



UNIVERSIDADE ESTADUAL DE CAMPINAS

IFGW – Instituto de Física "Gleb Wataghin"

PECIM – Programa de Pós-Graduação Multiunidades em Ensino de Ciências e
Matemática

Emiliano Augusto Chagas

DESEMPENHO EM ITENS DE MATEMÁTICA DO ENEM:

CARGA COGNITIVA E SEXO

PERFORMANCE IN ENEM MATHEMATIC ITEMS:

COGNITIVE LOAD AND GENDER

CAMPINAS

2019

EMILIANO AUGUSTO CHAGAS

DESEMPENHO EM ITENS DE MATEMÁTICA DO ENEM: CARGA COGNITIVA E SEXO

PERFORMANCE IN ENEM MATHEMATIC ITEMS: COGNITIVE LOAD AND GENDER

Tese apresentada ao Programa de Pós-graduação Multiunidades em Ensino de Ciências e Matemática – PECIM, da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Ensino de Ciências e Matemática, na Área de Ensino de Ciências e Matemática.

Dissertation presented to the Multi-unit Postgraduate Program in Science and Mathematics Teaching – PECIM, at State University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Science and Mathematics Teaching, in the Area of Science and Mathematics Teaching.

Orientador: Prof. Dr. Mauricio Urban Kleinke

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO EMILIANO AUGUSTO CHAGAS E ORIENTADA PELO PROF. DR. MAURICIO URBAN KLEINKE.

CAMPINAS

2019

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Física Gleb Wataghin
Lucimeire de Oliveira Silva da Rocha - CRB 8/9174

C346d Chagas, Emiliano Augusto, 1981-
Desempenho em itens de matemática do ENEM : carga cognitiva e sexo /
Emiliano Augusto Chagas. – Campinas, SP : [s.n.], 2019.

Orientador: Maurício Urban Kleinke.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Física
Gleb Wataghin.

1. Exame Nacional do Ensino Médio (Brasil). 2. Desempenho matemático.
3. Sexo - Diferenças. 4. Matemática - Estudo e ensino. 5. Avaliação de larga
escala. 6. Teoria da carga cognitiva. I. Kleinke, Maurício Urban, 1958-. II.
Universidade Estadual de Campinas. Instituto de Física Gleb Wataghin. III.
Título.

Informações para Biblioteca Digital

Título em outro idioma: Performance in ENEM mathematic items : cognitive load and
gender

Palavras-chave em inglês:

National Secondary Education Examination (Brazil)

Mathematics performance

Sex differences

Mathematics - Study and teaching

Large-scale assessment

Cognitive load theory

Área de concentração: Ensino de Ciências e Matemática

Titulação: Doutor em Ensino de Ciências e Matemática

Banca examinadora:

Maurício Urban Kleinke [Orientador]

Renato Fernandes Cantão

Carlos Miguel da Silva Ribeiro

Samuel Rocha de Oliveira

Jader Otávio Dalto

Data de defesa: 12-09-2019

Programa de Pós-Graduação: Multiunidades em Ensino de Ciências e Matemática

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0003-2510-1919>

- Currículo Lattes do autor: <http://lattes.cnpq.br/2550880151712885>

FOLHA DE APROVAÇÃO

COMISSÃO EXAMINADORA

Prof Dr Mauricio Urban Kleinke
Presidente da Comissão Examinadora

Prof Dr Renato Fernandes Cantão
Membro titular externo (UFSCAR)

Prof Dr Jader Otávio Dalto
Membro titular externo (UTFPR)

Prof Dr Carlos Miguel da Silva Ribeiro
Membro titular interno (Unicamp)

Prof Dr Samuel Rocha de Oliveira
Membro titular interno (Unicamp)

Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

“What is a man? A miserable little pile of secrets!”

Dracula – Symphony of The Night

Agradecimentos

Agradeço profundamente: a minha família e meus amigos, por todo o convívio e apoio durante todos esses anos; aos meus colegas do IFSP por toda a compreensão e ajuda nesses últimos semestres; a todos meus alunos e ex-alunos, por me moldarem o professor que sou hoje; ao meu orientador Mauricio, por me colocar no eixo no momento certo; e a Mariana, por estar comigo e me transformar todo dia em uma pessoa melhor.

RESUMO

Dentre todas as avaliações em larga escala no Brasil, o Exame Nacional do Ensino Médio (Enem) é aquela que possui maior relevância, na perspectiva do candidato, uma vez que essa avaliação passou a ser a maior porta de entrada para o ensino superior público, por meio do Sistema de Seleção Unificada (Sisu) a partir do ano de 2010. O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), disponibiliza em sua página eletrônica os microdados relativos às aplicações do Enem, das provas e dos questionários socioeconômicos. Esse banco de dados possibilita um campo vasto de pesquisas, uma vez que existem muitas variáveis disponíveis para análise. Muitos estudos abordam a diferença de desempenho entre homens e mulheres em avaliações de matemática, apontando em geral escores maiores para os homens. Em contrapartida, em provas de leitura, redação e interpretação, as mulheres se sobressaem em relação aos homens. Os microdados do Enem fornecem resultados em conformidade com esses estudos. Uma vez que a contextualização permeia os itens (questões) de matemática do Enem, seria esperado que esse pudesse ser um fator relevante para um melhor desempenho feminino. Nesse trabalho exploramos os efeitos da modificação de itens de matemática do Enem no desempenho de homens e mulheres, utilizando a teoria da carga cognitiva e técnicas de redução de carga cognitiva estranha. Para tal propósito, foram selecionados 24 itens de matemática do Enem, entre os anos de 2010 e 2015, e cada um deles foi submetido à técnicas de redução de carga cognitiva. Foram montados quatro instrumentos de avaliação com proporções diferentes de itens do Enem e itens modificados, com carga cognitiva estranha reduzida, além de um pequeno questionário socioeconômico. Esse instrumento foi aplicado para 691 estudantes distribuídos em oito localidades. Os resultados indicam que a redução de carga cognitiva estranha foi benéfica para o desempenho dos estudantes em seis itens, enquanto que para dois itens o escore diminuiu. Para o desempenho entre os sexos, a redução de carga cognitiva estranha se mostrou mais eficaz para o desempenho das mulheres, nove itens cujo desempenho masculino em suas versões originais eram superiores aos femininos, após a redução de carga cognitiva estranha, se mostraram com escore indistinguível em relação ao sexo. Para quatro itens o efeito foi inverso, ou seja, a redução de carga cognitiva beneficiou os homens. Ao recortar duas provas, uma exclusivamente com itens do Enem e outra com itens exclusivamente submetidos à redução de carga cognitiva, o efeito inteiro da prova teve um impacto maior para as mulheres, cujo escore aumentou mais do que os homens. Variáveis socioeconômicas como escolaridade do pai, escolaridade da mãe e renda familiar também foram estudadas. Existem indícios de que a redução de carga cognitiva beneficia mais estudantes cujos pais possuem mais escolaridade ou renda maior. Chegamos a conclusão que a construção do item possui um impacto considerável no desempenho dos estudantes e, uma vez que o Enem atualmente possui o papel de selecionar alunos para a maioria das vagas em instituições públicas de ensino superior, a confecção dos itens pode ser determinante na vida dos candidatos.

Palavras-Chave: Exame Nacional do Ensino Médio (Brasil). Desempenho Matemático. Sexo – Diferenças. Matemática – Estudo e Ensino. Avaliação de Larga Escala. Teoria da Carga Cognitiva.

ABSTRACT

Among all large-scale assessments in Brazil, the National High School Exam (Enem) is the one that has the most relevance from the candidate's perspective, since this exam has become the largest gateway to public higher education, through the Unified Selection System (Sisu) from 2010 on. The National Institute of Educational Studies and Research Anísio Teixeira (Inep), makes available on its website the microdata related to Enem applications, tests and socioeconomic questionnaires. This database enables a vast field of research as there are many variables available for analysis. Many studies address the difference in performance between men and women on math tests, generally pointing higher scores for men. In contrast, in reading, writing and interpretation tests, women excel over men. Enem microdata provides results in accordance with these studies. Since contextualization permeates Enem's math items (questions), it would be expected that this could be a relevant factor for better female performance. In this paper we explore the effects of Enem's mathematical item modification on the performance of men and women using cognitive load theory and strange cognitive load reduction techniques. For this purpose, 24 Enem math items were selected between 2010 and 2015, and each of them underwent cognitive load reduction techniques. Four evaluation instruments were assembled with different proportions of Enem items and modified items, with reduced strange cognitive load, and a small socioeconomic questionnaire. This instrument was applied to 691 students distributed in eight locations. The results indicate that the reduction of strange cognitive load was beneficial for students' performance in six items, while for two items the score decreased. For performance between the sexes, the reduction of strange cognitive load was more effective for women's performance, nine items whose male performance in its original versions were superior to the female, after the reduction of strange cognitive load, showed indistinguishable score in relation to sex. For four items the effect was inverse, ie, the reduction in cognitive burden benefited men. By cutting out two assessments, one exclusively with Enem items and one with items exclusively subjected to cognitive load reduction, the entire effect of the test had a greater impact on women, whose score increased more than men. Socioeconomic variables such as father's education, mother's education and family income were also studied. There is evidence that reduced cognitive burden benefits more students whose parents have more education or higher income. We concluded that the construction of the item has a considerable impact on student performance and, since Enem currently has the role of selecting students for most positions in public higher education institutions, the making of items can be a determining factor in candidates' lives.

Keywords: National Secondary Education Examination (Brazil). Mathematics Performance. Sex Differences. Mathematics – Study and Teaching. Large-scale Assessment. Cognitive Load Theory.

Lista de Tabelas

Tabela 1 - Estratégias de redução de carga cognitiva em itens	63
Tabela 2 - Formato dos Itens do Enem por Objeto de Conhecimento	72
Tabela 3 - Competências da Matriz de Referência do Novo Enem no Período de 2009 a 2016	72
Tabela 4 - Distribuição dos respondentes em cada localidade pelo sexo	80
Tabela 5 - Distribuição dos respondentes pela escolaridade do pai e da mãe	81
Tabela 6 - Distribuição dos respondentes pela renda familiar total	82
Tabela 7 - Grupo focal e grupo de referência com candidatos	88
Tabela 8 - Grupo focal e grupo de referência com chance	88
Tabela 9 - Taxa de acerto e desvio padrão em cada item por tipo de prova	91
Tabela 10 - Alfa de Cronbach em cada uma das provas	92
Tabela 11 - Taxa de acerto médio em itens originais e reduzidos	93
Tabela 12 - Número de respondentes por sexo em cada prova	93
Tabela 13 - Escore de homens e mulheres em cada prova	94
Tabela 14 - Diferença de desempenho em cada item categorizado por sexo e versão do item	94
Tabela 15 - Escore médio em cada prova por renda familiar	96
Tabela 16 - Escore médio em cada prova por escolaridade da mãe e do pai	98
Tabela 17 - Escore médio em cada prova por categoria de escolaridade familiar	99
Tabela 18 - Distância de Cohen entre os tipos de prova	100
Tabela 19 - Distância de Cohen entre os sexos para cada prova	101
Tabela 20 - Distância de Cohen entre os sexos entre as provas	101
Tabela 21 - Distância de Cohen entre os estratos escolares dos pais para cada prova ..	102
Tabela 22 - Distância de Cohen entre os estratos escolares dos pais entre as provas ..	103
Tabela 23 - Distância de Cohen entre os estratos socioeconômicos para cada prova ..	103
Tabela 24 - Distância de Cohen entre os estratos socioeconômicos entre as provas ...	104

Tabela 25 - Categorização da qualidade de um item pela sua discriminação	160
Tabela 26 - Categorização de valores de Alfa de Cronbach por autores	161

Lista de Quadros

Quadro 1 - Item 2 do artigo em sua forma tradicional e submetido à redução de carga cognitiva	74
Quadro 2 - Item 13 do artigo em sua forma tradicional e submetido à redução de carga cognitiva	75
Quadro 3 - Item 8 da Prova e sua respectiva versão Reduzida	77
Quadro 4 - Item 17 da Prova e sua respectiva versão Reduzida	78
Quadro 5 - Item 1 da Prova e sua respectiva versão Reduzida	109
Quadro 6 - Valores de Razão de Chance e p-valor para o Item 1 por tipo de item e sexo, e taxa de acerto por sexo	109
Quadro 7 - Item 3 da Prova e sua respectiva versão Reduzida	111
Quadro 8 - Valores de Razão de Chance e p-valor para o Item 3 por tipo de item e sexo, e taxa de acerto por sexo	111
Quadro 9 - Item 5 da Prova e sua respectiva versão Reduzida	112
Quadro 10 - Valores de Razão de Chance e p-valor para o Item 5 por tipo de item e sexo, e taxa de acerto por sexo	112
Quadro 11 - Item 6 da Prova e sua respectiva versão Reduzida	114
Quadro 12 - Valores de Razão de Chance e p-valor para o Item 6 por tipo de item e sexo, e taxa de acerto por sexo	114
Quadro 13 - Item 9 da Prova e sua respectiva versão Reduzida	115
Quadro 14 - Valores de Razão de Chance e p-valor para o Item 9 por tipo de item e sexo, e taxa de acerto por sexo	116
Quadro 15 - Item 10 da Prova e sua respectiva versão Reduzida	117
Quadro 16 - Valores de Razão de Chance e p-valor para o Item 10 por tipo de item e sexo, e taxa de acerto por sexo	117
Quadro 17 - Item 18 da Prova e sua respectiva versão Reduzida	118
Quadro 18 - Valores de Razão de Chance e p-valor para o Item 18 por tipo de item e sexo, e taxa de acerto por sexo	119

