RODA GIGANTE, reproduzido na figura 6 e o item em computador CM008Q01 – MEDIDA. O item PM934Q02 faz parte do bloco PM6B, um dos blocos fáceis aplicados no Brasil, sendo o mais fácil de todos os itens comparados nesse grupo, com parâmetro de dificuldade igual a 481 ficando, portanto, posicionado no nível 2 da escala. O Relatório Técnico do PISA 2012 apresenta uma breve descrição qualitativa da demanda desse item: “Interpretar texto para compreender tarefas, extrair e utilizar os dados do gráfico para formular um modelo simples, envolvendo raciocínio sobre frações de um círculo” (OECD, 2014b, p. 295)¹⁴. A questão requer que o aluno estime o local em que João estará depois de decorrido certo tempo, com base na rotação da roda gigante. Para tanto, analisando a relação entre o processo *Formular* e as capacidades fundamentais da Matemática (OECD, 2013), verifica-se que o aluno deverá ativar principalmente sua capacidade de *comunicar e delinear estratégias* para reformular o problema contextualizado transformando-o em um modelo matemático. Diferentemente do item PM464Q01 que requer nível 5 de proficiência, o item PM934Q02 não possui uma alta demanda de *matematização*, exigida também no item PM406Q02, que solicita ainda que o aluno demonstre sua capacidade de *representação* e de *utilizar linguagem e operações simbólicas, formais e técnicas*, o que provavelmente fez com que ficasse posicionado no nível 6 da escala.

¹⁴ Tradução livre. No original: “ Interpret text to understand task, extract and use data from graphic to formulate simple model, involving reasoning about fractions of a circle.”

Figura 6. Item PM934Q02.

RODA GIGANTE

Na margem do rio fica uma roda gigante.
Veja a foto e o diagrama abaixo.



A roda gigante tem um diâmetro de 140 metros e o seu ponto mais alto está a 150 metros acima do leito do rio, em uma das margens do rio. Ela gira na direção indicada pela seta.

PM934Q02: RODA GIGANTE

A roda gigante gira em velocidade constante. A roda faz uma rotação completa em exatamente 40 minutos.

João inicia o passeio na roda gigante na plataforma de embarque P .

Onde João estará depois de meia hora?

- A Em R
- B Entre R e S
- C Em S
- D Entre S e P

Fonte: INEP, c2011.

De todos os itens em papel relacionados, o item PM464Q01 é o que mais se aproxima do item em computador CM008Q01, seja pelo posicionamento na escala, seja pelo conceito Matemático exigido do aluno. No entanto, para que o aluno respondesse corretamente ao item no computador, era necessário compreender as instruções e reproduzir a informação gerada pela sua interação com a plataforma, o que talvez tenha contribuído para elevar a sua dificuldade, já que, comparado ao item em papel, a demanda de ativação das capacidades Matemáticas foi menor.

Por último, na categoria de processo *Interpretar*, contexto *Pessoal*, foram comparados três itens: os item em papel PM033Q01 – DESENHO DE UM CÔMODO e os itens em computador da unidade CM035 – NAVEGAÇÃO GPS, questões 01 e 03, todos com resposta do tipo múltipla escolha simples. Os itens da unidade CM035 apesar de não serem estáticos, podem praticamente ser considerados como tal, pois permitem uma interação pequena que apenas facilita a compreensão da imagem. Embora a questão 01 tenha ficado posicionada no nível 1 da escala geral de letramento em Matemática, o seu parâmetro de dificuldade ficou muito próximo do limite inferior do nível 2, em que ficaram posicionados os demais itens. No entanto, para os itens em computador, as habilidades exigidas do aluno estão mais bem descritas no nível 3 da subescala *Espaço e Forma*, nível em que ficou posicionado o item PM305Q01 - MAPA, que apesar de pertencer à categoria de processo *Empregar*, contexto *Social*, possui uma abordagem semelhante. Para o item em papel, analisando a relação entre o processo *Interpretar* e as capacidades fundamentais da Matemática, verifica-se uma demanda cognitiva com foco basicamente na ativação da *representação*, envolvendo “comparar ou avaliar duas ou mais representações em relação a uma situação” (OECD 2013, p. 32)¹⁵, com complexidade condizente com o seu posicionamento na escala.

Em síntese, não foi possível estabelecer uma relação de causa e efeito entre o nível de dificuldade dos itens e o modo de aplicação, haja vista essa característica não ter sido isolada. Ademais, dos 41 itens utilizados na prova de Matemática em computador, apenas sete podem ser considerados completamente estáticos, não exigindo do aluno nem mesmo o uso da barra de rolagem. Isso evidencia as diferenças substanciais existentes nos instrumentos aplicados, inviabilizando uma comparação direta dos processos cognitivos necessários à resolução dos itens.

¹⁵ Tradução livre. No original: “compare or evaluate two or more representations in relation to a situation.”

5 CONCLUSÃO

Embora se reconheça as inúmeras vantagens dos testes em computador, a possibilidade de incorporar a tecnologia às avaliações educacionais levanta uma série de questões importantes, dentre elas a sua equivalência com os testes aplicados tradicionalmente no formato “papel e lápis”. Este trabalho se propôs justamente a discutir a comparabilidade das avaliações educacionais nesses dois formatos de aplicação. Para tanto, foram utilizadas as provas de Matemática em papel e em computador do PISA 2012 aplicadas no contexto brasileiro.

O primeiro passo se deu pela análise da dimensionalidade do conjunto de itens, que sugeriu não haver violação do pressuposto da unidimensionalidade, indicando a equivalência dos instrumentos aplicados no que se refere ao traço latente avaliado. Apesar de algumas estatísticas apontarem para um ganho significativo de informação com a adição do segundo fator, a tentativa de encontrar indícios que pudessem levar à compreensão dessas dimensões não revelou nenhuma característica que pudesse estar associada a elas. Assim, a predominância da Matemática foi considerada condição suficiente para se presumir a existência de um único fator dominante, demonstrando não haver influência do modo de aplicação no construto avaliado.

Na sequência, o efeito do modo de aplicação no desempenho dos alunos foi verificado pela comparação das médias em papel e em computador de grupos específicos, que foram definidos considerando características individuais do aluno e da escola: posse de computadores, ano de ensino, sexo e localização da escola (urbana/rural; região/estado). Os resultados revelaram um impacto positivo no desempenho dos alunos de praticamente todos os grupos avaliados quando as provas foram aplicadas em computador, condizente com as médias gerais de desempenho do Brasil. Em geral, essa diferença se mostrou ainda mais acentuada quando considerados os alunos com menor desempenho, exceto para a variável posse de computadores e sexo. Na busca de explicações para esses achados, julgou-se necessária uma investigação mais detalhada dos instrumentos aplicados, realizada por meio de uma análise das provas.

A análise dos itens de Matemática em computador evidenciou uma concentração desses itens nos níveis mais altos da escala e um quantitativo reduzido nos níveis mais baixos, em que se encontram a maioria dos alunos brasileiros. A curva de informação dos testes comprovou esse deslocamento das provas em computador em relação às provas em papel, demonstrando o seu elevado nível de dificuldade. Embora aparentemente contraditória, essa constatação permitiu identificar uma possível razão para as médias de

desempenho dos alunos brasileiros na avaliação em computador terem sido melhores que na avaliação em papel. Uma vez que os itens em computador não estavam adequados ao nível de proficiência da maioria dos alunos, as médias de desempenho nesse formato de teste foram superestimadas pela alta dificuldade dos itens, se tornando ainda mais discrepantes quanto menor o desempenho dos alunos. Isso explica também porque para a variável sexo as diferenças de desempenho observadas nas provas em papel e em computador mantiveram praticamente a mesma proporção, já que a proficiência dos dois grupos se mostraram muito próximas.

Posteriormente, a análise comparativa do posicionamento dos itens em papel e em computador apontou para as diferenças inerentes aos dois modos de aplicação, trazendo à tona a discussão a respeito das inúmeras possibilidades dos testes em computador, que permitem ampliar os limites das avaliações educacionais e medir competências que até então, no formato em papel, seriam impossíveis.

Verifica-se, portanto, que, apesar do conjunto de itens em papel e em computador terem sido considerados unidimensionais, sugerindo a equivalência dos dois testes, o mesmo não ocorreu em relação às médias de desempenho obtidas pelos alunos brasileiros, o que admitiu-se estar associado aos limites de aplicação dos instrumentos e ao recorte deste estudo. Ressalta-se ainda que, embora os mesmos alunos tenham respondido ao teste em ambos os formatos, os itens em computador eram, em sua maioria, interativos, impossibilitando análises comparativas mais precisas.

Desta forma, os resultados finais desta pesquisa foram inconclusivos quando à comparabilidade dos dois modos de aplicação, sendo necessário dar continuidade à este trabalho incluindo novos dados que permitam efetivamente investigar a validade dos testes em computador em uma possível transição. Além disso, vale salientar que a aplicação de testes em computador é um tema praticamente inexplorado no Brasil, mais ainda no que se refere ao campo das avaliações educacionais em larga escala. Assim, muitas são as questões com potencial para se constituir em objetos de pesquisa, dentre elas:

- A possibilidade de serem desenvolvidas avaliações em computador melhor adaptadas aos alunos com necessidades especiais;
- O comportamento diferencial dos itens (DIF) em determinados grupos populacionais, que podem considerar, por exemplo, idade, sexo e nível socioeconômico;
- A análise dos resultados do PISA 2015, com foco em Ciências, cuja avaliação foi completamente aplicada em computador, até mesmo para verificar o efeito das avaliações em computador nas diferentes áreas

do conhecimento, podendo se valer ainda dos dados do questionário de TICs para compreender melhor como o acesso e uso das tecnologias pode impactar no desempenho dos alunos;

- Um estudo comparativo do “letramento digital” dos mais diversos países, incluindo itens que permitam avaliar diferentes competências e processos cognitivos;
- Uma investigação da comparabilidade das provas em papel e em computador quando os mesmos itens são utilizados e quando pequenas alterações são implementadas.

Por fim, não obstante a importância de se estabelecer a equivalência dos testes em computador, é preciso destacar que essa provavelmente será uma questão transitória, tendo em vista os rápidos avanços da tecnologia e a universalização do seu acesso, que poderão, em um futuro próximo, até mesmo inverter o foco de análise.

REFERÊNCIAS BIBLIOGRÁFICAS

AGUIAR, Glauco da Silva. O Funcionamento Diferencial do Item (DIF) como estratégia para captar ênfases curriculares diferenciadas em matemática. **Est. Aval. Educ.**, São Paulo, v. 45, n. 21, p.169-190, 2010.

AKAIKE, Hirotugu. Information theory and an extension of the Maximum Likelihood principle. In: Petrov B, Csaki F (Org.) **Second International Symposium on Information Theory**, 1973, Akademiai Kiado, Budapest, p. 267-281.

AL-AMRI, Saad Saeed. **Computer-based testing vs paper-based testing: establishing the comparability of Reading tests through the evolution of a new comparability model in a Saudi EFL context**. 2009. 336 f. Tese (Doctor of Philosophy and Linguistics). Department of Language and Linguistics, University of Essex, 2009.

ALAVARSE, Ocimar Munhoz; MELO, Wolney Candido. Avaliação Educacional e Testes Adaptativos Informatizados (TAI): Desafios presentes e futuros. In: **TIC Educação 2012: Pesquisa sobre o Uso das Tecnologias de Informação e Comunicação nas Escolas Brasileiras**. São Paulo: Comitê Gestor da Internet no Brasil, 2013, p. 103-112. Disponível em: <http://www.cetic.br/media/docs/publicacoes/2/tic-educacao-2012.pdf>. Acesso em: 15 abr. 2015.

ALDERSON, J. Charles. Technology in testing: the present and the future. **System**, v. 28 (4) p. 593-603, 2010.

ALLEN, Mary J.; YEN, Wendy M. **Introduction to measurement theory**. Illinois: Waveland Press, 2002.

ANDRADE, Marcia; FRANCO, Creso; CARVALHO, João Pitombeira. Gênero e Desempenho em Matemática ao Final do Ensino Médio: quais as relações? **Estudos em Avaliação Educacional**, n. 27, p. 77- 96, 2003.

ANDRADE, Josemberg Moura de; LAROS, Jacob Arie; GOUVEIA, Valdiney Veloso. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. **Aval. psicol.** Porto Alegre, v. 9, n. 3, p. 421-435, 2010.

ANDRADE, Dalton Francisco de; TAVARES, Heliton Ribeiro; VALLE, Raquel da Cunha. Teoria da Resposta ao Item: Conceitos e Aplicações. In: **SINAPE**, 164p, 2000.

ARAUJO, Eutalia Aparecida Candido de; ANDRADE, Dalton Francisco de; BORTOLOTTI, Silvana Ligia Vincenzi. Teoria da Resposta ao Item. **Rev Esc Enferm USP**, 43(Esp), p. 1000-8, 2009.

BELLER, Michal. Technologies in Large-Scale Assessments: New Directions, Challenges, and Opportunities. In: **The Role of International Large-Scale Assessments: Perspectives from Technology, Economy and Educational Research**. von Davier, M., Gonzalez, E., Kirsch, I., Yamamoto, K. (Org.). Springer, 2013, p. 25-45.

BENNETT, Randy Elliot. **Reinventing Assessment: Speculations on the future of large-scale educational testing**. ETS, 1998.

_____. Inexorable and Inevitable: The Continuing Story of Technology and Assessment. **The Journal of Technology, Learning, and Assessment (JTLA)**. v. 1, n. 1, 2002.

_____. **Online Assessment and the Comparability of Score Meaning**. Educational Testing Service (ETS), Princeton, New Jersey, 2003.

BEZERRA, Márcio Garcia; KASSOUF, Ana Lúcia. Análise dos fatores que afetam o desempenho escolar nas escolas das áreas urbanas e rurais do Brasil. In: **XLIV Congresso da SOBER**, Fortaleza, 2006.

BLAZER, Christie. Computer-based Assessments. In: **Information Capsule**, Research Services, v. 0918, Flórida, 2010.

BOCK, R. Darrell., AITKIN, Murray. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. **Psychometrika**, v. 46 (4), p.443-459, 1981.

CASTRO, Maria Helena Guimarães de. **Sistemas Nacionais de Avaliação e de Informações Educacionais**. São Paulo em Perspectiva, 14(1), 2000.

CASTRO, Wilton. A era da aplicação de provas em computador. **Jornal do CESPE/UnB**, ano 5, n. 23, p. 6-8, 2013.

CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO (CETIC.br). O uso da internet por alunos brasileiros do Ensino Fundamental e Médio. **Panorama Setorial da Internet**. Ano 5, n. 2, São Paulo: Comitê Gestor da Internet no Brasil, 2013. Disponível em: <<http://cetic.br/publicacao/tecnologias-e-educacao-o-uso-da-internet-por-alunos-brasileiros-de-ensino-fundamental-e-medio/>> Acesso em: 15 abr. 2015.

_____. **30 anos de tecnologia da informação e comunicação no Brasil** [livro eletrônico]. São Paulo: Comitê Gestor da Internet no Brasil, 2014a. Disponível em: <http://cetic.br/media/docs/publicacoes/2/TIC_DOM_EMP_2013_livro_eletronico.pdf> Acesso em: 15 abr. 2015

_____. **30 anos de tecnologia da informação e comunicação nas escolas brasileiras** : [livro eletrônico]. São Paulo: Comitê Gestor da Internet no Brasil, 2014b. Disponível em: <<http://www.cetic.br/media/docs/publicacoes/2/tic-educacao-2013.pdf>> Acesso em: 15 abr. 2015

CHALMERS, Robert Philip, mirt: A Multidimensional Item Response Theory Package for the R Environment, **Journal of Statistical Software**, v.48, n.6, p.1-29, 2012.

CHEEK, Dennis W.; AGRUSO, Susan. Gender and equity issues in computer-based science assessment. **Journal of Science Education and Technology**, v. 4 (1), p. 75-79, 1995.

CHOI, Inn-Chull; KIM, Kyoung Sung; BOO, Joeyool. Comparability of a paper-based language test and a computer-based language test. **Language Testing**, 20(3), p. 295-320, 2003.

CLARIANA, Roy; WALLACE, Patricia. Paper-based versus computer-based assessment: key factors associated with the test mode effect. **British Journal of Educational Technology**, v. 33, n. 5, p. 593-602, 2002.

CLARKE, Jody; DEDE, Chris. Assessment, technology, and change. **Journal of Research on Technology in Education**. v. 42, n. 3, p. 309–328, 2010.

CONDÉ, Frederico Neves; LAROS, Jacob Arie. Unidimensionalidade e a propriedade de invariância das estimativas da habilidade pela TRI. **Avaliação Psicológica**, 6(2), p. 205-215, 2007.

CONOLE, Gráinne; WARBURTON, Bill. A review of computer-assisted assessment. **ALT-J, Research in Learning Technology**, v. 13, n. 1, p. 17-31, 2005.

DEL PORTO, Fabíola Brigante; FERREIRA, Clécio da Silva. Os fatores socioeconômicos associados ao desempenho dos alunos no PISA 2003 (Brasil, México, Espanha e Portugal). In: **XIII Congresso Brasileiro de Sociologia**, Recife, 2007.

DE-SIQUEIRA, Jose Macario; PERIS-FAJARNES, Guillermo; GIMENEZ, Fernando; MAGAL-ROYO, Teresa. Spanish students and teachers' preferences towards computer-based and paper-and-pencil test at universities. **Procedia Social and Behavioral Sciences**. v. 1, p. 814-817, 2009.

DEMO, Pedro. **Metodologia Científica em Ciências Sociais**. 3ª ed. rev. e ampl. São Paulo: Atlas, 1995.

DEMPSTER, Arthur P.; LAIRD, Nan M.; RUBIN, Donald B. Maximum Likelihood from Incomplete Data via the EM Algorithm. **Journal of the Royal Statistical Society. Series B (Methodological)**. v. 39, n. 1, pp. 1-38, 1977.

FARCOT, Matthieu; LATOUR, Thibaud. Transitioning to Computer-Based Assessments: A Question of Costs. In F. Scheuermann & J. Björnsson (Org.). **The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing**. Luxembourg: Office for Official Publications of the European Communities, JRC, 2009.