Quadro 19 - Item 2 da Prova e sua respectiva versão Reduzida	153
Quadro 20 - Valores de Razão de Chance e p-valor para o Item 2 por tipo de item e sexo, e taxa de acerto por sexo	153
Quadro 21 - Item 4 da Prova e sua respectiva versão Reduzida	155
Quadro 22 - Valores de Razão de Chance e p-valor para o Item 4 por tipo de item e sexo, e taxa de acerto por sexo	155
Quadro 23 - Item 7 da Prova e sua respectiva versão Reduzida	156
Quadro 24 - Valores de Razão de Chance e p-valor para o Item 7 por tipo de item e sexo, e taxa de acerto por sexo	156
Quadro 25 - Item 8 da Prova e sua respectiva versão Reduzida	157
Quadro 26 - Valores de Razão de Chance e p-valor para o Item 8 por tipo de item e sexo, e taxa de acerto por sexo	158
Quadro 27 - Item 11 da Prova e sua respectiva versão Reduzida	159
Quadro 28 - Valores de Razão de Chance e p-valor para o Item 11 por tipo de item e sexo, e taxa de acerto por sexo	159
Quadro 29 - Item 12 da Prova e sua respectiva versão Reduzida	160
Quadro 30 - Valores de Razão de Chance e p-valor para o Item 12 por tipo de item e sexo, e taxa de acerto por sexo	161
Quadro 31 - Item 13 da Prova e sua respectiva versão Reduzida	162
Quadro 32 - Valores de Razão de Chance e p-valor para o Item 14 por tipo de item e sexo, e taxa de acerto por sexo	162
Quadro 33 - Item 14 da Prova e sua respectiva versão Reduzida	164
Quadro 34 - Valores de Razão de Chance e p-valor para o Item 14 por tipo de item e sexo, e taxa de acerto por sexo	165
Quadro 35 - Item 15 da Prova e sua respectiva versão Reduzida	166
Quadro 36 - Valores de Razão de Chance e p-valor para o Item 15 por tipo de item e sexo, e taxa de acerto por sexo	166
Quadro 37 - Item 16 da Prova e sua respectiva versão Reduzida	167

Quadro 38 - Valores de Razão de Chance e p-valor para o Item 16 por tipo de item e sexo, e taxa de acerto por sexo	168
Quadro 39 - Item 17 da Prova e sua respectiva versão Reduzida	169
Quadro 40 - Valores de Razão de Chance e p-valor para o Item 17 por tipo de item e sexo, e taxa de acerto por sexo	169
Quadro 41 - Item 19 da Prova e sua respectiva versão Reduzida	171
Quadro 42 - Valores de Razão de Chance e p-valor para o Item 19 por tipo de item e sexo, e taxa de acerto por sexo	171
Quadro 43 - Item 20 da Prova e sua respectiva versão Reduzida	172
Quadro 44 - Valores de Razão de Chance e p-valor para o Item 20 por tipo de item e sexo, e taxa de acerto por sexo	173
Quadro 45 - Item 21 da Prova e sua respectiva versão Reduzida	174
Quadro 46 - Valores de Razão de Chance e p-valor para o Item 21 por tipo de item e sexo, e taxa de acerto por sexo	175
Quadro 47 - Item 22 da Prova e sua respectiva versão Reduzida	176
Quadro 48 - Valores de Razão de Chance e p-valor para o Item 22 por tipo de item e sexo, e taxa de acerto por sexo	177
Quadro 49 - Item 23 da Prova e sua respectiva versão Reduzida	178
Quadro 50 - Valores de Razão de Chance e p-valor para o Item 23 por tipo de item e sexo, e taxa de acerto por sexo	179
Quadro 51 - Item 24 da Prova e sua respectiva versão Reduzida	180
Quadro 52 - Valores de Razão de Chance e p-valor para o Item 24 por tipo de item e sexo, e taxa de acerto por sexo	180

LISTA DE ABREVIATURAS

ANA	Avaliação Nacional de Alfabetização
ANEB	Avaliação Nacional da Educação Básica
COTUCA	Colégio Técnico da Unicamp
Enem	Exame Nacional do Ensino Médio
IFES	Institutos Federais de Ensino Superior
IFSP	Instituto Federal de São Paulo
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais
LDB	Lei de Diretrizes e Bases
PCN	Parâmetros Curriculares Nacionais
PISA	Programa Internacional de Avaliação de Alunos
Prouni	Programa Universidade Para Todos
RC	Razão de Chance
SAEB	Sistema de Avaliação da Educação Básica
Sisu	Sistema Unificado de Vagas
SM	Salários Mínimos
TCT	Teoria Clássica dos Testes
TRI	Teoria de Resposta ao Item
Unicamp	Universidade Estadual de Campinas

Sumário

Introdução.....	17
1. Teoria da Carga Cognitiva	21
1.1. Introdução à Teoria da Carga Cognitiva	22
1.1.1. Memória	24
1.2. Categorias de Carga Cognitiva	28
1.3. Efeitos da Modificação da Carga Cognitiva	34
1.4. Considerações sobre a Carga Cognitiva.....	39
2. Avaliação.....	41
2.1. Desempenho, Avaliação em Larga Escala e Sexo - resumo	43
2.2. Avaliações em Larga Escala - panorama	44
2.2.1. Enem - panorama	46
2.2.2. Pisa - panorama.....	51
2.3. Desempenho e Sexo em Avaliações em Larga Escala.....	53
3. Contextualização e TCC	61
3.1. Contextualização e Carga Cognitiva Estranha.....	62
3.2. Contextualização no Enem e no Pisa	64
3.3. Letramento Matemático - panorama	64
3.3.1. Contextualização no Pisa	65
3.3.2. Contextualização no Enem	70
3.3.3. Contexto em Testes de Múltipla Escolha e Carga Cognitiva Estranha.....	74
3.3.4. Aplicações em Itens do Enem	76
4. Procedimentos metodológicos	79
4.1. Sujeitos de Pesquisa	80
4.2. Instrumento de Avaliação.....	82
4.3. Métodos Estatísticos - panorama	86
5. Resultados	90
5.1. Estatística Descritiva e Teoria Clássica dos Testes	90
5.2. Distância de Cohen.....	99
5.3. Testes de Hipótese	104
6. O Efeito da Redução de Carga Cognitiva nos Itens	108
7. Considerações Finais.....	122
8. Referências Bibliográficas	128
APÊNDICE I - TALE e TCLE	137

APÊNDICE II - Folha de Respostas e Questionário Socioeconômico	139
APÊNDICE III - Teoria Clássica dos Testes e Alfa de Crombach	140
APÊNDICE IV - Estatística clássica dos testes para cada prova	144
APÊNDICE V - Desempenho de Homens e Mulheres por Item em Cada Prova.....	148
APÊNDICE VI - Desempenho em cada item estratificado por sexo e por tipo	149
APÊNDICE VII - Análise dos outros Itens da Prova	153
ANEXO I - Histórico do SAEB	182
ANEXO II - Matrizes: Competência, Habilidade e Análise de Desempenho	186
ANEXO III - Matriz de Referência de Matemática e Suas Tecnologias	190
ANEXO IV – Parecer Consubstanciado do Comitê de Ética em Pesquisa.....	190

Introdução

As avaliações em larga escala suscitam discussões que transcendem o propósito inicial destas, que em geral são aferições de qualidade em sistemas de educação ou avaliações classificatórias em concursos. A avaliação pode se dar sobre sistemas educacionais ou sobre os indivíduos, não sendo simples para um indivíduo ser avaliado, bem como para os gestores dos distintos sistemas educacionais. Dada a dimensão, a ordem de grandeza e as implicações sociais e políticas que essas avaliações podem ter, não são apenas os candidatos que estão sob uma lupa.

No caso de avaliações educacionais examina-se não só o aluno, mas toda uma cadeia na qual o estudante está posicionado na ponta. As respostas fornecidas pelos alunos, bem como os construtos diversos exigidos pela prova, estão direta ou indiretamente ligados às condições físicas de seu colégio, às condições socioeconômicas do estudante e de seus colegas, ao corpo docente, à carga de trabalho dos professores, ao perfil da coordenação e da direção, ao currículo associado a essa avaliação e outros elementos.

As discussões mencionadas anteriormente tomam dimensões em diversas esferas, não apenas educacionais, uma vez que as avaliações em larga escala, especialmente pelo número de pessoas atingidas e por seus resultados, envolvem questões políticas. Outro uso dessas avaliações está na esfera científica e acadêmica, já que o número de informações produzidas nas aplicações dessas provas é muito grande.

Esse trabalho é construído sob a perspectiva da utilização de microdados de avaliações em larga escala, juntamente com a análise de variáveis socioeconômicas e pinceladas políticas. Estamos particularmente interessados na questão da diferença no desempenho entre homens e mulheres em provas de matemática, que é um assunto de interesse acadêmico, político e social. Estudos nesse campo são realizados há mais de 60 anos (ANASTASI, 1958), e já naquela época os pesquisadores combinavam o uso dos resultados com as variáveis socioeconômicas e étnicas. Pesquisas mais recentes nesse tema se utilizam de diversas avaliações em larga escala em meta análise, além de fazerem divisões de desempenho por países e comparações desses escores com medidas de igualdade entre sexos (ELSE-QUEST; HYDE; LINN, 2010). O ponto que as autoras desse último trabalho levantam é que uma consequência da discrepância entre o

desempenho nas provas de matemática, favoráveis aos homens frente às mulheres, é o papel social da mulher na sociedade em que ela está inserida. Em outras palavras, o desempenho das mulheres em avaliações de ciências exatas (não apenas matemática) acaba proporcionando uma restrição no mercado de trabalho feminino, particularmente em cargos associados à ciência, tecnologia, engenharia e matemática, que no meio internacional também é conhecido por STEM (Science, Technology, Engineering and Mathematics). Ocorre então um ciclo vicioso: moças possuem desempenho médio inferior aos rapazes em provas de matemática, o que acarreta menos entrada de mulheres em cursos de exatas, que implica em menos mulheres no mercado de trabalho em áreas de exatas (por gosto ou autocracia), que levam jovens mulheres a não se identificarem nessas carreiras, o que pode induzir a um desempenho médio menor das mulheres em provas de matemática.

Em contrapartida, as autoras chamam a atenção para o fato de que as mulheres possuem um desempenho robustamente melhor em provas de leitura, interpretação e redação frente aos homens. Essa solidez dos dados transcende os países quanto à riqueza, distribuição de renda e religião.

Uma avaliação em larga escala, embora não tenha como papel principal evocar todas essas discussões, acaba fornecendo informações valiosas. No caso do Brasil, é possível encontrar dados comparativos entre países e também entre estados brasileiros por meio dos relatórios do PISA (OECD, 2013). Esses relatórios contemplam – além do desempenho dos estudantes nas áreas de matemática, ciências e linguagens – as análises dos resultados dos questionários sobre a situação familiar e escolar respondidos por alunos, pais e professores. Detalhes dessa prova serão discutidos no decorrer do presente trabalho, sendo que para nossos propósitos as informações essenciais desses relatórios são as seguintes: os alunos brasileiros na prova de matemática possuem desempenho superior às alunas, e alunas brasileiras na prova de leitura possuem desempenho superior aos alunos (INEP, 2014; INEP, 2016).

Esses últimos dados são bem interessantes, especialmente ao considerarmos que a contextualização está presente na prova de matemática PISA. Os enunciados contêm uma quantidade considerável de informação para que os candidatos avaliem a situação problema e cheguem a uma conclusão. Se as mulheres possuem um desempenho superior comparadas aos homens na área de leitura e interpretação essas habilidades, em teoria,

poderiam fornecer uma vantagem às meninas em problemas de matemática com enunciados contextualizados.

Uma maneira de verificar esse questionamento seria trabalhar com os itens do próprio PISA, mas infelizmente poucos itens são abertos e o acesso aos microdados não é simples. Por outro lado, o Enem possui um banco de itens de matemática bem vasto desde sua reformulação em 2009, além do que os microdados estão abertos na página eletrônica do INEP. Esse microdados fornecem, além da informação sobre a prova de cada aluno, as respostas dos questionários socioeconômicos. Mais ainda, assim como o PISA, o Enem utiliza situações problema e contextualização em seus itens de matemática.