FERNANDES, Paula Gabriela de Medeiros. **Sistema Computadorizado de Avaliação Adaptativa em Larga Escala (SCAALE)**. 2009. 152f. Monografia apresentada como requisito parcial para conclusão do Bacharelado em Ciência da Computação, Universidade de Brasília, 2009.

FRAGOSO, Tiago de Miranda. **Modelos Multidimensionais da Teoria de Resposta ao Item**. 2010. 111 f. Dissertação (Mestrado). Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2010.

GALLAGHER, Ann M.; BRIDGEMAN, Brent; CAHALAN, Cara. The effect of computer-based tests on racial/ethnic, gender, and language groups. **Research Report** n. 00-8, Educational Testing Service, USA, 2000.

GEEKIE. [internet], c2015. Disponível em: <<http://www.geekie.com.br>> Acesso em: 15 abr. 2015

GHADERI, Marzieh; MOGHOLI, Marzieh; AFSHIN, Soori. Comparing Between Computer based Tests and Paper-and-Pencil based Tests. **IJELS** 2(4), p. 36-38, 2014.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4ª ed., São Paulo: Atlas, 2002.

HAMBLETON, Ronald K.; SWAMINATHAN, Hariharan. **Item response theory: Principles and applications**. Boston: Kluwer, 1985.

_____; ROGERS, H. Jane. **Fundamentals of item response theory**. Sage Publications, 1991.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). Programa Internacional de Avaliação de Estudantes [internet]. c2011. Disponível em: <<http://portal.inep.gov.br/pisa-programa-internacional-de-avaliacao-de-alunos>>. Acesso em: 15 abr. 2015.

_____. Diretoria de Avaliação da Educação Básica (DAEB). **Teoria de Resposta ao Item**. Andrade, DF, Karino, CA. Nota técnica, 2012. Disponível em: <http://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_tri_enem_18012012.pdf>. Acesso em: 15 abr. 2015.

_____. **Relatório Nacional PISA 2012: Resultados brasileiros**. Fundação Santillana, 2013

KANG, Taehoon; COHEN, Allan S. IRT Model Selection Methods for Dichotomous Items. **Applied Psychological Measurement**. v. 31, n. 4, p. 331–358, 2007.

KLEIN, Ruben. Uma re-análise dos resultados do PISA: problemas de comparabilidade. **Ensaio: aval. pol. públ. Educ.** [online]. v.19, n.73, p. 717-

768, 2011. Disponível em: < <http://www.scielo.br/pdf/ensaio/v19n73/02.pdf>>
Acesso em: 15 abr. 2015.

_____; FONTANIVE, Nilma Santos. Avaliação em larga escala: uma proposta inovadora. **Em Aberto**, Brasília, ano 15, n. 66, abr./jun., 1995.

LAROS, Jacob Arie. O uso da Análise Fatorial: algumas diretrizes para pesquisadores. In: Luiz Pasquali (Org.) **Análise fatorial para pesquisadores**. Brasília: LabPAM Saber e Tecnologia, p. 141-160, 2004.

_____; PASQUALI, Luiz; RODRIGUES, Margarida M. M. **Análise da unidimensionalidade das provas do SAEB**. Relatório Técnico. Brasília: Centro de Pesquisa em Avaliação Educacional, UnB, 2000.

LOTTRIDGE, Susan; NICEWANDER, Alan; SCHULZ, Matt; MITZEL, Howard. **Comparability of Paper-based and Computer-based Tests: A Review of the Methodology (draft)**. CCSSO Technical Issues in Large Scale Assessment Comparability Research Group, Pacific Metrics Corporation, California, 2008.

LUECHT, Richard M.; SIRECI, Stephen G. **A Review of Models for Computer-Based Testing**. College Board Research Reports, 2011-12.

MC DONALD, Angus S. The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. **Computers & Education**, 39, p. 299-312, 2002.

MOE, Eli. Introducing Large-scale Computerised Assessment: Lessons Learned and Future Challenges. In: F. Scheuermann e J. Björnsson (Org.). **The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing**. Luxembourg: Office for Official Publications of the European Communities, JRC, 2009.

NUNES, Stella Maris Lemos. **A proficiência Matemática dos alunos brasileiros no PISA 2003: uma análise dos itens de incerteza**. 2013. 218 f. Tese (Doutorado). Programa de Pós-Graduação em Educação: Conhecimento e Inclusão Social, Faculdade de Educação, Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **Are students ready for a technology-rich world?** OECD Publishing, 2005.

_____. **PISA Data Analysis Manual: SPSS®** (Second Edition), OECD Publishing, 2009.

_____. **PISA Computer-Based Assessment of Student Skills in Science**, OECD Publishing, 2010.

_____. **PISA 2009 Technical Report**, OECD Publishing, 2012. Disponível em: <<http://dx.doi.org/10.1787/9789264167872-en>>. Acesso em: 15 abr. 2015.

_____. **PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy**, OECD Publishing, 2013. Disponível em: <http://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book_final.pdf>. Acesso em: 15 abr. 2015.

_____. **PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014)**, OECD Publishing, 2014a. Disponível em: <<http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-I.pdf>>. Acesso em: 15 abr. 2015.

_____. **PISA 2012 Technical Report**, OECD Publishing, 2014b. Disponível em: <<http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>>. Acesso em: 15 abr. 2015.

OSHIMA, T. Chris. The Effect of Speededness on Parameter Estimation in Item Response Theory. **Journal of Educational Measurement**. v. 31, p. 200–219, 1994.

PARSHALL, Cynthia G.; SPRAY, Judith A.; KOLOHN, John C.; DAVEY, Tim. **Practical Considerations in Computer-Based Testing**. Springer, 2002.

PASQUALI, Luiz. **Psicometria: Teoria dos testes na Psicologia e na Educação**. Petrópolis: Editora Vozes, 2003, 397p.

_____. **Psicometria**. **Rev Esc Enferm, USP**; 43(Esp), p. 992-9, 2008.

_____; PRIMI, Ricardo. Fundamentos da Teoria da Resposta ao Item – TRI. **Avaliação Psicológica**, (2)2, p. 99-110, 2003.

PELLEGRINO, James W.; QUELLMALZ, Edys S. Perspective on the Integration of Technology and Assessment. **Journal of Research on Technology in Education**, v. 43, n. 2, p. 119-134, 2010.

PEARSON. **Computer-Based & Paper-Pencil Test Comparability Studies**. Test, Measurement & Research Services, Bulletin, Issue 9, November 2009a.

_____. **Methods of Comparability Studies for Computerized and Paper-Based Tests**. Test, Measurement & Research Services, Bulletin, Issue 10, December 2009b.

_____. **Comparability of Computerized Adaptive and Paper-Pencil Tests**. Wang H, Shin, DC. Test, Measurement & Research Services, Bulletin, Issue 13, March 2010.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2012.
Disponível em: <<http://www.R-project.org/>>. Acesso em: 15 abr. 2015.

RECKASE, Mark D. Multidimensional Item Response Theory. **Statistics for Social and Behavioral Sciences**, Springer, 2009, 355p.

REGAZZI, Adair José; SILVA, Carlos Henrique Osório. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear. I. Dados no delineamento inteiramente casualizado. **Rev. Mat. Estat.**, São Paulo, v.22, n.3, p.33-45, 2004.

SCHWARZ, Gideon. Estimating the dimension of a model. **The Annals of Statistics**, v.6, n.2, p. 461–464, 1978.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação**. 4ª ed. rev. e atual. Florianópolis: UFSC, 2005, 138p.

SINGLETON, Chris. Computer-based assessment in education. **Educational and Child Psychology**, v. 18, n. 3, p. 58-74, 2001.

TEXAS EDUCATION AGENCY (TEA). **A Review of Literature on the Comparability of Scores Obtained from Examinees on Computer-Based and Paper-Based Tests**, 2008. Disponível em:

http://ritter.tea.state.tx.us/student.assessment/resources/techdigest/Technical_Reports/2008_literature_review_of_comparability_report.pdf. Acesso em 15 abr. 2015.

TEZZA, Rafael. **Modelagem multidimensional para mensurar qualidade em website de e-commerce utilizando a teoria da resposta ao item**. 2012. 182 f. Tese (Doutorado). Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2012.

THURSTONE, Louis Leon. **Multiple factor analysis**. Chicago, University of Chicago, Press, 1947.

VALLE, Raquel da Cunha. Teoria de Resposta ao Item. **Est. Aval. Educ.**, n.21, p. 7-92, 2000.