Os dados do Enem relativos à comparação de desempenho na prova de matemática entre homens e mulheres está em conformidade com os resultados do Brasil no PISA, ou seja, homens possuem desempenho melhor frente às mulheres. No espectro de humanidades, nossa análise dos dados do Enem indica um melhor desempenho em redação e linguagens por parte das mulheres, assim como ocorre no PISA, o que sugere que poderia vir a ocorrer um melhor desempenho feminino em questões de matemática que apresentassem maior grau de contextualização.

Portanto, uma característica em média positiva para as mulheres, como a leitura, não possui efeito benéfico nas provas de matemática que contém leitura e contextualização. Levantamos uma hipótese de que fosse possível que, exatamente a presença de leitura e interpretação nas questões de matemática, possa ser um fator de baixo desempenho das mulheres frente aos homens.

Na literatura buscamos uma teoria que contemplasse questões sobre mobilização de processos cognitivos, exemplos e aplicações tanto em matemática quanto em avaliações e que apresentasse relevância acadêmica. A Teoria da Carga Cognitiva (TCC) atendeu a todos esses requisitos, portanto utilizamos resultados dessa teoria para testar a nossa hipótese. Tendo essas informações em mente, junto com as discussões anteriores, surge a questão orientadora desse estudo.

A redução, ou até a eliminação, do texto e/ou do contexto nos itens de matemática do Enem impacta de que maneira o desempenho de homens e mulheres?

Aqui a redução do texto significa efetivamente reduzir a quantidade de leitura e interpretação dos respondentes, de maneira geral pode implicar também em redução do

contexto do problema. Mais para frente será discutido que esse texto e informações em demasia, para a TCC, estão intimamente ligados com uma carga cognitiva irrelevante ao processo de resolução, chamada de carga cognitiva estranha. A pergunta orientadora seria equivalente fazendo essa troca.

Ao analisar o desempenho de estudantes em itens de matemática com contexto estamos interessados nas mobilizações cognitivas dos alunos face a esses itens, sendo que o referencial utilizado será o da TCC criado por John Sweller. Construiremos novos itens utilizando redução de carga cognitiva estranha, baseado no trabalho feito por Gillmore, Poggio e Embretson (2015). Por último, as análises de desempenho entre homens e mulheres em problemas de matemática serão baseadas nos trabalhos de Else-Quest, Hyde e Linn.

Para essa pesquisa, a estrutura da Tese foi dividida da seguinte maneira: No capítulo 1 apresentamos a TCC, suas categorias e algumas considerações referentes à modificações nas cargas cognitivas; No capítulo 2 são apresentadas as avaliações em larga escala, dentre elas o PISA e o Enem, em especial as provas de matemática desses exames, e também a questão do desempenho entre homens e mulheres em avaliações em larga escala.

O capítulo 3 descreve o papel da contextualização na TCC, e como é possível utilizar maneiras de modificar a carga cognitiva em itens de matemática, utilizando exemplos de referencial teórico e mostrando aplicação em alguns itens da prova do Enem. As metodologias, incluindo sujeitos de pesquisa e construção do instrumento de avaliação, são discutidas no capítulo 4.

Os resultados são apresentados no capítulo 5, as discussões sobre os efeitos da redução de carga cognitiva estranha em situações problema com contexto estão no capítulo 6 e as conclusões do estudo, bem como as reflexões finais, são apresentadas no capítulo 7.

1. Teoria da Carga Cognitiva

Ciências cognitivas tem por finalidade estudar como o nosso cérebro consegue aprender, bem como compreender suas capacidades, limitações e particularidades. Uma das ações importantes do cérebro no ambiente escolar é a memória, sendo que dentre alguns dos formatos descritos para memória, temos Memória de Curto Prazo, Memória de Trabalho ou Memória de Longo Prazo (COWAN, 2008). Essas separações são resultado de pesquisas da área, e aparecem com certa frequência no cotidiano. A capacidade limitada do nosso cérebro e de seu processamento (MILLER, 1956) traz consequências na área de ensino. Diversos pesquisadores se dedicam a compreender melhor a arquitetura cognitiva humana e estudar os efeitos – benéficos e prejudiciais – associados aos processos de ensino e aprendizagem.

Dentre essas pesquisas, a Teoria da Carga Cognitiva (TCC) está presente nas cinco maiores revistas de psicologia do mundo e conta com um contingente internacional e diversificado de pesquisadores, como apontado por Souza (2010). Como essa teoria contempla exemplos baseados em conceitos matemáticos aplicados em testes, além de possuir mecanismos para a melhoria de métodos de instrução, ela foi escolhida para fornecer suporte teórico para nossa pesquisa.

Considere o seguinte exemplo, estudado na TCC, que contempla os aspectos mais relevantes desse campo, além de motivar uma reflexão sobre o tema uma vez que se trata de um exercício de matemática para alunos da educação básica. Em 2001, Ayres propôs apresentar o seguinte experimento para alunos do oitavo e nono anos: efetuar a operação $-3.(-4 - 5x) - 2.(3x - 4)$.

Professores de matemática, em algum momento do ensino fundamental, devem apresentar aos alunos expressões algébricas e operações elementares com essas expressões. Portanto, a operação proposta por Ayres poderia estar em qualquer lista de exercícios ou até mesmo na prova de um professor que estivesse administrando esse conteúdo, uma vez que esse é exatamente o construto¹ pretendido.

¹ “traços, aptidões ou características supostamente existentes e abstraídos de uma variedade de comportamentos que tenham significado educacional (ou psicológico). Assim, fluência verbal, rendimento

Para resolver esse problema, o aluno deve aplicar a multiplicação de cada número antes dos dois parêntesis de modo distributivo. Ele deve também acertar o sinal dos resultados das multiplicações, e por fim deve somar os termos (polinômios) semelhantes. Durante a correção de uma questão desse tipo, o professor consegue verificar, dentre os alunos que não chegaram na resposta, quais foram as etapas problemáticas.

O trabalho de Ayres (2001) mostrou que os alunos não erravam o resultado da operação multiplicativa da distributiva de maneira uniforme. A primeira multiplicação, $-3 \cdot -4 = 12$, tinha uma taxa de erro menor do que $-3 \cdot -5x = 15x$, o produto $-2 \cdot 3x = -6x$ tinha uma taxa de erro menor do que $-2 \cdot -4 = 8$. Além disso, as operações do segundo parêntese tinham uma taxa de erro maior do que as operações do primeiro parêntese. Uma reflexão que deve ser feita nesse momento: estudantes que acertam, por exemplo, duas multiplicações, e erram as outras duas multiplicações, dominam ou não dominam o construto pretendido? Esse estudo se predispôs a mostrar que existem fatores cognitivos que afetam o desempenho dos alunos e que fogem do escopo do construto.

Pensando no outro espectro, professores de matemática com bastante experiência, e às vezes até alguns alunos, conseguem resolver essas expressões algébricas como se elas não oferecessem qualquer dificuldade. Como será discutido nas próximas subseções, a TCC tenta abarcar de maneira eficaz a maior quantidade de explicações desses efeitos colocados em reflexão, sejam elas em avaliações ou conteúdos.

1.1. Introdução à Teoria da Carga Cognitiva

A TCC foi desenvolvida por John Sweller e seus colegas na Universidade de Nova Gales do Sul a partir da década de 80. O termo Carga Cognitiva aparece pela primeira vez no artigo "O que os alunos aprendem enquanto resolvem problemas de matemática?" (OWEN; SWELLER, 1985) e se consolidou no trabalho de Sweller de 1988 intitulado "A Carga Cognitiva durante a resolução de problemas: efeitos sobre a aprendizagem".

Em seu trabalho de 1998, Sweller define a Carga Cognitiva como: "Um construto representando a carga que a realização de uma tarefa particular impõe no sistema

escolar, aptidão mecânica, inteligência, motivação, agressividade, entre outros, são construtos." (VIANNA, 2013 p. 35)

cognitivo [...] A Carga mental se refere à carga imposta pela demanda da tarefa” (SWELLER, 1998, p. 266, apud Souza 2010, p. 42). Tomemos o exemplo apresentado anteriormente para uma melhor compreensão: simplificar a expressão $-3.(-4 - 5x) - 2.(3x - 4)$ pode ter um impacto diferente para distintos indivíduos. Para um estudante do sétimo ano, que está iniciando seus estudos em cálculo com variáveis, essa será uma tarefa que exigirá muito mais recursos da Memória de Trabalho do que para um estudante do ensino médio que já está trabalhando com funções quadráticas, e este por sua vez ainda reserva mais recursos para a Memória de Trabalho do que para um professor de matemática. Desse modo, o termo Carga Cognitiva não se refere apenas ao volume de informações a serem processadas pela Memória de Trabalho, mas a demanda que essas informações produzem sobre a Memória de Trabalho de um indivíduo.

O pressuposto fundamental da TCC é que para os métodos de instrução sejam efetivos, os professores, materiais, ou instruções, precisam levar em consideração o sistema cognitivo humano. Mais ainda, alerta para que os métodos instrucionais sejam confeccionados em conformidade com os princípios operacionais básicos do sistema cognitivo humano (CHANDLER; SWELLER, 1991; SWELLER, 1988; 1994; 2003; SWELLER; CHANDLER, 1991) e desse modo, possibilitem a integração dos conhecimentos cognitivos necessários e reduza cargas cognitivas desnecessárias para facilitar e melhorar processos cognitivos associados ao ensino e à aprendizagem.

Assim, a TCC amplia a eficácia do material instrucional agindo como facilitadora de aprendizagem ao filtrar e direcionar recursos cognitivos em direção a atividades que são efetivamente relevantes à aprendizagem. Em outras palavras, trata-se de um provedor de diretrizes e/ou protocolos que auxiliam na construção e apresentação da informação de tal forma que, ao otimizar a performance intelectual de um indivíduo, encoraje suas atividades de aprendizado (KIRSCHNER, 2002).

Muito embora a TCC tenha se desenvolvido desde a década de 80, portanto independente do advento computacional em larga escala, atualmente um grande foco da teoria da carga cognitiva está em materiais instrucionais de informática, além de acadêmicos que trabalham com a interface entre educação e estrutura cognitiva, e também por designers instrucionais², que são profissionais especializados em desenvolver

² Do inglês *instructional designer*.

métodos, técnicas e recursos para maior eficiência dos processos de ensino e aprendizagem.

Em particular, para os propósitos dessa tese, a TCC no contexto da matemática aparece em alguns trabalhos, cuja estrutura lógica de apresentação dos conceitos nos utilizamos. Katheeb (2008) e Takir (2011) discutem com profundidade, além da TCC e matemática, arquitetura cognitiva. Em português vale citar os trabalhos de Rauber (2016) pela apresentação didática dos conceitos, e de Souza (2010) que contempla a evolução histórica da TCC em que ele utiliza informações trocadas diretamente com John Sweller.

No Brasil, a professora Liane Tarouco e diversos colegas têm desenvolvido trabalhos sobre a importância de estudar a teoria do grupo de John Sweller na era digital (SANTOS; TAROUCO, 2007), muito embora de Souza já tenha apontado que as pesquisas brasileiras sobre a TCC sejam quase inexistentes.

1.1.1. Memória

Para uma boa compreensão da TCC é necessário abordar dois elementos muito importantes que integram o sistema cognitivo e a aprendizagem: a Memória de Trabalho, (antigamente conhecida como Memória de Curto Prazo) (COWAN, 2008), e a Memória de Longa Duração, às vezes chamada de memória de Longo Prazo.

A estrutura cognitiva nos seres humanos inclui esses dois principais sistemas de memória mencionados acima. Todas as atividades humanas relacionadas ao mundo do trabalho ou aprendizagem dependem da Memória de Trabalho, da Memória de Longa Duração e da interação entre as duas (CLARKE et al., 2005)

A definição padrão de Memória de Trabalho refere-se ao armazenamento e manipulação momentânea de informações que são processadas para realizar tarefas cognitivas (BADDELEY, 1996). Refere-se também a um sistema que foi desenvolvido para manutenção e gerenciamento de curto prazo da informação necessária para o desempenho das tarefas complexas, como a aprendizagem (BADDELEY, 1998).

Memória de Longa Duração é o repositório de conhecimentos e habilidades mais permanentes e inclui todos os elementos contidos na memória que atualmente não estão sendo usados, mas que são necessários (KIRSCHNER, 2002). Algumas linhas de

pesquisa sugerem que características tais como capacidade ilimitada: itens armazenados na Memória de Longa Duração podem ser permanentes, confrontando a etimologia do nome da memória, e que os itens armazenados podem ser prontamente recuperados quando um estímulo for proporcionado (BADDELEY, 1982; BADDELEY, 1996). Além disso, o autor aponta que a informação na Memória de Longa Duração nunca se apaga, apenas torna-se menos acessível ao longo do tempo.

John Sweller (2003) ainda propõe que o conteúdo da Memória de Longa Duração seja desconhecido até que essa informação seja alocada para a memória de trabalho visando o processamento de uma determinada tarefa.