VITÓRIA, Florbela; ALMEIDA, Leandro S.; PRIMI, Ricardo. Unidimensionalidade em testes psicológicos: conceito, estratégias e dificuldades na sua avaliação. **Revista de Psicologia**, Vetor Editora, v. 7, n. 1, p. 1-7, 2006

APÊNDICE A

Frequências de respostas (continua)

| Item | Categoria de resposta | | | | Item | Categoria de resposta | | | |
|----------|-----------------------|-----|----|---------|----------|-----------------------|-----|-----|---------|
| | 0 | 1 | 2 | Válidas | | 0 | 1 | 2 | Válidas |
| PM00GQ01 | 389 | 6 | | 395 | PM915Q01 | 314 | 66 | | 380 |
| PM00KQ02 | 338 | 16 | | 354 | PM915Q02 | 237 | 142 | | 379 |
| PM033Q01 | 164 | 247 | | 411 | PM934Q01 | 332 | 57 | | 389 |
| PM034Q01 | 309 | 62 | | 371 | PM934Q02 | 241 | 147 | | 388 |
| PM155Q01 | 215 | 188 | | 403 | PM936Q01 | 277 | 138 | | 415 |
| PM155Q02 | 254 | 57 | 95 | 406 | PM936Q02 | 285 | 126 | | 411 |
| PM155Q03 | 371 | 12 | 19 | 402 | PM939Q01 | 195 | 209 | | 404 |
| PM155Q04 | 250 | 149 | | 399 | PM939Q02 | 264 | 138 | | 402 |
| PM192Q01 | 345 | 68 | | 413 | PM942Q01 | 248 | 157 | | 405 |
| PM273Q01 | 315 | 116 | | 431 | PM942Q02 | 362 | 42 | | 404 |
| PM305Q01 | 266 | 152 | | 418 | PM942Q03 | 346 | 14 | 41 | 401 |
| PM406Q01 | 388 | 25 | | 413 | PM948Q01 | 87 | 339 | | 426 |
| PM406Q02 | 401 | 10 | | 411 | PM948Q02 | 203 | 216 | | 419 |
| PM408Q01 | 337 | 90 | | 427 | PM948Q03 | 373 | 45 | | 418 |
| PM411Q01 | 322 | 76 | | 398 | PM949Q01 | 245 | 159 | | 404 |
| PM411Q02 | 264 | 126 | | 390 | PM949Q02 | 341 | 60 | | 401 |
| PM420Q01 | 305 | 121 | | 426 | PM949Q03 | 336 | 7 | 58 | 401 |
| PM423Q01 | 117 | 299 | | 416 | PM955Q01 | 216 | 174 | | 390 |
| PM442Q02 | 317 | 63 | | 380 | PM955Q02 | 348 | 39 | | 387 |
| PM446Q01 | 293 | 132 | | 425 | PM955Q03 | 371 | 10 | 3 | 384 |
| PM446Q02 | 415 | 6 | | 421 | PM957Q01 | 207 | 193 | | 400 |
| PM447Q01 | 246 | 186 | | 432 | PM957Q02 | 273 | 125 | | 398 |
| PM462Q01 | 371 | 2 | 2 | 375 | PM957Q03 | 372 | 23 | | 395 |
| PM464Q01 | 394 | 22 | | 416 | PM961Q02 | 381 | 28 | | 409 |
| PM474Q01 | 206 | 202 | | 408 | PM961Q03 | 270 | 138 | | 408 |
| PM496Q01 | 285 | 131 | | 416 | PM961Q05 | 235 | 70 | 101 | 406 |
| PM496Q02 | 214 | 202 | | 416 | PM967Q01 | 331 | 70 | | 401 |
| PM559Q01 | 215 | 205 | | 420 | PM967Q03 | 369 | 25 | | 394 |
| PM564Q01 | 259 | 128 | | 387 | PM982Q01 | 85 | 303 | | 388 |
| PM564Q02 | 280 | 107 | | 387 | PM982Q02 | 300 | 87 | | 387 |
| PM571Q01 | 289 | 118 | | 407 | PM982Q03 | 188 | 198 | | 386 |

Frequências de respostas (conclusão)

| Item | Categoria de resposta | | | |
|----------|-----------------------|-----|----|---------|
| | 0 | 1 | 2 | Válidas |
| PM603Q01 | 272 | 138 | | 410 |
| PM800Q01 | 66 | 349 | | 415 |
| PM803Q01 | 361 | 27 | | 388 |
| PM828Q01 | 379 | 41 | | 420 |
| PM828Q02 | 255 | 163 | | 418 |
| PM828Q03 | 346 | 72 | | 418 |
| PM906Q01 | 215 | 162 | | 377 |
| PM906Q02 | 283 | 32 | 58 | 373 |
| PM909Q01 | 122 | 287 | | 409 |
| PM909Q02 | 220 | 187 | | 407 |
| PM909Q03 | 344 | 63 | | 407 |
| CM002Q01 | 368 | 16 | 0 | 384 |
| CM002Q02 | 290 | 41 | | 331 |
| CM004Q01 | 354 | 112 | | 466 |
| CM004Q02 | 368 | 97 | | 465 |
| CM004Q03 | 343 | 118 | | 461 |
| CM005Q01 | 387 | 80 | | 467 |
| CM005Q02 | 325 | 137 | | 462 |
| CM005Q03 | 263 | 197 | | 460 |
| CM005Q04 | 344 | 78 | 29 | 451 |
| CM006Q01 | 229 | 214 | | 443 |
| CM006Q02 | 234 | 189 | 8 | 431 |
| CM006Q03 | 379 | 30 | | 409 |
| CM008Q01 | 337 | 27 | 24 | 388 |
| CM008Q02 | 355 | 10 | | 365 |
| CM011Q01 | 210 | 247 | | 457 |
| CM011Q02 | 436 | 16 | | 452 |
| CM011Q03 | 422 | 18 | | 440 |
| CM014Q01 | 323 | 85 | | 408 |
| CM014Q02 | 372 | 24 | | 396 |
| CM014Q03 | 351 | 21 | | 372 |
| CM015Q01 | 278 | 187 | | 465 |
| PM982Q04 | 272 | 113 | | 385 |
| PM985Q01 | 92 | 315 | | 407 |
| PM985Q02 | 254 | 153 | | 407 |
| PM985Q03 | 0 | 0 | | 0 |
| PM991Q01 | 173 | 232 | | 405 |
| PM991Q02 | 385 | 7 | 12 | 404 |
| PM992Q01 | 167 | 217 | | 384 |
| PM992Q02 | 346 | 36 | | 382 |
| PM992Q03 | 372 | 9 | | 381 |
| PM998Q02 | 203 | 180 | | 383 |
| PM998Q04 | 253 | 128 | | 381 |
| CM015Q02 | 447 | 0 | 14 | 461 |
| CM015Q03 | 374 | 38 | 34 | 446 |
| CM016Q01 | 359 | 106 | | 465 |
| CM020Q01 | 315 | 119 | 19 | 453 |
| CM020Q02 | 261 | 184 | | 445 |
| CM020Q03 | 374 | 69 | | 443 |
| CM020Q04 | 294 | 136 | | 430 |
| CM025Q01 | 324 | 118 | | 442 |
| CM025Q02 | 259 | 172 | | 431 |
| CM028Q03 | 365 | 56 | | 421 |
| CM035Q01 | 180 | 276 | | 456 |
| CM035Q02 | 356 | 96 | | 452 |
| CM035Q03 | 254 | 195 | | 449 |
| CM035Q04 | 433 | 13 | | 446 |
| CM036Q01 | 228 | 200 | | 428 |
| CM036Q02 | 417 | 2 | | 419 |
| CM036Q03 | 348 | 40 | | 388 |
| CM038Q03 | 185 | 242 | | 427 |
| CM038Q05 | 348 | 73 | | 421 |
| CM038Q06 | 334 | 43 | | 377 |

Fonte: elaboração da autora.

APÊNDICE B

Calibração dos itens (continua)

| Item | Uni | Bidimensional | | Item | Uni | Bidimensional | |
|----------|-------|---------------|--------|----------|-------|---------------|--------|
| | a1 | a1 | a2 | | a1 | a1 | a2 |
| PM00KQ02 | 1.294 | 1.439 | 0.59 | PM915Q02 | 1.884 | 2.167 | -1.643 |
| PM033Q01 | 1.025 | 1.031 | -0.165 | PM934Q01 | 1.772 | 1.921 | 0.78 |
| PM034Q01 | 1.383 | 1.393 | -0.466 | PM934Q02 | 0.573 | 0.565 | 0.001 |
| PM155Q01 | 1.486 | 1.844 | 0.563 | PM936Q01 | 1.973 | 1.962 | -0.221 |
| PM155Q02 | 2.115 | 2.112 | -0.027 | PM936Q02 | 2.248 | 2.22 | -0.329 |
| PM155Q03 | 3.26 | 3.529 | 0.731 | PM939Q01 | 1.202 | 1.216 | -0.103 |
| PM155Q04 | 0.874 | 0.86 | -0.38 | PM939Q02 | 1.792 | 1.79 | -0.172 |
| PM192Q01 | 1.446 | 1.475 | 0.25 | PM942Q01 | 0.649 | 0.677 | 0.261 |
| PM273Q01 | 1.03 | 1.007 | -0.272 | PM942Q02 | 2.147 | 2.119 | 0.28 |
| PM305Q01 | 0.85 | 0.829 | -0.195 | PM942Q03 | 2.551 | 2.501 | 0.068 |
| PM406Q01 | 2.456 | 2.434 | -0.291 | PM948Q01 | 1.199 | 1.234 | -1.141 |
| PM408Q01 | 1.191 | 1.181 | -0.206 | PM948Q02 | 1.252 | 1.217 | -0.487 |
| PM411Q01 | 2.616 | 2.606 | -0.324 | PM948Q03 | 2.231 | 2.21 | -0.102 |
| PM411Q02 | 0.436 | 0.426 | -0.154 | PM949Q01 | 1.187 | 1.28 | 0.286 |
| PM420Q01 | 1.236 | 1.213 | -0.284 | PM949Q02 | 0.984 | 0.968 | -0.115 |
| PM423Q01 | 1.161 | 1.209 | 0.106 | PM949Q03 | 1.908 | 1.842 | -0.094 |
| PM442Q02 | 1.681 | 1.739 | -0.714 | PM955Q01 | 1.033 | 1.06 | 0.128 |
| PM446Q01 | 1.912 | 1.927 | -0.249 | PM955Q02 | 1.378 | 1.371 | 0.198 |
| PM447Q01 | 1.044 | 1.19 | 0.251 | PM955Q03 | 1.684 | 1.625 | 0.536 |
| PM464Q01 | 2.822 | 2.741 | -0.111 | PM957Q01 | 0.79 | 0.86 | 0.429 |
| PM474Q01 | 1.096 | 1.11 | -0.116 | PM957Q02 | 0.811 | 0.937 | 0.717 |
| PM496Q01 | 1.574 | 1.58 | 0.017 | PM957Q03 | 2.06 | 2.123 | 0.398 |
| PM496Q02 | 1.261 | 1.303 | 0.131 | PM961Q02 | 2.287 | 2.338 | 0.423 |
| PM559Q01 | 0.79 | 0.753 | -0.394 | PM961Q03 | 1.272 | 1.278 | -0.593 |
| PM564Q01 | 0.543 | 0.543 | 0.038 | PM961Q05 | 1.008 | 0.994 | -0.194 |
| PM564Q02 | 0.829 | 0.836 | 0.021 | PM967Q01 | 1.798 | 1.78 | -0.261 |
| PM571Q01 | 1.007 | 1.075 | 0.311 | PM967Q03 | 1.199 | 1.205 | 0.266 |
| PM603Q01 | 1.311 | 1.313 | -0.028 | PM982Q01 | 1.052 | 1.12 | -2.068 |
| PM800Q01 | 0.747 | 0.719 | -0.229 | PM982Q02 | 1.359 | 1.35 | -0.747 |

Calibração dos itens (conclusão)

| Item | Uni | | Bidimensional | | Item | Uni | | Bidimensional | |
|----------|-------|-------|---------------|----|----------|-------|-------|---------------|----|
| | a1 | a1 | a1 | a2 | | a1 | a1 | a1 | a2 |
| PM803Q01 | 2.34 | 2.369 | -0.542 | | PM982Q03 | 0.897 | 0.854 | -0.489 | |
| PM828Q01 | 1.672 | 2.146 | -2.092 | | PM982Q04 | 0.699 | 0.693 | -0.13 | |
| PM828Q02 | 1.325 | 1.392 | -1.266 | | PM985Q01 | 1.299 | 1.302 | -0.351 | |
| PM828Q03 | 1.293 | 1.344 | -1.154 | | PM985Q02 | 1.502 | 1.503 | -0.274 | |
| PM906Q01 | 1.063 | 1.008 | -0.462 | | PM991Q01 | 1.099 | 1.085 | -0.049 | |
| PM906Q02 | 1.819 | 1.796 | -0.763 | | PM991Q02 | 2.052 | 2.07 | 0.464 | |
| PM909Q01 | 1.584 | 1.658 | 0.238 | | PM992Q01 | 0.918 | 0.893 | -0.437 | |
| PM909Q02 | 1.085 | 1.077 | -0.153 | | PM992Q02 | 1.36 | 1.34 | -0.869 | |
| PM909Q03 | 3.054 | 3.333 | -1.021 | | PM998Q02 | 1.328 | 1.657 | 0.732 | |
| PM915Q01 | 0.2 | 0.281 | 0.325 | | PM998Q04 | 0.287 | 0.282 | -0.023 | |
| CM002Q01 | 2.552 | 2.66 | 0.218 | | CM015Q01 | 0.821 | 0.81 | -0.275 | |
| CM002Q02 | 1.103 | 1.175 | 0.203 | | CM015Q03 | 2.504 | 2.585 | -0.826 | |
| CM004Q01 | 0.653 | 0.647 | -0.352 | | CM016Q01 | 1.95 | 2.122 | 0.286 | |
| CM004Q02 | 1.244 | 1.253 | 0.268 | | CM020Q01 | 0.969 | 0.959 | -0.3 | |
| CM004Q03 | 1.51 | 1.651 | 0.58 | | CM020Q02 | 0.071 | 0.033 | -0.32 | |
| CM005Q01 | 0.777 | 0.795 | -0.052 | | CM020Q03 | 1.395 | 1.381 | -0.24 | |
| CM005Q02 | 0.613 | 0.641 | 0.059 | | CM020Q04 | 0.53 | 0.542 | 0.062 | |
| CM005Q03 | 0.352 | 0.317 | -0.333 | | CM025Q01 | 1.598 | 1.648 | -0.876 | |
| CM005Q04 | 1.708 | 1.728 | -0.797 | | CM025Q02 | 1.631 | 1.618 | -0.524 | |
| CM006Q01 | 1.722 | 1.79 | -0.068 | | CM028Q03 | 1.212 | 1.208 | -0.017 | |
| CM006Q02 | 1.508 | 1.499 | -0.251 | | CM035Q01 | 0.987 | 0.991 | -0.153 | |
| CM006Q03 | 1.663 | 2.33 | 1.242 | | CM035Q02 | 0.999 | 1.07 | 0.277 | |
| CM008Q01 | 1.111 | 1.096 | -0.601 | | CM035Q03 | 0.61 | 0.623 | 0.005 | |
| CM011Q01 | 1.077 | 1.056 | -0.183 | | CM036Q01 | 0.347 | 0.303 | -0.355 | |
| CM011Q02 | 2.149 | 2.17 | 0.118 | | CM036Q03 | 0.25 | 0.342 | 0.547 | |
| CM011Q03 | 1.857 | 1.858 | 0.005 | | CM038Q03 | 0.534 | 0.505 | -0.297 | |
| CM014Q01 | 1.097 | 1.165 | 0.425 | | CM038Q05 | 1.03 | 1.036 | -0.685 | |
| CM014Q02 | 1.622 | 1.641 | -0.173 | | CM038Q06 | 1.859 | 1.869 | -0.033 | |
| CM014Q03 | 1.352 | 1.387 | 0.494 | | | | | | |

Fonte: elaboração da autora.

ANEXO

Demais itens de Matemática do PISA 2012 liberados para divulgação¹.

Disponível em:

<http://portal.inep.gov.br/internacional-novo-pisa-itens>

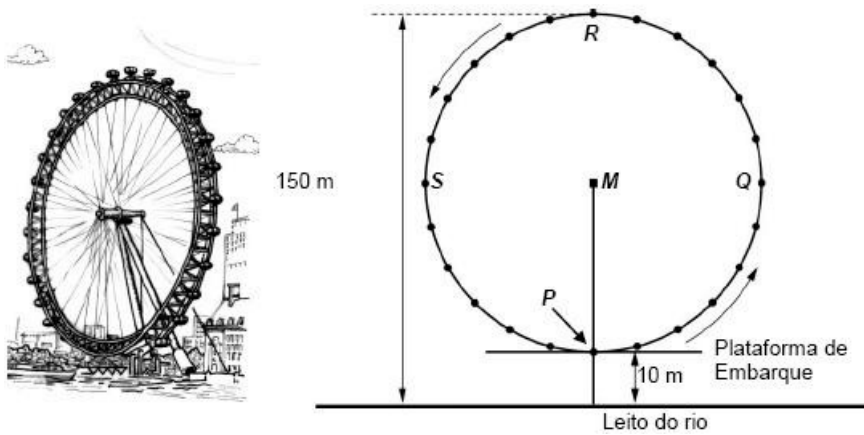
<http://www.oecd.org/pisa/pisaproducts/pisa-test-questions.htm>

<Acesso em: 15 jun. 2014>

¹ A ficha técnica dos itens foi elaborada pela autora com base nas informações dos documentos disponíveis nas páginas da internet citadas e no Relatório Técnico do PISA 2012 (OECD, 2014b, p. 406-409; 415-416). Por vezes as informações dos documentos foram divergentes e, nesses casos, considerou-se a informação final apresentada pela OCDE, mantendo a coerência com o que foi apresentado ao longo deste trabalho.

RODA GIGANTE

Na margem do rio fica uma roda gigante.
Veja a foto e o diagrama abaixo.



A roda gigante tem um diâmetro de 140 metros e o seu ponto mais alto está a 150 metros acima do leito do rio, em uma das margens do rio. Ela gira na direção indicada pela seta.

PM934Q01: RODA GIGANTE

A letra M , no diagrama, indica o centro da roda gigante.

Quantos metros (m) sobre o leito do rio está o ponto M ?

Resposta: m

| Item PM934Q01 | Correção |
|---|--|
| <p>Descrição: Calcular o comprimento baseado na informação em um desenho em 2D.</p> <p>Conteúdo matemático: Espaço e forma.</p> <p>Processo: Empregar.</p> <p>Contexto: Social.</p> | <p>Crédito completo Código 1: 80.</p> <p>Nenhum crédito Código 0: Outras respostas. Código 9: Em branco.</p> |
| <p>Tipo de resposta: Resposta construída codificada manualmente.</p> | <p>Dificuldade: 592,3 (nível 4).</p> |

ESCALANDO O MONTE FUJI

O Monte Fuji é um famoso vulcão inativo, no Japão.