Estes dois sistemas de Memória (de Trabalho e de Longa Duração) operam em conjunto: à medida que a Memória de Trabalho processa novos aprendizados num indivíduo, os novos conhecimentos e habilidades são armazenados na Memória de Longa Duração. Desse modo, quando esse indivíduo adquire experiência em domínio, seu repositório de conhecimento em Memória de Longa Duração se expande (CLARKE et al., 2005).

Uma das grandes preocupações dos pesquisadores da área está na limitação dos processos cognitivos humanos. Como já reportado em pesquisas, a cognição humana é limitada por sete informações³, com um desvio padrão de dois, a serem mantidos em Memória de Trabalho (MILLER, 1956) e por quatro informações a serem processadas (COWAN, 2001). Face a essa situação, a Teoria da Carga Cognitiva é desenhada para adaptar os limites da Memória de Trabalho e explorar suas potencialidades. CLARKE et al (2005) definem a teoria da carga cognitiva como:

"um conjunto universal de princípios instrucionais e guias baseados em evidências que oferecem os métodos mais eficientes de projetar e entregar ambientes instrucionais de modo que tentam utilizar da melhor maneira a limitação da capacidade da memória de trabalho" (p. 342, tradução nossa⁴)

Como apresentado nos últimos parágrafos, a TCC está preocupada em estudar a relação entre a Memória de Longa Duração, a Memória de Trabalho e o aprendizado, e

³ Em seu original, Miller chama de itens. Como no nosso trabalho a palavra item está associada à questão (ou problema), substituímos por informação.

⁴ a universal set of instructional principles and evidence based guidelines that offer the most efficient methods to design and deliver instructional environments in ways that best utilize the limited capacity of working memory

como materiais instrucionais interagem com esse sistema cognitivo. Isso implica que a instrução e os *designers* instrucionais precisam considerar as limitações da Memória de Trabalho, para que por fim, informação possa ser armazenada de modo eficaz na Memória de Longa Duração (CHANDLER; SWELLER, 1991; 1988; 1994; SWELLER; CHANDLER, 1991).

Uma vez que o sistema cognitivo humano possui limitações intrínsecas, é necessário se perguntar se existe alguma maneira de contornar essa restrição. Materiais e construtos, após serem familiarizados por indivíduos através de interações duradouras, passam a ser utilizados com naturalidade e pouco esforço. O que está por trás dessa situação são os esquemas, que são estruturas inseridas na Memória de Longa Duração, elementos de informação bem categorizados e que essencialmente são a base da experiência (SWELLER; Van MERRIENBOER; PAAS, 1998).

Na perspectiva da TCC, os esquemas têm duas funções. A primeira é fornecer um mecanismo para a organização do conhecimento e armazenamento, sendo esta função essencial para organizar o conhecimento existente em um modo categorizado (esquema). A segunda função é reduzir a carga da Memória de Trabalho, uma vez que os esquemas permitem que muitos elementos sejam tratados como um único elemento na Memória de Trabalho; Como resultado desse efeito, mais capacidade de Memória de Trabalho fica disponível (SWELLER, 2003).

Outro mecanismo capaz de reduzir a Memória de Trabalho é a automação, definida por Sweller como "a habilidade de processar informação sem o controle consciente da Memória de Trabalho" (SWELLER, 2003). Dirigir um carro com câmbio manual ou digitar textos no computador podem ser considerados exemplos de automação. Um exemplo escolar em matemática é a tabuada, em que alunos são submetidos ao processo de repetição das multiplicações numéricas entre dois números de um dígito até que os resultados dessas multiplicações apareçam de maneira imediata.

Com a aquisição e automação de esquemas, tarefas familiares são realizadas com precisão, enquanto tarefas desconhecidas podem ser aprendidas com a máxima eficiência. Sem automação, uma tarefa já encontrada anteriormente pode ser totalmente realizada, mas o desempenho dessa atividade será provavelmente mais lento (SWELLER; Van MERRIENBOER; PAAS, 1998). Portanto, a capacidade de automatizar um esquema é uma condição crítica para a utilização de esquemas e redução da carga cognitiva.

Em outras palavras, o objetivo do programa de instrução é liberar a Memória de Trabalho – que é limitada – do esforço mental irrelevante necessário para integrar novos conhecimentos e habilidades nos esquemas em Memória de Longa Duração (CLARKE et al., 2005). No exemplo de Ayres (2001) visto no começo desse capítulo, o efeito que levou ao erro parcial de algumas multiplicações executadas pelos estudantes é explicado pelo autor por meio da sobrecarga da Memória de Trabalho. Em contrapartida, os professores, que são especialistas no assunto, já desenvolveram esquemas e automação para essa tarefa de operações algébricas simples.

1.1.1.1. Elemento e interação

O conceito de elementos, e a interação destes, é importante para a compreensão da carga cognitiva. Um elemento é essencialmente algo que precisa ser entendido e aprendido. Clarke et al. (2005) definem como

"uma propriedade do conteúdo instrucional que reflete a medida em que vários componentes de conteúdo devem ser mantidos e processados simultaneamente na Memória de Trabalho para serem aprendidos ou alcançar um objetivo de desempenho" (p.344, tradução nossa⁵).

Sweller (1994) propôs que todas as informações possam ser classificadas de acordo com o grau em que seus elementos interagem. Aprender fatos simples está no espectro de baixa interação de elementos, como por exemplo nomes de pessoas ou objetos, aprender o abecedário ou contar de um até dez. No outro extremo, a informação pode ter alta interatividade de elementos, como aprender equações complicadas em matemática, a gramática de uma segunda língua ou qualquer tipo de aprendizagem que possa ser chamada de "aprender com entendimento". Nessa parte do espectro, existe uma exigência de interação de elementos em processos simultâneos, situação essa que pode acarretar alta demanda, ou seja, alta carga cognitiva (SWELLER, 2003).

⁵ a property of instructional content that reflects the extent to which multiple content components must be held and processed simultaneously in WM in order to be learned or to achieve a performance objective

1.2. Categorias de Carga Cognitiva

A carga cognitiva pode ser definida como uma construção multidimensional que representa a carga que a execução de uma determinada tarefa impõe ao sistema cognitivo do aluno (PAAS et al., 2003; PAAS; Van MERRIENBOER, 1994). Serão abordados agora os tipos de carga cognitiva que os procedimentos de instrução impõem dependendo de sua função. A carga cognitiva pode ser classificada em três categorias: carga cognitiva intrínseca, estranha e pertinente. A carga cognitiva total é determinada pela soma total dessas três fontes de carga.

1.2.1. Carga Cognitiva Intrínseca

A carga cognitiva intrínseca está associada ao desafio inerente ou ao nível de dificuldade do material a ser processado. O nível de carga cognitiva intrínseca de um dado conteúdo pode variar entre distintos alunos, devido a pré-requisitos e esquemas adquiridos, e salvo casos particulares, não pode ser modificado através do design instrucional (PAAS; RENKL; SWELLER, 2003).

Um conceito importante que permeia a carga cognitiva intrínseca é o nível de interatividade dos elementos, acredita-se que a elevada interatividade dos elementos provoca alta carga cognitiva intrínseca. Este tipo de carga cognitiva depende do número de elementos do material a ser aprendido que devem ser processados simultaneamente na Memória de Trabalho (SWELLER, 2010).

Quanto mais elementos individuais, dependentes um dos outros, tiverem de ser processados de forma conjunta na Memória de Trabalho durante o processo de aprendizagem relacionado a um determinado material, maior é a carga cognitiva intrínseca imposta, e reciprocamente, quanto menos elementos estiverem presentes na situação, menor a carga cognitiva intrínseca e portanto, esse material deve ser mais facilmente compreendido (CHANDLER; SWELLER 1996).

Essas próximas situações ajudam a elucidar a interrelação de elemento e carga cognitiva intrínseca (SWELLER, 2010). São exemplos de tarefas com baixa interatividade de elemento: o aprendizado de símbolos de elementos químicos ou substantivos de uma língua estrangeira, uma vez que eles são entendidos apenas como elementos individuais que podem ser aprendidos de maneira independente sem interação

com outros elementos externos, portanto, não sobrecarregam a Memória de Trabalho. Por outro lado, a junção de elementos químicos para formar substâncias ou compostos químicos, ou a composição de uma frase em uma língua estrangeira é considerada como uma tarefa de alta interatividade de elemento porque todos os elementos químicos, ou todas as palavras, devem ser consideradas em relação entre si e com as regras de química ou gramática.

Em matemática tomemos o exemplo da resolução da equação $x^2 - y^2 = 9$, onde x e y são números inteiros. Como observado por SWELLER (2010), ao trabalhar com equações, todos os elementos associados a uma equação devem ser considerados simultaneamente porque todos os elementos interagem. Nesse exemplo, existe uma grande quantidade de elementos associados e inter-relacionados: a fatoração da diferença de quadrados, a fatoração do número 9 em números primos, a estruturação da quantidade de soluções e a resolução de sistemas de equações. A interatividade de elementos é muito alta nesse caso.

A compreensão de uma informação ocorre quando é possível processar todos os seus elementos de interação na Memória de Trabalho. A falta de compreensão de uma informação acontece quando não é possível processar os elementos apropriados na Memória de Trabalho. Uma informação é difícil de entender quando existem mais elementos interativos a serem tratados do que podem ser processados na Memória de Trabalho. Finalmente, uma informação de interatividade de elementos baixos é fácil de entender porque pode ser processada de forma fácil e adequada na Memória de Trabalho (SWELLER et al., 1998; SWELLER et al., 2011).

A aprendizagem de um conceito ou procedimento, por exemplo, inclui a transformação de um conjunto de elementos interagentes que são tratados como elementos múltiplos na Memória de Trabalho em um número menor, ou até único. Reduzir a interatividade dos elementos, incorporando elementos interagentes em esquemas, portanto, reduzindo também a carga cognitiva intrínseca, é a principal função do aprendizado. Ao reduzir a carga cognitiva intrínseca os recursos da Memória de Trabalho são liberados para outras atividades. O aprendizado através de aquisição de esquema reduz, ou até elimina, a carga de Memória de Trabalho imposta por informações de interatividade de elementos elevados (SWELLER et al., 2011).

Esquemas possuem um papel importante no aprendizado, sendo geralmente elementos múltiplos interagindo e uma vez que um esquema foi construído, torna-se outro elemento único, com uma carga de Memória de Trabalho leve (SWELLER et al., 2011).

1.2.2. Carga Cognitiva Estranha

A carga cognitiva estranha⁶ é definida como qualquer coisa apresentada em um construto que ocupa capacidade de Memória de Trabalho, mas é irrelevante para o construto pretendido. Uma das maiores tarefas dos designers de instrução é estudar e aplicar métodos de redução de carga cognitiva estranha para liberar espaço na Memória de Trabalho, para efetivamente aprender e resolver problemas associados ao material, ou construto, de relevância. (CLARK et al., 2005; SWELLER 2010).

Uma carga cognitiva intrínseca de um material não é a única responsável pela carga da memória de trabalho. Se o material, ou o construto, possuir instruções ruins ou instruções que transbordam a pretensão do material ou construto, a Memória de Trabalho é afetada, e esses procedimentos que não atingem o seu objetivo são aqueles que exercem uma chamada carga cognitiva estranha (SWELLER et al., 2011).

Se uma determinada área de estudo ou material possui uma alta interatividade entre seus elementos, reduzir a carga cognitiva estranha se torna muito importante. No outro espectro, se um determinado material ou área de estudo possui pouca interação de elementos, portanto carga cognitiva intrínseca baixa, reduzir a carga cognitiva estranha não é crucial uma vez que a Memória de Trabalho não deve transbordar (CLARK et al., 2005). Uma das maiores preocupações dos pesquisadores em TCC é o desenvolvimento e aperfeiçoamento de técnicas e metodologias que permitam reduzir a carga cognitiva estranha (SWELLER, 2010).

Sweller (2010) ainda sugere uma distinção entre a interação dos elementos nos casos das cargas cognitivas intrínseca e estranha: se a interatividade dos elementos pode ser reduzida sem modificar o que o indivíduo aprende, então a carga cognitiva é estranha, mas se a interatividade dos elementos só pode ser alterada modificando o que o indivíduo aprende, a carga cognitiva é intrínseca.

⁶ Em português também encontramos as traduções de Carga Cognitiva Externa, ou ainda irrelevante

Tanto a carga cognitiva intrínseca quanto a estranha são aditivas, e contribuem simultaneamente para a carga cognitiva total imposta por uma tarefa de aprendizagem. Como já explicitado anteriormente, se a carga cognitiva total ultrapassar a capacidade limitada da Memória de Trabalho a aprendizagem será prejudicada, assim como a possibilidade da construção de esquemas e automação (SWELLER et al., 2011). Como a apresentação do material, tarefa ou construto, possui um papel importante na presença ou não da carga cognitiva estranha, então é essa que os designers instrucionais conseguem alterar mais facilmente para que a Memória de Trabalho consiga suportar a carga cognitiva e proporcionar aprendizado. Um pouco mais adiante serão listadas técnicas de redução de carga cognitiva estranha.

1.2.3. Carga Cognitiva Pertinente

Esta categoria de carga cognitiva⁷ está associada a processos que são relevantes para a aprendizagem, está relacionada com a capacidade livre remanescente na Memória de Trabalho para a aquisição e automação de esquemas (SWELLER et al., 2011). Mais ainda, dentre as cargas cognitivas, é aquela que contribui diretamente para o aprendizado do indivíduo através da construção de estruturas cognitivas e processos que visam a melhoria do desempenho do indivíduo (CLARKE et al., 2005). De maneira sintética, é o esforço mental que os alunos dedicam para aprender ou resolver um problema. O aumento da carga pertinente contribui para novos conhecimentos armazenados para o aluno.

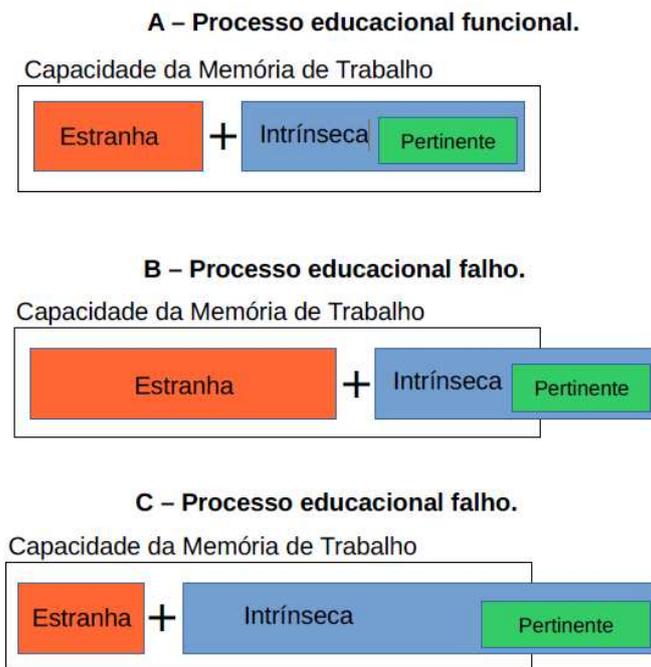
Como um dos objetivos da TCC temos a construção de esquemas e automação, mas para que essas construções sejam adequadas é preciso que a carga cognitiva pertinente atue no sentido de que essas construções de esquemas e automações sejam armazenadas na Memória de Longa Duração (KIRSCHNER, 2002). De modo razoável, a carga cognitiva pertinente aumenta com a motivação dos alunos para participar do processo de aprendizagem.

A carga cognitiva pertinente atua em conjunto com as outras cargas cognitivas. Uma vez que as cargas cognitivas intrínseca e estranha são aditivas, ao aumentar a carga estranha ocorre redução nos recursos na Memória de Trabalho disponíveis para atuar com a carga intrínseca e, portanto, existe uma redução na carga pertinente. Entretanto, a

⁷ Em português também encontramos as traduções de Carga Cognitiva Natural, ou ainda Carga Cognitiva Relevante

diminuição da carga estranha libera recursos para atuar na carga intrínseca, o que conseqüentemente aumenta a carga pertinente. Sob a ótica do design instrucional, é sempre relevante colocar as cargas cognitivas estranha e pertinente na balança, já que a redução da carga cognitiva estranha libera recursos cognitivos para um aumento na carga cognitiva pertinente. (PAAS et al., 2003; PAAS et al., 2010)

Figura 1 - Carga cognitiva e capacidade de Memória de Trabalho



Fonte: RAUBER, 2016

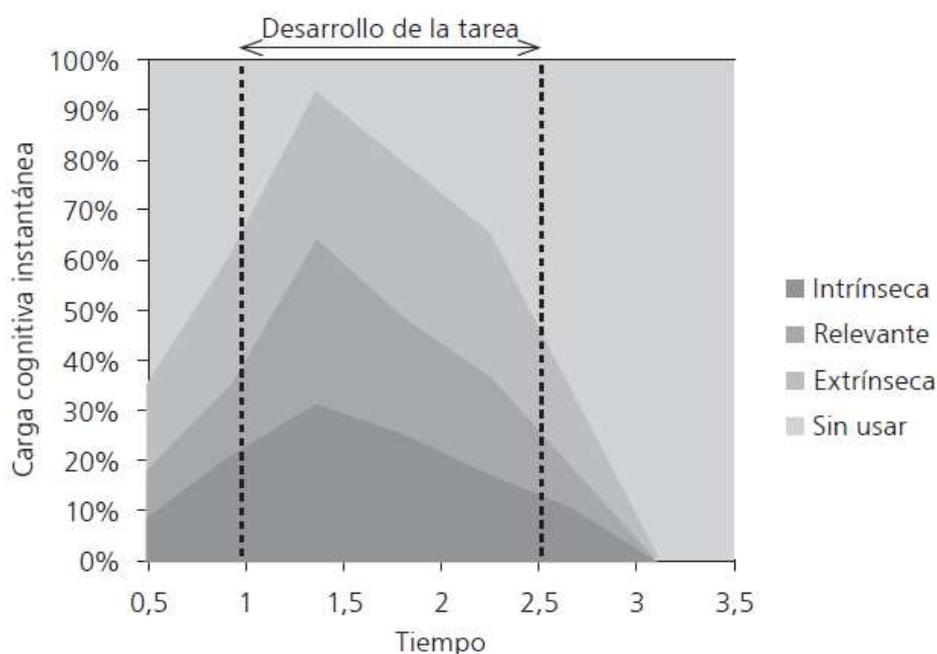
De um modo sintético, o cerne do embate entre as cargas cognitivas e a Memória de Trabalho pode ser compreendido na figura 1. O processo instrucional pode falhar, pelo menos em parte, se a Memória de Trabalho tiver sua capacidade excedida. A tarefa do designer instrucional é encontrar maneiras de adequar os processos às cargas cognitivas em jogo.

A carga cognitiva intrínseca através da interatividade do elemento é determinada por uma interação entre a natureza do material sendo aprendido e a experiência dos aprendizes. Não pode ser diretamente influenciado por designers instrucionais. A carga cognitiva estranha é a carga extra além da carga cognitiva intrínseca resultante da instrução principalmente mal projetada, enquanto a carga cognitiva pertinente é a carga relacionada aos processos que contribuem para a construção e automação dos esquemas. Tanto a carga estranha quanto a pertinente

estão sob o controle direto dos projetistas instrucionais. (PAAS et al, 2003 p. 65, tradução nossa⁸)

Um diagrama de todas essas cargas cognitivas atuando simultaneamente pode ser encontrado em alguns trabalhos. Com base no artigo de Paas et al (2003), Andrade-Lotero (2012) confeccionou um esquema adequado, utilizando apenas os conceitos das cargas cognitivas apresentadas até o momento, para apresentar a interação da Memória de Trabalho sob efeito de uma tarefa.

Figura 2 - Gráfico da Memória de Trabalho de um sujeito sob a resolução de uma tarefa



Fonte: ANDRADE-LOTTERO, 2012

A figura 2 é uma representação do processo da carga cognitiva total que uma tarefa é desenvolvida em um sujeito. A carga cognitiva total é separada em seus três tipos, Intrínseca, Relevante e Estranha, o complementar da carga cognitiva total é a Memória de Trabalho resultante. A tarefa apresentada e a carga inserida no sistema possuem uma duração consideravelmente curta, já que "a Memória de Trabalho permite reter e manipular informação por períodos curtos, de 15 a 30 segundos" (ANDRADE-LOTTERO, 2012, p. 79). Desse modo, picos de carga cognitiva total podem afetar o desenvolvimento

⁸ Intrinsic cognitive load through element interactivity is determined by an interaction between the nature of the material being learned and the expertise of the learners. It cannot be directly influenced by instructional designers. Extraneous cognitive load is the extra load beyond the intrinsic cognitive load resulting from mainly poorly designed instruction, whereas germane cognitive load is the load related to processes that contribute to the construction and automation of schemas. Both extraneous and germane load are under the direct control of instructional designers

de uma tarefa, e a modificação da carga cognitiva para que não ocorra essa sobrecarga, pode trazer benefícios nos processos mentais, incluindo aprendizagem e realização de tarefas.

1.3. Efeitos da Modificação da Carga Cognitiva

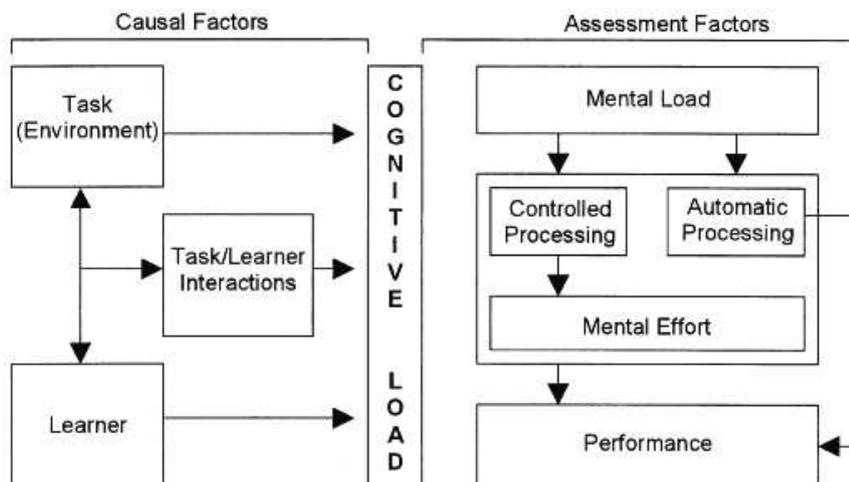
Uma vez apresentadas as categorias de carga cognitiva e suas relações conjuntas e com a Memória de Trabalho, é importante atentar para fatores que contribuem para aumento ou redução de carga cognitiva. Além desses fatores, existem efeitos de instrução que podem contribuir para o aumento ou redução de carga cognitiva.

Quanto aos fatores, eles podem ser causais, ou seja, características ambientais da tarefa, como a estrutura ou temperatura do lugar, características inerentes ao sujeito, como sua capacidade cognitiva, e o meio em que ele interage, inclusive com a própria tarefa, como sua motivação para a realização da tarefa. Esses fatores causais são interdependentes.

Os fatores também podem ser de avaliação, que no caso deste trabalho possui um papel muito importante. Esses fatores de avaliação contêm carga mental, que é imposta pela dimensão da tarefa, e esforço mental, que está na dimensão do sujeito e remete sua capacidade de dedicar recursos cognitivos para a resolução da tarefa. Finalmente, o desempenho, ou performance, é uma manifestação da carga mental, do esforço mental e dos fatores causais discutidos nos últimos parágrafos.

A performance de um indivíduo em uma determinada tarefa (avaliação, por exemplo) depende da carga cognitiva da tarefa e do esforço mental investido pelo indivíduo. Portanto o desempenho é uma manifestação do conjunto formado pela carga mental, esforço mental e dos fatores causais. De um ponto de vista empírico, o desempenho pode ser mensurado em termos das realizações dos indivíduos, como número de itens corretos ou tempo de tarefa, por exemplo. A Figura 3 a seguir mostra em um esquema os fatores, causais e de avaliação, que determinam o nível de carga cognitiva.

Figura 3 - Fatores que determinam o nível de carga cognitiva.



Fonte: PAAS & Van MERRIENBOER, 1994

Como visto, muitos fatores impactam o nível de carga cognitiva. A partir de pesquisas, possíveis modificações da carga cognitiva foram identificadas, induzindo o desenvolvimento de técnicas para a melhoria dos métodos instrucionais (SWELLER, 2008). Algumas dessas técnicas incluem a percepção de distintos efeitos sobre a carga cognitiva: a falta de objetividade; a redundância; a conclusão do problema; a interatividade do elemento; o exemplo trabalhado e a atenção dividida (SWELLER, 2003). Este último mencionado possui um papel importante nesse trabalho e por esse motivo será o enfoque dessa seção.

Efeito sem objetivo (Goal-free Effect) - Fazer uma substituição de problemas convencionais por problemas sem metas, isso propicia ao indivíduo o alcance de um objetivo específico. Ocorre uma redução de carga cognitiva estranha causada por relacionar um estado de problema atual com um estado de objetivo, portanto a meta é tentar reduzir as diferenças entre esses dois estados.

Efeito de Exemplo Trabalhado (Worked Example Effect) - Ao substituir problemas convencionais por exemplos trabalhados, que devem ser cuidadosamente estudados previamente, ocorre redução de carga cognitiva estranha por solução de problemas de método fraco. Os indivíduos focalizam a atenção nos estados problemáticos e etapas úteis da solução.

Efeito de Conclusão do Problema (Completion Problem Effect) - Substituir os problemas convencionais por problemas de conclusão, apresentar uma solução parcial que deve ser completada pelos indivíduos. Reduz a carga cognitiva estranha, porque dar

parte da solução reduz o tamanho do problema; é preciso focar a atenção do aluno em etapas de solução úteis.