PM942Q01: ESCALANDO O MONTE FUJI

O Monte Fuji está aberto ao público para escaladas a partir de 1º de julho até 27 de agosto, todos os anos. Cerca de 200 000 pessoas escalam o Monte Fuji nesse período.

Em média, aproximadamente quantas pessoas escalam o Monte Fuji por dia?

- A 340
- B 710
- C 3400
- D 7100
- E 7400

| Item PM942Q01 | Correção |
|---|--|
| <p>Descrição: Identificar uma taxa média diária, dados um número total e um período específico de tempo (datas fornecidas).</p> <p>Conteúdo matemático: Quantidade.</p> <p>Processo: Formular.</p> <p>Contexto: Social.</p> | <p>Crédito completo Código 1: C 3400.</p> <p>Nenhum crédito Código 0: Outras respostas. Código 9: Em branco.</p> |
| <p>Tipo de resposta: Múltipla escolha simples.</p> | <p>Dificuldade: 464,0 (nível 2).</p> |

PM942Q02: ESCALANDO O MONTE FUJI

A trilha Gotemba, que leva até o alto do Monte Fuji, tem cerca de 9 quilômetros (km) de comprimento.

Os caminhantes precisam retornar da caminhada de 18 km até às 8h da noite.

Toshi calcula que ele pode caminhar uma média de 1,5 km por hora, montanha acima, e, montanha abaixo, o dobro dessa velocidade. Essas velocidades incluem pausa para refeições e descanso.

Usando as velocidades calculadas por Toshi, qual é o último horário no qual ele pode iniciar sua caminhada de modo que ele possa estar de volta até às 8h da noite?

.....

| Item PM942Q02 | Correção |
|---|---|
| <p>Descrição: Calcular o início do tempo levado para uma caminhada, dadas duas velocidades diferentes, uma distância total para viajar e um tempo final.</p> <p>Conteúdo matemático: Mudanças e relações.</p> <p>Processo: Formular.</p> <p>Contexto: Social.</p> | <p>Crédito completo Código 1: 11h (da manhã) [11 h com ou sem o período ou equivalente escrita de tempo para o exemplo].</p> <p>Nenhum crédito Código 0: Outras respostas. Código 9: Em branco.</p> |
| <p>Tipo de resposta: Resposta construída codificada por especialistas.</p> | <p>Dificuldade: 641,6 (nível 5).</p> |

PM942Q03: ESCALANDO O MONTE FUJI

Toshi usou um pedômetro para contar seus passos ao longo da trilha Gotemba.

O pedômetro mostrou que ele havia dado 22 500 passos montanha acima.

Calcule o comprimento médio dos passos de Toshi em sua caminhada de 9 km na trilha Gotemba, montanha acima. Dê sua resposta em centímetros (cm).

Resposta: cm

| Item PM942Q03 | Correção |
|---|--|
| <p>Descrição: Dividir um comprimento dado em km por um número específico e expressar o quociente em cm.</p> <p>Conteúdo matemático: Quantidade.</p> <p>Processo: Empregar.</p> <p>Contexto: Social.</p> | <p>Crédito completo Código 2: 40.</p> <p>Crédito parcial Código 1: Resposta com o dígito 4 baseada numa conversão incorreta para centímetros. <input type="checkbox"/>0.4 [resposta dada em metros] <input type="checkbox"/>4000 [conversão incorreta]</p> <p>Nenhum crédito Código 0: Outras respostas. Código 9: Em branco.</p> |
| <p>Tipo de resposta: Resposta construída codificada manualmente.</p> | <p>Dificuldade: Crédito completo: 610,0 (nível 5). Crédito parcial: 591,3 (nível 4).</p> |

A CICLISTA HELENA



Helena acabou de receber uma nova bicicleta, com um velocímetro fixado no guidão.

O velocímetro pode indicar a distância que Helena percorre e sua velocidade média no trajeto.

PM957Q01: A CICLISTA HELENA

Em um passeio, Helena pedalou 4 km durante os 10 primeiros minutos e em seguida 2 km durante os 5 minutos seguintes.

Dentre as afirmações abaixo, qual está correta?

- A A velocidade média de Helena, durante os 10 primeiros minutos, foi superior à velocidade média durante os 5 minutos seguintes.
- B A velocidade média de Helena, durante os 10 primeiros minutos, foi igual à velocidade média durante os 5 minutos seguintes.
- C A velocidade média de Helena, durante os 10 primeiros minutos, foi inferior à velocidade média durante os 5 minutos seguintes.
- D Não é possível dizer nada sobre a velocidade média de Helena, a partir das informações fornecidas.

| Item PM957Q01 | Correção |
|---|--|
| <p>Descrição: Comparar as velocidades médias em função das distâncias e da duração do percurso.</p> <p>Conteúdo matemático: Mudanças e relações.</p> <p>Processo: Empregar.</p> <p>Contexto: Pessoal.</p> | <p>Crédito completo Código 1: B A velocidade média de Helena, durante os 10 primeiros minutos, foi igual à velocidade média durante os 5 minutos seguintes.</p> <p>Nenhum crédito Código 0: Outras respostas. Código 9: Em branco.</p> |
| <p>Tipo de resposta: Múltipla escolha simples.</p> | <p>Dificuldade: 440,5 (nível 2).</p> |

PM957Q02: A CICLISTA HELENA

Helena pedalou 6 km até a casa de sua tia. O velocímetro indicou que sua velocidade média foi de 18 km/h para todo o trajeto.

Dentre as afirmações abaixo, qual está correta?

- A Foram necessários 20 minutos para Helena chegar à casa de sua tia.
- B Foram necessários 30 minutos para Helena chegar à casa de sua tia.
- C Foram necessárias 3 horas para Helena chegar à casa de sua tia.
- D Não é possível dizer quanto tempo foi necessário para Helena chegar à casa de sua tia.

| Item PM957Q02 | Correção |
|---|---|
| <p>Descrição: Calcular a duração do trajeto a partir de uma velocidade média e de uma distância percorrida.</p> <p>Conteúdo matemático: Mudanças e relações.</p> <p>Processo: Empregar.</p> <p>Contexto: Pessoal.</p> | <p>Crédito completo Código 1: A Foram necessários 20 minutos para Helena chegar à casa de sua tia.</p> <p>Nenhum crédito Código 0: Outras respostas. Código 9: Em branco.</p> |
| Tipo de resposta: Múltipla escolha simples. | Dificuldade: 510,6 (nível 3). |

PM957Q03: A CICLISTA HELENA

Helena pedalou sua bicicleta de casa até o rio, que se encontra a 4 km. Foram necessários 9 minutos. Ela pedalou de volta para casa, usando um caminho mais curto, que levou 3 km. Ela precisou de somente 6 minutos.

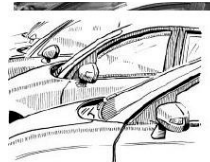
Qual foi a velocidade média de Helena (em km/h) neste passeio de ida e volta ao rio?

Velocidade média do passeio: km/h

| Item PM957Q03 | Correção |
|--|---|
| <p>Descrição: Calcular uma velocidade média de dois trajetos a partir de duas distâncias percorridas e da duração do percurso.</p> <p>Conteúdo matemático: Mudanças e relações.</p> <p>Processo: Empregar.</p> <p>Contexto: Pessoal.</p> | <p>Crédito completo Código 1: 28.</p> <p>Nenhum crédito Código 0: Outras respostas. <input type="checkbox"/> 28,3 [<i>Método incorreto: média das velocidades de 2 percursos (26,67 e 30)</i>].</p> |
| Tipo de resposta: Resposta construída codificada manualmente. | Código 9: Em branco. |
| | Dificuldade: 696,6 (nível 6). |

QUAL CARRO?

Cris acabou de receber sua carteira de habilitação e quer comprar seu primeiro carro.



A tabela abaixo mostra os detalhes de quatro carros que ela viu em uma concessionária de veículos local.

| Modelo: | Argentum | Brisa | Corinto | Doral |
|---------------------------------|----------|---------|---------|---------|
| Ano | 2003 | 2000 | 2001 | 1999 |
| Preço anunciado (em zeds) | 4800 | 4450 | 4250 | 3990 |
| Quilometragem (em quilômetros) | 105 000 | 115 000 | 128 000 | 109 000 |
| Capacidade do motor (em litros) | 1,79 | 1,796 | 1,82 | 1,783 |

PM985Q01: QUAL CARRO?

Cris quer um carro que preencha **todas** as seguintes condições:

- A quilometragem **não** deve ser maior do que 120 000 quilômetros.
- O ano de fabricação deve ser de 2000 ou mais recente.
- O preço anunciado **não** deve ser maior do que 4500 zeds.

Qual carro preenche as condições de Cris?

- A Argentum
- B Brisa
- C Corinto
- D Doral

| Item PM985Q01 | Correção |
|---|---|
| <p>Descrição: Selecionar um valor que preencha quatro condições numéricas/conjunto de afirmações dentro de um contexto financeiro.</p> <p>Conteúdo matemático: Incerteza e dados.</p> <p>Processo: Interpretar.</p> <p>Contexto: Pessoal.</p> | <p>Crédito completo Código 1: B Brisa.</p> <p>Nenhum crédito Código 0: Outras respostas. Código 9: Em branco.</p> |
| <p>Tipo de resposta: Múltipla escolha simples.</p> | <p>Dificuldade: 327,8 (abaixo do nível 1).</p> |

PM985Q02: QUAL CARRO?

Qual carro tem o motor de menor capacidade?

- A Argentum
- B Brisa
- C Corinto
- D Doral

| Item PM985Q02 | Correção |
|---|---|
| <p>Descrição: Escolher o menor número decimal em um conjunto de quatro, em um dado contexto.</p> <p>Conteúdo matemático: Quantidade.</p> <p>Processo: Empregar.</p> <p>Contexto: Pessoal.</p> | <p>Crédito completo Código 1: D Doral.</p> <p>Nenhum crédito Código 0: Outras respostas. Código 9: Em branco.</p> |
| Tipo de resposta: Múltipla escolha simples. | Dificuldade: 490,9 (nível 3). |

PM985Q03: QUAL CARRO?

Cris terá que pagar um adicional de 2,5% em taxas sobre o preço anunciado do carro.

Quanto é a taxa adicional para o Argentum?

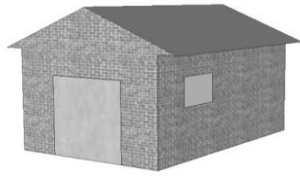
Taxa adicional em zeds:

| Item PM985Q03 | Correção |
|--|---|
| <p>Descrição: Calcular 2,5% de um valor em milhares de zeds, em um contexto financeiro.</p> <p>Conteúdo matemático: Quantidade.</p> <p>Processo: Empregar.</p> <p>Contexto: Pessoal.</p> | <p>Crédito completo Código 1: 120.</p> <p>Nenhum crédito Código 0: Outras respostas. <input type="checkbox"/> 2,5% de 4800 zeds [<i>Necessita de cálculo</i>]. Código 9: Em branco.</p> |
| Tipo de resposta: Resposta construída codificada manualmente. | Dificuldade: 552,6 (nível 4). |

GARAGEM

A série "básica" de um fabricante de garagens inclui modelos com apenas uma janela e uma porta.

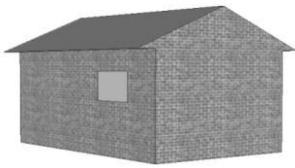
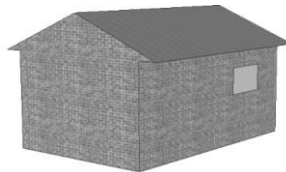
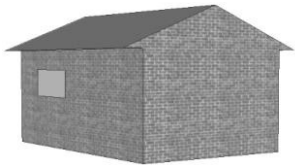
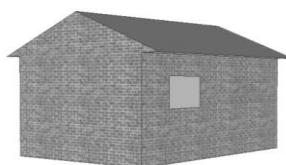
Jorge escolhe o seguinte modelo da série "básica". A posição da janela e da porta são as mostradas aqui.



PM991Q01: GARAGEM

As ilustrações abaixo mostram modelos "básicos" diferentes, vistos de trás. Apenas uma destas ilustrações corresponde ao modelo acima, escolhido por Jorge.

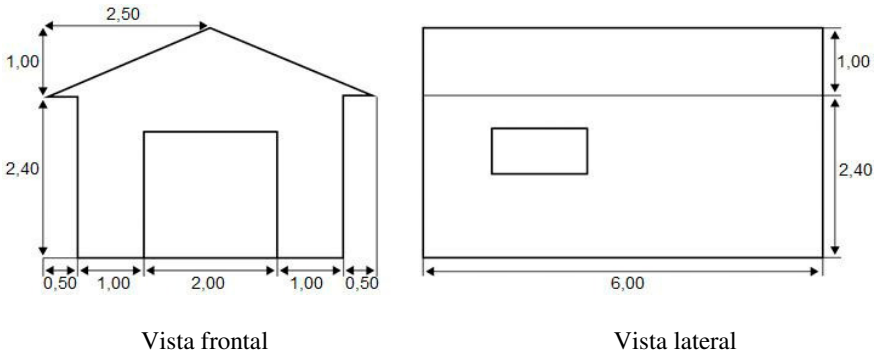
Qual modelo Jorge escolheu? Circule A, B, C ou D.

A**B****C****D**

| Item PM991Q01 | Correção |
|--|--|
| <p>Descrição: Usar habilidade espacial para identificar a imagem em 3D correspondente a outra imagem dada em 3D.</p> <p>Conteúdo matemático: Espaço e forma.</p> <p>Processo: Interpretar.</p> <p>Contexto: Ocupacional.</p> | <p>Crédito completo Código 1: C [Ilustração C].</p> <p>Nenhum crédito Código 0: Outras respostas. Código 9: Em branco.</p> |
| Tipo de resposta: Múltipla escolha simples. | Dificuldade: 419,6 (nível 1). |

PM991Q02: GARAGEM

As duas plantas abaixo mostram as dimensões, em metros, da garagem que Jorge escolheu.



O telhado é feito de duas partes retangulares idênticas.

Calcule a área **total** do telhado. Demonstre seu raciocínio.

.....

.....

.....

.....

| Item PM991Q02 | Correção |
|---|---|
| <p>Descrição: Interpretar um plano e calcular a área do retângulo utilizando o teorema de Pitágoras ou cálculos de medida.</p> <p>Conteúdo matemático: Espaço e forma.</p> <p>Processo: Empregar.</p> <p>Contexto: Ocupacional.</p> | <p>Crédito completo Código 21: Qualquer valor entre 31 e 33, com ou sem a demonstração do raciocínio. <i>[Unidades (m²) não são necessárias].</i> <input type="checkbox"/> $12 \times 2,6 = 31,2$. <input type="checkbox"/> $\sqrt{\square\square\square\square\square\square}$. <input type="checkbox"/> $12 \times 2,69 = 31,28 \text{ m}^2$. <input type="checkbox"/> $12 \times 2,7 = 32,4 \text{ m}^2$.</p> <p>Crédito parcial Código 11: Os cálculos mostram o uso correto do teorema de Pitágoras, mas há erros de cálculo, ou usa comprimento incorreto ou não dobra a área do telhado. <input type="checkbox"/> $2,5^2 + 1^2 = 6$, <input type="checkbox"/> $\sqrt{6} = \square 9,39$ <i>[Usa corretamente o teorema de Pitágoras com erro de cálculo].</i> <input type="checkbox"/> $2^2 + 1^2 = 5$, <input type="checkbox"/> $6 \times \sqrt{5} = \square 6,8 \text{ m}^2$ <i>[Usa o comprimento incorreto].</i> <input type="checkbox"/> $6 \times 2,6 = 15,6$ <i>[Não dobra a área do telhado].</i> Código 12: Os cálculos não mostram o uso do teorema de Pitágoras, mas usa valores de largura do telhado que fazem sentido (por exemplo, qualquer valor entre 2,6 e 3) e completa o resto do cálculo corretamente. <input type="checkbox"/> $2,75 \times 12 = 33$. <input type="checkbox"/> $3 \times 6 \times 2 = 36$.</p> <p>Nenhum crédito Código 00: Outras respostas. <input type="checkbox"/> $2,5 \times 12 = 30$ <i>[Considera a largura do telhado fora da faixa aceitável, que é de 2,6 a 3].</i> <input type="checkbox"/> $3,5 \times 6 \times 2 = 42$ <i>[Considera a largura do telhado fora da faixa aceitável, que é de 2,6 a 3].</i> Código 99: Em branco.</p> |
| <p>Tipo de resposta: Resposta construída codificada por especialistas.</p> | <p>Dificuldade Crédito completo: 687,3 (nível 6). Crédito parcial: 663,2 (nível 5).</p> |

CM015Q01: PRODUÇÃO DE CD

ICPT Programme for International Student Assessment 2012

PRODUÇÃO DE CD

A Zedtec fornece um serviço de cópias de CD.
 Há dois métodos para copiar CD: a duplicação e a replicação.
 O gráfico e a calculadora de preço mostram os preços para os dois métodos, em função da quantidade de CD a copiar.
 Podes escrever valores diferentes na célula «Número de cópias», para descobrires o custo exato da duplicação e da replicação.

Preço da replicação
 Preço da duplicação

O teu orçamento: replicação
 O teu orçamento: duplicação

CALCULADORA DE PREÇO

Preço da replicação: 420,00 zeds
 Preço da duplicação: 360,00 zeds

Número de cópias: 100

Questão 1: PRODUÇÃO DE CD CM015Q01
 Qual é a diferença de preço entre a duplicação e a replicação de 500 CD?

80 zeds
 110 zeds
 140 zeds
 940 zeds

Item CM015Q01

Conteúdo matemático: Quantidade.

Processo: Empregar.

Contexto: Ocupacional.

Tipo de resposta: Múltipla escolha simples.

Dificuldade: 498,5 (nível 3).