Efeito de Modalidade (Modality Effect) - Modalidade aqui está associada a canais de comunicação, nesse caso é útil substituir explicações escritas e outras fontes de informações visuais, como diagramas, por explicações por áudio e visuais (multimodais). Existe uma redução de carga cognitiva pelo fato da Memória de Trabalho processar melhor os canais visual e auditivo simultaneamente.

Efeito de Redundância (Redundancy Effect) - Substituir múltiplas fontes de informação que possuem sobreposição por apenas uma fonte de informação. Ocorre uma redução de carga cognitiva estranha pelo processo desnecessário de informações redundantes.

Efeito de Reversão da Experiência (Expertise Reversal Effect) - Materiais instrucionais que são benéficos aos novatos podem se tornar prejudiciais a especialistas uma vez que, para novatos, um conjunto de elementos que interagem reflete Carga Cognitiva Intrínseca porque eles são essenciais para a compreensão. Com o aumento da experiência, os mesmos elementos de interação refletem Carga Cognitiva Estranha porque eles não são mais necessários.

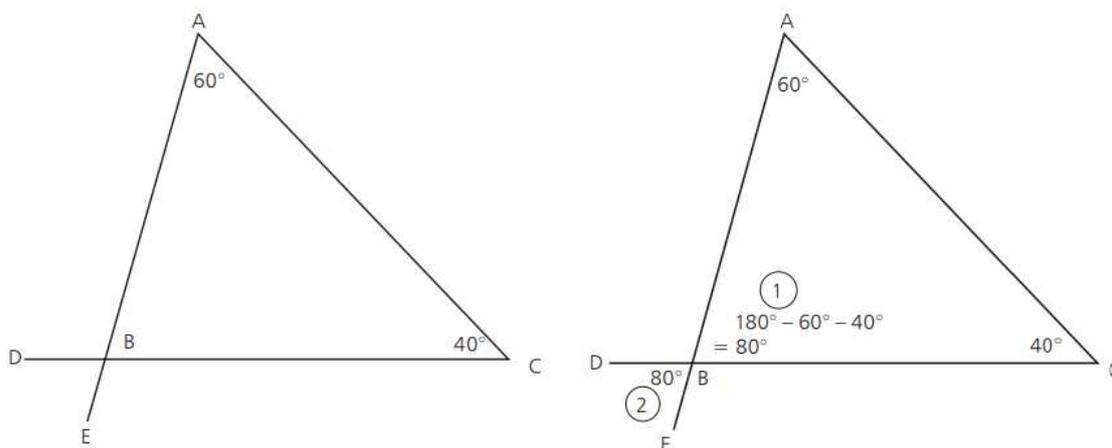
Efeito de Atenção Dividida (Split Attention Effect) - A atenção dividida foi definida por Clarke et al. (2005) como uma "fonte de carga cognitiva estranha causada pela separação de elementos de instrução relacionados que devem ser processados em conjunto para a compreensão" (p.351, tradução nossa⁹). Essa atenção dividida ocorre quando os estudantes se veem obrigados a dividir a atenção entre duas ou mais fontes de informação simultânea que foram separadas, espacial ou temporalmente (SWELLER et al., 2011). Um diagrama qualquer e suas instruções fornecidas são um exemplo de atenção dividida, um outro exemplo seria de um item de múltipla escolha onde a informação necessária para acertar o problema esteja em um gráfico e em um texto. Como é de se esperar, para que o estudante compreenda as instruções do diagrama e das instruções ou acertar o problema que envolve o gráfico e o texto, ele deve integrar mentalmente essas informações (SWELLER et al., 2011).

⁹ source of extraneous cognitive load caused by separation of related instructional elements that must be processed together for understanding

Ou seja, fontes individuais de informação não podem ser compreendidas isoladamente e devem ser integradas na Memória de Trabalho. É partido do pressuposto que cada fonte de informação deve ser essencial no aprendizado e compreensão do conteúdo particular a ser aprendido pelo aluno. No entanto, ao exigir que os estudantes integrem diversas fontes de informação deslocadas e separadas no espaço e no tempo, surge a carga cognitiva estranha. Portanto, para evitar que o aluno vivencie a atenção dividida, essas diferentes fontes de informação devem ser integradas fisicamente (SWELLER et al., 2011).

Existe um conjunto de requisitos comuns a todas as situações nos quais a atenção dividida emerge. A mais importante delas é que as múltiplas fontes de informação não conseguem ser identificáveis isoladamente. Por exemplo, se um diagrama fornece todas as informações necessárias para compreender determinado assunto, integrar qualquer texto adicional no diagrama não será benéfico, no caso de um item em uma avaliação, se em um parágrafo toda a informação relevante para a resolução do problema já é fornecida, proporcionar um ou mais parágrafos sem propósito para a aferição da compreensão do construto pode acarretar em uma taxa menor de acertos daquele item, poluindo o objetivo de avaliar um determinado construto. Sob essas situações, por exemplo, o texto deve ser eliminado.

Figura 4 - Exemplo de atenção dividida



$\angle ABC = 180^\circ - \angle BAC - \angle BCA$ (A soma dos ângulos internos de um triângulo é 180°)

$\angle ABC = 180^\circ - 60^\circ - 40^\circ = 80^\circ$

$\angle DBE = \angle ABC$ (Ângulos OPV são congruentes) = 80°

Fonte: Sweller, 2002, adaptado

Vamos analisar um exemplo de Sweller (2002) sobre a instrução de uma resolução de problema de geometria. Na figura 4, o problema em questão é encontrar o valor do ângulo DBE.

O triângulo da esquerda possui alguns de seus elementos geométricos na própria figura, como a marcação de ângulos nos vértices A e C, a resolução algébrica está logo em seguida. No triângulo da direita, com respeito à resolução, tem-se uma integração melhor da resposta com o problema geométrico, uma vez que é marcado o passo 1, com as contas para se encontrar o ângulo, e o passo 2 com a ênfase em ângulos opostos pelo vértice.

Particularmente em geometria, a integração de resolução da parte geométrica com algébrica, é muito complicada. É necessário observar a figura para se compreender os passos de marcação de ângulo, surge então a atenção dividida. No caso do triângulo da direita, na própria figura a álgebra se justifica apenas pelo contexto geométrico, desse modo qualquer elemento da expressão pode ser compreendido olhando na vizinhança do desenho.

Esse efeito de atenção dividida efetivamente ocorre quando se verifica que estudantes que estudaram com informações integradas tiveram uma performance maior que aqueles estudantes que se prepararam com materiais no formato de atenção dividida. Nesses estudos foram obtidos e categorizados alguns tipos de efeito de atenção dividida: exemplos trabalhados e atenção dividida, diagramas e explicações escritas, múltiplas fontes de texto, mais de duas fontes de informação, e atenção dividida ao aprender a usar um computador (SWELLER et al, 2011). Essa última pode ter sido contundente no desempenho das avaliações do PISA de 2012 em diante, uma vez que as avaliações passaram a ser aplicadas em computadores e as amostras dos alunos contemplavam estudantes de baixa renda com pouco ou nenhum acesso a tecnologia informática no Brasil, e provavelmente muitos países subdesenvolvidos.

Para o nosso trabalho, que envolve o design e aplicação de itens, o efeito da atenção dividida é muito importante, uma vez que nos proporciona a cautela em momentos de apresentar mais de uma informação relevante na resolução de algum problema, e a chave está em apresentar formas de suprimir a separação espacial ou temporal (SWELLER et al, 2011)

O estudo do efeito da atenção dividida indica também por que o efeito do exemplo trabalhado pode falhar, apenas apresentar exemplos trabalhados não é garantida de uma redução de carga cognitiva estranha (SWELLER; CHANDLER, 1991). É possível que ocorra uma propagação de carga cognitiva estranha em um exemplo trabalhado, por esse motivo a estrutura desses exemplos devem garantir que esse efeito adverso não ocorra.

É preciso salientar que o efeito de atenção dividida decorre diretamente da teoria da carga cognitiva, uma vez que, ao ser apresentado por um conteúdo, ocorre simultaneamente uma busca de referenciais e armazenamento de informações na Memória de Trabalho, desse modo surge uma carga cognitiva estranha pesada que deve ser reduzida (SWELLER et al., 2011).

Uma vez que o efeito da atenção dividida é prejudicial, diretrizes foram desenvolvidas por pesquisadores para evitar atenção dividida (CLARKE et al., 2005). A chave dos métodos instrucionais para reduzir a carga cognitiva estranha causada pelo efeito de atenção dividida não está somente em designers procurarem mecanismos para focar atenção, mas também evitar a atenção. Para focar a atenção, são utilizados métodos que direcionam recursos cognitivos para os conteúdos relevantes em materiais. Algumas das diretrizes estipuladas, e que foram utilizadas nesse trabalho são: usar sugestões e sinais para focar a atenção para conteúdo visual e textual importante, e integrar texto explicativo próximo de imagens relacionadas nas páginas.

Um exemplo prático e bem frequente de atenção dividida é a tomada de notas em uma palestra (CLARKE et al., 2005). A carga cognitiva necessária para a tomada de notas reduz a capacidade de processamento mental que poderia ser dedicada ao processamento do conteúdo de maneiras que levem à aprendizagem. Uma sugestão dos autores é que a carga cognitiva pode ser minimizada ao fornecer aos ouvintes resumos prévios de conteúdo, e utilizar o tempo de palestra efetivamente para atividades de aprendizado.

1.4. Considerações sobre a Carga Cognitiva

Ao considerarmos os três tipos de carga cognitiva, intrínseca, pertinente e estranha, e também os efeitos da teoria da carga cognitiva discutidos anteriormente, temos que a carga cognitiva depende: da interação do conteúdo; do objetivo de aprendizagem; do conhecimento prévio dos alunos e do ambiente instrucional.

Dependendo da interatividade dos elementos, a carga cognitiva intrínseca de uma tarefa pode ser alta ou baixa, e a grande questão está em como lidar com o caso da carga cognitiva intrínseca ser alta. O cerne da questão é minimizar a carga cognitiva estranha. Clarke et al. (2005) sugerem que os efeitos da teoria da carga cognitiva conseguem reduzir o impacto da carga cognitiva estranha somente na aprendizagem de tarefas complexas, desse modo, a carga cognitiva estranha não é um fator limitante na aprendizagem de tarefas de baixa complexidade. Em vias de regra, a diretriz geral na busca da eficiência na aprendizagem é "minimizar a carga cognitiva estranha em materiais de instrução quando as tarefas de aprendizagem são complexas" (CLARKE et al., 2005, tradução nossa¹⁰).

Uma vez que os especialistas possuem um repertório maior de habilidades em sua memória que os novatos, e isso lhes permite, comparativamente, realizar tarefas com menos esforço, a orientação geral é "evitar carga cognitiva estranha quando as aulas envolvem conteúdo complexo e os alunos são novatos " (CLARKE et al., 2005, tradução nossa¹¹), como descrito no parágrafo anterior. E como visto nos efeitos de carga cognitiva, outra tarefa dos designers é mudar as estratégias de instrução à medida que os alunos adquirem e desenvolvem expertise durante os cursos.

¹⁰ minimize extraneous cognitive load in instructional materials when learning tasks are complex

¹¹ avoid extraneous cognitive load when lessons involve complex content and the learners are novices

2. Avaliação

A avaliação educacional é uma parte muito importante de qualquer sistema e programa na área de Educação, ela permeia todo o espectro escolar, do básico ao superior. Está presente ativamente na vida do jovem aluno do Ensino Fundamental até postulantes a título de Doutor, passando necessariamente pelos professores, de todas as disciplinas e níveis, e também em esferas governamentais, como secretarias de educação e políticas públicas.

Segundo a Associação Americana de Pesquisas Educacionais (AERA) e também a Associação Americana de Psicólogos (APA) em seu guia padrão de testes educacionais e psicológicos¹², avaliação é definida como "um método sistemático de obter informações a partir de testes e outras fontes, utilizado para extrair inferências sobre características de pessoas, objetos ou programas"¹³ (AERA et al, 1999). A importância da avaliação está em obter um retorno do conjunto de objetivos e proficiências pré-estabelecidos pelos agentes responsáveis. Muito embora existam muitos instrumentos de avaliação disponíveis, cada um com seus méritos e deficiências, o propósito desse capítulo está em apresentar elementos e discussões das avaliações, exames de larga escala e o efeito do fator sexo no desempenho nessas avaliações.

Primeiramente, é importante distinguir duas dimensões para a avaliação educacional. A avaliação interna da aprendizagem é aquela realizada por professores para suas turmas, geralmente é uma prova composta por questões dissertativas, e cujo propósito é verificar o desenvolvimento de seus alunos referentes a um certo conjunto de proficiências e habilidades, para que o professor possa tomar alguma ação referente a readequação de conceitos ou metodologias, visando o bom desenvolvimento de seus alunos.

A avaliação externa, também chamada de avaliação em larga escala, é uma avaliação padronizada confeccionada por agentes externos ao ambiente de sala de aula, conduzida em uma escala regional, nacional ou internacional, envolvendo uma grande quantidade de pessoas. Dentro do escopo de cada uma dessas duas dimensões para a

¹² The Standards for Educational and Psychological Testing, tradução nossa

¹³ any systematic method of obtaining information from tests and other sources, used to draw inferences about characteristics of people, objects, or programs, tradução nossa

avaliação, informações preciosas podem ser obtidas. Para Penin (2009), ambas são importantes.

[...] no âmbito interno, possibilita a avaliação como instrumento de ação formativa, levando instituições e os professores a refletirem a respeito de suas práticas e de seus objetivos e, assim, a melhorar sua ação docente e sua identidade profissional. Por outro, em âmbito externo, oferece informações para que tanto os pais quanto a sociedade, especialmente os sistemas de ensino, possam efetivar um relacionamento produtivo com a instituição escolar. Apurar os usos da avaliação, comparar resultados e comportamento de entrada dos alunos em cada situação e contexto social e institucional é da maior importância para não homogeneizar processos que são de fato diferentes.

Nessa perspectiva, é crucial que os colégios façam o uso das avaliações interna e externa. Concordamos com Depresbiteris (2001) que, se um colégio usufrui das duas dimensões avaliativas, pode-se criar uma cultura avaliativa que permitiria chegar em objetivos básicos da avaliação já mencionados por Penin, a saber: o aperfeiçoamento contínuo da missão do sistema, que é educar, ser um instrumento de gestão educativa, que é crucial para políticas públicas, e proporcionar uma devolutiva a todos os atores escolares, particularmente professores, pais e estudantes.

Colocado dessa maneira, dado que ambas dimensões são importantes, é necessário buscar um equilíbrio entre elas, como aponta Depresbiteris (2001). Uma avaliação exclusivamente interna isolaria o colégio de modo a comparar e criticar sistemas de ensino, enquanto que uma avaliação puramente externa pode exacerbar competições interescolares, colocando a concorrência e a classificação dos colégios acima do propósito máximo da escola e seus fins educacionais.

Segundo Depresbiteris ainda, uma avaliação deve ter três propósitos centrais, de modo sintético: "fornecer resultados para a gestão da educação, subsidiar a melhoria de projetos pedagógicos das escolas e propiciar informações para a melhoria da própria avaliação".

Como já era de se esperar, a avaliação está diretamente ligada a muitos outros fatores e variáveis, é um processo com muitos atores. Muitos autores reconhecem a complexidade do assunto nesse sentido, para Depresbiteris (2001) essa questão não é simples "pelo papel que desempenha no julgamento de valor dos programas e sistemas, a

avaliação tem grande poder e, por isso, a importância de uma reflexão aprofundada sobre os diferentes aspectos que a permeiam".

Historicamente, alguns fatores ajudaram a moldar a avaliação em larga escala, vale mencionar a psicologia, as medidas do comportamento humano, processos sensoriais e habilidades mentais, comparação educação e medidas mentais, e a própria avaliação educacional interna.

2.1. Desempenho, Avaliação em Larga Escala e Sexo - resumo

As avaliações, internas ou externas, trabalham como uma interface entre o conhecimento e o indivíduo. Em matemática, poderíamos pensar como se a avaliação fosse uma função de várias variáveis, onde se colocam os dados relativos ao estudante: o quanto este estudou, se dormiu bem no dia anterior, etc; e o instrumento de avaliação em si: os tipos de itens, a seleção dos temas, etc. Desse modo, ao embutir os dados de entrada, teríamos valores numéricos esperados, relacionando indivíduos com uma métrica do que foi avaliado.

No final do processo avaliativo, o estudante está interessado nesse número, mas como sabemos, a quantidade de variáveis que afetam esse número pode ser muito grande, e algumas variáveis podem se comportar de maneira não esperada. Esse capítulo tem grande importância nesse trabalho, uma vez que a variável sexo não nos mostrou ter um comportamento previsível.

Os exames em larga escala conseguem gerar muitos dados, sendo que alguns deles apresentam bancos de dados abertos. Desse modo, pesquisadores conseguiram fazer estudos e buscar alguns padrões e inferências relativas à avaliações de matemática ou leitura, por exemplo. No que tange à variável sexo, o desempenho masculino é maior em matemática, e o desempenho feminino é maior em leitura. Algumas avaliações em larga escala trabalham com itens de matemática permeados com leitura e interpretação, misturando esses dois conceitos.

A seguir serão feitas algumas considerações sobre avaliações em larga escala, e algumas dessas avaliações que foram relevantes em nosso estudo, para então discutir as pesquisas relativas ao desempenho entre os sexos masculino e feminino.

2.2. Avaliações em Larga Escala - panorama

Existem diferentes redes de ensino em nosso país: municipal, estadual e nacional. Não é possível imaginar uma uniformidade das redes em relação a conteúdos e ensino, mesmo entre municípios, portanto, era de se esperar que em algum momento as avaliações em larga escala no Brasil aparecessem com suas preocupações em relação ao processo avaliativo escolar.

Segundo Gatti (2009), alguns profissionais receberam formação, alguns no exterior, na área de avaliação em rendimento escolar, e em 1966 no Rio de Janeiro, a Fundação Getulio Vargas criou um centro de testes educacionais. Este centro criou uma avaliação composta por questões objetivas, em diversas áreas do conhecimento para alunos do final do Ensino Médio. Neste processo também se incluía um questionário, com itens socioeconômicos e de aspirações. A autora considera esta como a primeira ação nacional de grande porte para o estudo e verificação de diversas variáveis no processo avaliativo escolar. Na mesma época a Fundação Carlos Chagas também começou a formar especialistas em testes e medidas, mas o enfoque foi em processos seletivos para universidades, cursos superiores e cargos públicos.

Ainda segundo a autora, as administrações públicas não se mostravam muito interessadas em avaliar sistematicamente o rendimento escolar dos alunos, dado que após a primeira avaliação de 1966, poucas ações foram tomadas. A situação toma um rumo "a partir de 1988, com alguns estudos exploratórios, ocorrendo a implementação de um sistema nacional de avaliação da educação básica apenas em 1990." (GATTI, 2009, p.9). Nessa época era intenso o debate sobre a repetência e a evasão escolar na escola básica, uma grande questão levantada era a falta de dados sobre rendimento escolar. Por este motivo "ao final de 1987, foi proposto que se fizesse uma avaliação de rendimento escolar em 10 capitais de estados do país, para se aquilatar se um processo de avaliação mais amplo por parte do Ministério seria viável e traria resultados relevantes." (GATTI, 2009, p.11).

Foi um estudo piloto para algumas séries do ensino fundamental, com provas de português (com redação), matemática e ciências. Para a análise foi utilizada a teoria clássica dos testes, e o grande desafio com relação a validade do teste foi abarcar a heterogeneidade cultural do nosso país, uma vez que não existia um currículo nacional integrado. Após este estudo piloto bem sucedido, o estudo avaliativo se estendeu para 20

capitais e, depois, mais algumas outras cidades. Os resultados, segundo relata a autora, não foram bons, provocando ressonância no Ministério e nas secretarias de educação, bem como na mídia. Esta devolutiva foi relevante para professores, alunos e sociedade no geral, para que fosse possível uma reflexão das práticas e contextos educacionais. "Estes primeiros estudos mais abrangentes de alunos e escolas serviram de base para a implantação do SAEB — Sistema de Avaliação da Educação Brasileira." (GATTI, 2009, p.11).

Segundo a página da internet do INEP:

O SAEB, instituído em 1990, é composto por um conjunto de avaliações externas em larga escala e tem como principal objetivo realizar um diagnóstico da educação básica brasileira e de alguns fatores que possam interferir no desempenho do estudante, fornecendo um indicativo sobre a qualidade do ensino ofertado. O levantamento produz informações que subsidiam a formulação, reformulação e o monitoramento das políticas públicas nas esferas municipal, estadual e federal, visando a contribuir para a melhoria da qualidade, equidade e eficiência do ensino. Além disso, procura também oferecer dados e indicadores sobre fatores de influência do desempenho dos alunos nas áreas e anos avaliados.

Em uma reestruturação metodológica do SAEB, em 1995, foram incorporadas à avaliação as técnicas da teoria de resposta ao item (TRI), uma ferramenta estatística que permite a comparação do desempenho dos estudantes (comparabilidade em escala) ao longo dos anos (RABELO, 2013). Ainda segundo o autor, essa experiência adquirida com o uso da TRI no SAEB permitiu que essa metodologia se difundisse para outros processos de avaliação no país, incluindo o Enem.

O SAEB foi novamente reestruturado em 2005 e passou a ser composto por duas avaliações: 1) a Avaliação Nacional da Educação Básica (Aneb), semelhante ao que havia sido realizado até o momento; 2) Avaliação Nacional do Rendimento Escolar (Anresc), conhecida como Prova Brasil, a qual tem como objetivo principal avaliar a educação da rede pública brasileira. Passando por outra mudança, em 2013 acrescenta-se outra avaliação em larga escala: a Avaliação Nacional da Alfabetização (ANA), que busca avaliar a alfabetização e o letramento em Língua Portuguesa e Matemática. Uma linha do tempo com as devidas modificações e observações do SAEB está disponível no **Anexo I**.

No âmbito global, a primeira movimentação para se estabelecer uma avaliação em larga escala ocorreu no encontro de 12 países para a fundação da Associação

Internacional para a Avaliação da Conquista Educacional¹⁴ (IAE). A primeira avaliação internacional foi de matemática e ocorreu em 1964 (OECD, 2015).

2.2.1. Enem - panorama

O Enem teve início em 1998 para ser aplicado aos alunos egressos e concluintes do ensino médio. A avaliação era aplicada anualmente com o objetivo fundamental de "avaliar o desempenho do aluno ao término da escolaridade básica, para aferir o desenvolvimento de competências fundamentais ao exercício pleno da cidadania" (BRASIL, 1998). Mais ainda:

As tendências internacionais, tanto em realidades mais próximas da nossa como nas mais distantes, acentuam a importância da formação geral na educação básica, não só para a continuidade da vida acadêmica como, também, para uma atuação autônoma do sujeito na vida social, com destaque à sua inserção no mercado de trabalho, que se torna mais e mais competitivo. Esta formação deve ser compreendida como uma sólida aquisição dos conteúdos tradicionais das ciências e das artes associada ao desenvolvimento de estruturas capazes de operacionalizá-los no enfrentamento de problemas apresentados pela realidade social, cada vez mais complexa, e numa dinâmica de tempo progressivamente acelerada. (INEP, 1999, p. 5)

Desde o surgimento desta avaliação, até os dias de hoje várias são as críticas direcionadas à aplicação desta avaliação. Muitos entendem que não é possível aplicar uma avaliação comum ao país todo, de modo a avaliar adequadamente as diferenças regionais presentes em um país tão grande. O fato das provas do Enem (entre 1988 e 2008) não utilizarem TRI faz ainda com que as provas de anos distintos não fossem comparáveis entre si, por não utilizarem uma escala de proficiência.

O Enem em sua concepção inicial ainda possuía objetivos específicos que envolviam a situação presente e futura dos candidatos, já tendo em vista possíveis utilizações do exame (BRASIL, 1998)

a) oferecer uma referência para que cada cidadão possa proceder a sua autoavaliação com vistas às suas escolhas futuras, tanto em relação ao mercado de trabalho quanto em relação à continuidade de estudos; b) estruturar uma avaliação da educação básica que sirva como modalidade alternativa ou complementar aos processos de seleção nos diferentes setores do mundo do trabalho; c) estruturar uma avaliação da educação básica que sirva como modalidade alternativa ou

¹⁴ International Association for the Evaluation of Educational Achievement

complementar aos exames de acesso aos cursos profissionalizantes pós-médios e ao ensino superior.

Desde sua origem, o exame tem como base de referência a LDB, os PCNs, a reforma do Ensino Médio e as matrizes de referência do SAEB. O Enem até 2008 era composto de uma única prova, com 63 itens de múltipla escolha e uma redação, tendo como base uma matriz de competências e habilidades desenvolvida por um grupo de pessoas que fazem parte da área da educação. A aplicação desta prova acontecia em um único dia. No relatório final do Exame Nacional do Ensino médio de 1998 (BRASIL, 1999) é possível encontrar, além da equipe técnica e de educadores responsáveis, as motivações da implementação do Enem. No **Anexo II** está disponível a Matriz de Referência, com as competências e habilidades, além da matriz da Análise de Desempenho.

A matriz entende que o domínio de ler, compreender e interpretar textos é uma competência que deve perpassar todas as disciplinas, não devendo estar presente somente em Língua Portuguesa. Para isso, os itens seguem padrões diferentes do que normalmente são trabalhados no ensino médio. Este é um dos pontos que causa mais estranhamento para quem começa a prestar o exame. Os itens são produzidos de modo a exigir leitura e interpretação, demandando mais atenção e conhecimento por parte dos candidatos.