CM015Q02: PRODUÇÃO DE CD

pe-P-T Programme for International Student Assessment 2012

PRODUÇÃO DE CD

A Zentec fornece um serviço de cópias de CD. Há dois métodos para copiar CD: a duplicação e a replicação. O gráfico e a calculadora de preço mostram os preços para os dois métodos, em função da quantidade de CD a copiar. Podes escrever valores diferentes na célula «Número de cópias», para descobrires o custo exato da duplicação e da replicação.

Preço da cópia de CD, usando a duplicação ou a replicação

Preço da replicação (zeds)

Preço da duplicação (zeds)

Número de cópias

○ O teu orçamento: replicação
■ O teu orçamento: duplicação

| CALCULADORA DE PREÇO | |
|----------------------|-------------|
| Preço da replicação | 420,00 zeds |
| Preço da duplicação | 360,00 zeds |

Número de cópias: 100

Questão 2: PRODUÇÃO DE CD - CM015Q02

Usa os gráficos e a calculadora de preço para encontrares a fórmula que permite determinar o preço da replicação. Escreve os dois valores que faltam na fórmula abaixo, para mostrares a relação entre o preço, P , e o número de cópias efetuadas, n , para a replicação.

$P = \square \cdot n + \square$

?

→

Item CM015Q02

Conteúdo matemático: Mudanças e Relações.

Processo: Formular.

Contexto: Ocupacional.

Tipo de resposta: Resposta construída autocodificada.

Dificuldade:

Crédito completo: 700,7 (nível 6).

Crédito parcial: 685,8 (nível 6).

CM015Q03: PRODUÇÃO DE CD

pt-PT Programme for International Student Assessment 2012

PRODUÇÃO DE CD

A Zedtec fornece um serviço de cópias de CD.
 Há dois métodos para copiar CD: a duplicação e a replicação.
 O gráfico e a calculadora de preço mostram os preços para os dois métodos, em função da quantidade de CD a copiar.
 Pode escrever valores diferentes na célula «Número de cópias», para descobrir o custo exato da duplicação e da replicação.

Preço da cópia de CD, usando a duplicação ou a replicação

Preço da replicação (zeds) vs. Número de cópias

Preço da duplicação (zeds) vs. Número de cópias

● O teu orçamento: replicação
 ■ O teu orçamento: duplicação

CALCULADORA DE PREÇO

Preço da replicação: 420,00 zeds
 Preço da duplicação: 360,00 zeds

Número de cópias: 100

Questão 3: PRODUÇÃO DE CD CM015Q03

Na sua publicidade a Zedtec afirma: *A duplicação fica mais barata para pequenas encomendas de cópias (até 500 CD).*
 Explica por que razão o número, 500 CD, da afirmação está incorreto.

Qual é o número máximo de cópias que tornaria esta afirmação correta?
 Número de cópias =

Item CM015Q03

Conteúdo matemático: Mudanças e Relações.

Processo: Interpretar.

Contexto: Ocupacional.

Tipo de resposta: Resposta construída codificada por especialistas.

Dificuldade:

Crédito completo: 658,6 (nível 5).

Crédito parcial: 577,2 (nível 4).

CM038Q03: ÍNDICE DE MASSA CORPORAL

pe-P7 Programme for International Student Assessment 2012

ÍNDICE DE MASSA CORPORAL

A Teresa e o Raul estão a fazer um projeto sobre massa corporal e saúde. Encontram o sítio da internet, Zedisaúde, sobre saúde e índice de Massa Corporal (IMC).

A página da internet tinha algumas estatísticas que mostravam de que forma os valores do IMC variam para rapazes e raparigas entre os 9 e os 19 anos.

Estadísticas sobre o Índice de Massa Corporal (IMC)
Estes gráficos mostram os valores do IMC para os 5% inferiores, o valor mediano (50%) e os 5% superiores, em raparigas e rapazes de diferentes idades.

Valores de IMC para rapazes e raparigas dos 9 aos 19 anos

Podes clicar nos botões abaixo, para mostrar ou esconder qualquer uma das seis curvas.

| Idade (anos) | 5% inferiores | Mediano | 5% superiores |
|--------------|---------------|---------|---------------|
| 9 | 14 | 16 | 18 |
| 10 | 14 | 17 | 19 |
| 11 | 14 | 18 | 20 |
| 12 | 14 | 19 | 21 |
| 13 | 14 | 20 | 22 |
| 14 | 14 | 21 | 23 |
| 15 | 14 | 22 | 24 |
| 16 | 14 | 23 | 25 |
| 17 | 14 | 24 | 26 |
| 18 | 14 | 25 | 27 |
| 19 | 14 | 26 | 28 |

RAPARIGAS
 - 5% inferiores
 - mediano
 - 5% superiores

RAPAZES
 - 5% inferiores
 - mediano
 - 5% superiores

Questão 1: ÍNDICE DE MASSA CORPORAL CM038Q03

A Teresa faz as seguintes afirmações acerca dos dados apresentados nos gráficos. As afirmações da Teresa são confirmadas pelos gráficos? Para cada afirmação, seleciona «Verdadeira» ou «Falsa».

Afirmção

Tanto para as raparigas como para os rapazes, a amplitude dos valores do IMC aumenta dos 9 aos 19 anos.

Depois dos 17 anos, o valor de IMC dos 5% inferiores é maior para as raparigas do que para os rapazes.

Verdadeira Falsa

Item CM038Q03

Conteúdo matemático: Incerteza e dados.

Processo: Interpretar.

Contexto: Social.

Tipo de resposta: Múltipla escolha complexa.

Dificuldade: 468,8 (nível 2).

CM038Q05: ÍNDICE DE MASSA CORPORAL

pe-PT Programme for International Student Assessment 2012

ÍNDICE DE MASSA CORPORAL

A Teresa e o Raul estão a fazer um projeto sobre massa corporal e saúde. Encontram o sítio da internet, Zedsaúde, sobre saúde e Índice de Massa Corporal (IMC).

Uma outra página da internet tinha alguns dados sobre as taxas de obesidade na Zedlândia.

O teu IMC **Estatísticas** **Dados da Zedlândia**

Taxas de Obesidade na Zedlândia

A tabela apresenta a percentagem de jovens de 12 a 19 anos, na Zedlândia, em cada uma das quatro classificações de IMC, em 2000, 2005 e 2010. O gráfico mostra as taxas de obesidade para cada um destes três anos.

| Classificação de IMC: | 2000 | 2005 | 2010 |
|-----------------------|-------|-------|-------|
| Peso abaixo do normal | 14,1% | 14,8% | 14,3% |
| Normal | 57,5% | 53,5% | 51,5% |
| Excesso de peso | 16,3% | 17,1% | 16,1% |
| Obeso | 12,1% | 14,6% | 18,1% |

Taxas de obesidade na Zedlândia

Classificação de IMC:

| anos | 2000 | 2005 | 2010 |
|-------|-------|-------|-------|
| Obeso | 12,1% | 14,6% | 18,1% |

Questão 2: ÍNDICE DE MASSA CORPORAL CM038Q05

Qual é uma alteração significativa das classificações de IMC nos jovens dos 12 aos 19 anos, na Zedlândia, entre 2000 e 2010? Justifica a tua resposta, baseando-te no(s) valor(es) da tabela.

Item CM038Q05

Conteúdo matemático: Incerteza e dados.

Processo: Interpretar.

Contexto: Social.

Tipo de resposta: Resposta construída codificada por especialistas.

Dificuldade: 641,1 (nível 5).

CM038Q06: ÍNDICE DE MASSA CORPORAL

PCPT Programme for International Student Assessment 2012

ÍNDICE DE MASSA CORPORAL

A Teresa e o Raul estão a fazer um projeto sobre massa corporal e saúde. Encontram o site da internet, Zeisau.de, sobre saúde e índice de Massa Corporal (IMC).

Uma outra página da internet tinha alguns dados sobre as taxas de obesidade na Zedlândia.

O teu IMC **Estatísticas** **Dados da Zedlândia**

Taxas de Obesidade na Zedlândia

A tabela apresenta a percentagem de jovens de 12 a 19 anos, na Zedlândia, em cada uma das quatro classificações de IMC, em 2000, 2005 e 2010.

O gráfico mostra as taxas de obesidade para cada um destes três anos.

| Classificação de IMC | 2000 | 2005 | 2010 |
|-----------------------|-------|-------|-------|
| Peso abaixo do normal | 14,1% | 14,8% | 14,3% |
| Normal | 57,5% | 53,5% | 51,5% |
| Excesso de peso | 16,3% | 17,1% | 16,1% |
| Obeso | 12,1% | 14,6% | 18,1% |

Taxas de obesidade na Zedlândia

Obesidade **CM038Q06**

Questão 3: ÍNDICE DE MASSA CORPORAL

A Teresa diz: «O gráfico mostra que a taxa de obesidade em 2010 é cerca de quatro vezes maior do que a taxa em 2000.»

A Teresa não tem razão. Explica o que há de enganador no gráfico, que pode ter levado a Teresa a tirar esta conclusão.

Item CM038Q06

Conteúdo matemático: Incerteza e dados.

Processo: Interpretar.

Contexto: Social.

Tipo de resposta: Resposta construída codificada por especialistas.

Dificuldade: 660,4 (nível 5).