O desempenho na prova foi separado em duas partes, cada uma valendo 100 pontos. Uma nota de desempenho geral representou uma das partes, que envolvia os 63 itens de múltipla escolha de conhecimentos gerais. A partir de cada uma das cinco competências e pelas relações associadas com as respectivas habilidades, conforme o modelo de Análise de Desempenho, gerando para cada competência uma nota de 0 a 100.

A classificação final ficou estabelecida em faixas de intervalos, sendo que ao atingir de 0 a 40% de acertos o aluno será considerado com desempenho insuficiente a regular; de 41 a 70% de regular a bom e de 71 a 100% de bom a excelente.

No ano 2005 (INEP, 2005) foi elaborado um documento intitulado *Exame Nacional do Ensino Médio (Enem): fundamentação teórico-metodológica*, que explicita as novas orientações e regras deste exame a partir daquela data. O primeiro capítulo deste trabalho tem como autor Lino de Macedo, um dos consagrados escritores e estudiosos da área da educação. O foco desta primeira parte é apresentar os eixos teóricos que fundamentam a criação e elaboração do Enem. Alguns temas são focos de apresentação

e discussão no início do capítulo mencionado: competências e habilidades; competição, competência e concorrência; autonomia; aprendizagem significativa; escola para todos ou escola de excelência; cooperação e competência relacional. O que há nesse trabalho são reflexões sobre os termos citados, de modo que o leitor entenda que estas são as bases para a criação e manutenção desta avaliação. Em seguida, há uma explicação teórica sobre a situação problema como forma de aprendizagem, sendo acompanhada por uma relação desta com as provas do Enem, de modo que haja uma reflexão sobre este tipo de atividade feita pelo professor.

Na sequência, Nilson José Machado discute sobre a interdisciplinaridade e a contextualização, apresentando teorias da psicologia e da filosofia. Também são elementos que são inseridos como forma de convencer o leitor que há um grande embasamento teórico como apoio ao Exame.

Somente no segundo capítulo deste documento é possível encontrar uma relação mais direta com o Enem e documentos/realidade que se referem ao ensino médio. São retomadas e reforçadas algumas ideias presentes no documento de 1999, já citado anteriormente: importância da presença da compreensão e interpretação da leitura em todas as disciplinas e a as competências necessárias para este eixo de linguagem e códigos. Também é reforçada neste texto a relação entre as três áreas do conhecimento (Linguagem e Código e suas tecnologias, Ciências da Natureza, Matemática e suas tecnologias e Ciências Humanas e suas tecnologias) em que é dividido o Enem, assim como as convergências de concepções de matriz e objetivos presentes no Exame e nos PCNs. Como forma de comprovação das relações entre as áreas e a avaliação, destacam-se os objetivos da primeira e terceira competências do exame:

Dominar a norma culta da Língua Portuguesa e fazer uso das linguagens matemática, artística e científica.

Relacionar informações, representadas em diferentes formas, e conhecimentos disponíveis em situações concretas, para construir argumentação consistente. (INEP, 2005, p. 63)

A área do conhecimento de Ciências Humanas e suas Tecnologias se relaciona com as mudanças sociais, econômicas e culturais do Brasil. Por se tratar de assuntos atuais e envolver também os jovens estudantes do ensino médio, os PCNs ressaltam a importância deste estudo, também presente na avaliação do Enem:

[...] as Ciências Humanas têm um importante papel na compreensão do significado das tecnologias para as sociedades. Apontam tanto os

processos sociais que levam os homens a buscarem respostas e ferramentas para a resolução de problemas concretos, quanto avaliam o impacto que as tecnologias promovem sobre essas mesmas sociedades (INEP, 1999, p. 34).

A formação do cidadão crítico deve ser uma das preocupações da educação, sendo assim, esta área precisa estar voltada para questões sociais, inserindo os jovens nas discussões sobre os contextos em que estes estão inseridos. O Exame tem como proposta avaliar habilidades de compreensão da realidade social e de posicionamento crítico, conforme o documento:

19. Confrontar interpretações diversas de situações ou fatos de natureza histórico-geográfica, técnico-científica, artístico-cultural ou do cotidiano, comparando diferentes pontos de vista, identificando os pressupostos de cada interpretação e analisando a validade dos argumentos utilizados
20. Comparar processos de formação socioeconômica, relacionando-os com seu contexto histórico e geográfico;
21. Dado um conjunto de informações sobre uma realidade histórico-geográfica, contextualizar e ordenar os eventos registrados, compreendendo a importância dos fatores sociais, econômicos, políticos ou culturais (INEP, 1999, p. 9).

Desde sua criação, em 1998, o número de inscritos no Enem cresceu consideravelmente devido a diversos fatores. No ano em que foi lançado, o Enem contou com 157 mil inscritos e já passou para 346 mil inscritos em 1999, quando 96 IES utilizaram de algum modo a nota do Enem na composição da nota final de alguns vestibulares tradicionais. Em 2001, com a gratuidade da taxa de inscrição, o número de candidatos saltou para 1,6 milhões, e após a criação do Prouni, Programa Universidade para Todos, em 2004, o número de inscritos saltou para 3 milhões em 2005.

Para entender os rumos do Enem antes da sua reestruturação, é importante entender o público que começou a aderir essa avaliação ao longo dos anos, e quais foram as finalidades desses candidatos. Como bem apontado por Corti (2013), ocorreu uma heterogenização do público após a criação do Prouni, uma vez que a disputa de bolsas em IES privadas provocou o aumento de candidatos mais velhos, interessados nessa oportunidade. Esse caráter cada vez mais forte associado ao ingresso no ensino superior transformou o Enem em algo distante da proposta inicial, uma vez que o público majoritariamente não ficou mais caracterizado como concluintes do ensino médio.

O Enem sofre em 2009 uma mudança estrutural profunda, é publicada a portaria n. 109 de 27 de maio daquele mesmo ano (BRASIL, 2009). Nesse documento estão presentes o almejo do governo avaliar a situação do ensino médio brasileiro e aferir devolutivas educacionais para mudanças que visassem melhorias efetivas na estrutura do ensino, seja no macro, em secretarias e sistemas de educação, quanto no micro, em salas de aula, entre professores e alunos, bem como competências e habilidades também relacionados ao exercício da cidadania.

Desde 2009 o exame é composto por 180 itens de múltipla escolha mais uma redação, são 45 itens para cada grande área: Ciências Humanas e suas Tecnologias, Ciências da Natureza e suas Tecnologias, Matemática e suas Tecnologias e Linguagens, Códigos e suas Tecnologias. Os itens em cada uma das provas utilizam TRI, a régua foi calibrada em 500 pontos na prova de 2009 como a média de estudantes da escola pública, em cada prova.

Como já mencionado anteriormente, muito embora o exame seja do ensino médio, vale destacar duas características que contribuíram para inflar o número de candidatos. A possibilidade de conseguir uma certificação do ensino médio, obtendo pelo menos 450 pontos na parte de conhecimento geral e 500 na redação, e o caráter de ingresso em IES, que se consolidou com a criação do Sisu, Sistema de Seleção Unificada. Naquele mesmo ano 51 instituições utilizaram exclusivamente o Enem como forma de ingresso. Em 2010 o Enem teve mais de 4,6 milhões de alunos inscritos, nesse ano foi criado o Fies, Financiamento ao Estudante do Ensino Superior, para candidatos de baixa renda e pontuações mínimas estipuladas, e também mais de 600 IES passaram a utilizar o Enem com esse mesmo propósito.

No ano de 2015, o Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), publicou um Relatório pedagógico, em que são estudados os números obtidos após a aplicação da prova, seguido por algumas discussões sobre a proposta e estrutura do exame. Além de apresentações sobre as orientações e estrutura do Enem, já inseridas em documentos anteriores, algumas alterações aparecem neste momento e são explicitadas no documento. Uma das mudanças diz respeito à matriz e a distribuição dos itens.

O que foi apresentado até o momento foram dados específicos da estrutura de produção e aplicação do Enem. Quando se pensa na recepção da sociedade atual podemos perceber que ainda há diversas críticas sobre esta avaliação. Um aspecto que é

amplamente criticado são os tipos de itens que são elaborados, já que a maioria das escolas brasileiras não conseguem aprofundar o processo de aprendizagem no mesmo nível de cobrança que aparece na prova. O grande problema do Enem e do ensino de maneira geral está associado à interpretação de textos. Muitos não conseguem entender o que está sendo solicitado no enunciado e por isso não conseguem responder de modo adequado.

2.2.2. Pisa - panorama

O Programme for International Student Assessment (PISA), que pode ser traduzido para o português como Programa Internacional de Avaliação de Alunos, é uma avaliação internacional em larga escala que foi proposta e lançada em 1997 pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que é uma organização atualmente composta por 34 países com o propósito de efetuarem um "intercâmbio de informações e alinharem políticas, com o objetivo de potencializar seu crescimento econômico e colaborar com o desenvolvimento de todos os demais países membros".

Sobre a avaliação do PISA, suas especificidades e metodologias, é importante citar que (OCDE 2016; OCDE 2013):

- É uma avaliação internacionalmente padronizada, desenvolvida em conjunto pelos países participantes e administrada para alunos com idade entre 15 anos e 2 meses e 16 anos e 3 meses no momento da avaliação, matriculados em alguma instituição educacional a partir do sétimo ano do ensino fundamental.
- Cobre conteúdos de matemática, leitura e ciências e é aplicada a cada 3 anos (ciclo) onde em cada ciclo uma das disciplinas recebe um foco maior.
- É aplicada aos países membros da OCDE e mais países/economias parceiras. Em 2012 foram 65 países/economias participantes e em 2015 foram 72 países/economias.
- O teste é administrado para uma amostra entre 4500 e 10000 estudantes em cada país/economia em pelo menos 150 colégios de cada um desses países/economias. Em 2015 participaram 23141 estudantes de 841 escolas na amostra do Brasil, com o propósito de produzir resultados mais confiáveis por unidade da Federação.

- São testes de mesa, papel e caneta num total de 2 horas de prova para cada aluno. Em alguns países/economias, a partir de 2012 foram aplicados itens adicionais de matemática e leitura via computador em um teste adicional de 40 minutos.
- Em 2015, no Brasil, os testes e questionários contextuais do PISA foram aplicados integralmente por computador.
- Os itens são uma mistura de questões de múltipla escolha e questões dissertativas.
- Os alunos possuem 30 minutos para preencher questionários socioeconômicos.

Com o surgimento de diversas avaliações nacionais em larga escala, era de se esperar que em algum momento fosse pensado e concretizado um processo de avaliação entre países, e um dos propósitos do PISA era exatamente este. Muito embora pareça natural querer comparar desempenhos entre países, esta proposta esbarra nas políticas educacionais e currículos particulares de cada país, portanto um ponto chave de uma avaliação internacional é contornar essa questão da localidade dos currículos.

O PISA, enquanto tenta acessar um conhecimento baseado em um componente curricular comum entre os países, também "examina a habilidade dos alunos em refletir e aplicar o seu conhecimento e experiências em problemas do cotidiano de maneira reflexiva" (OCDE 2013, p.13). Essa ideia de trabalhar o conceito de conhecimento não como a posse de uma informação, mas também como a avaliação e aplicação dela, está associado ao termo letramento (literacy), que é

A conjunção de um conceito mais amplo de conhecimentos e habilidades, e a avaliação do PISA tem como objetivo determinar a extensão no qual alunos de 15 anos podem ativar diversos processos cognitivos que lhes permitam fazer o uso efetivo de leitura, matemática e conhecimento científico e habilidades que eles adquiriram através de sua escolaridade e experiências de aprendizagem relacionadas até aquele ponto.

Vale mencionar uma avaliação em larga escala internacional que é um contraponto ao PISA: o Trends in International Mathematics and Science Study (TIMSS), Tendências em Estudos Internacionais de Matemática e Ciência, mas que diferentemente do PISA, possui um enfoque no alcance do currículo, ou mais precisamente, no que o aluno aprendeu em sala de aula. Em suma, o TIMSS é uma avaliação mais baseada no currículo enquanto o PISA está mais preocupado com a compreensão e aplicação de ideias,

que é uma característica do Enem, e por este motivo vamos abordar as facetas da prova de matemática do PISA.

2.3. Desempenho e Sexo em Avaliações em Larga Escala

Avaliações internacionais bem consolidadas como o PISA tiveram grande influência nas avaliações em larga escala nacionais. Pesquisas qualitativas e quantitativas sobre o desempenho entre homens e mulheres em avaliações de matemática acontecem já há um bom tempo. Anne Anastasi, uma pioneira em testes psicométricos, descobriu que muito embora os rapazes se sobressaíssem em conhecimentos numéricos, em relação às meninas, essa diferença de desempenho não aparecia até o final do ensino fundamental. Mais ainda, a pesquisa afirmou que se as diferenças de desempenho entre meninos e meninas aparecessem em computação de contas, elas favorecem as mulheres, enquanto os homens se destacavam nos testes de raciocínio (ANASTASI, 1956).

Os testes sobre desempenho entre homens e mulheres inicialmente foram aplicados em contextos locais e em populações reduzidas. Análises de desempenho em matemática entre homens e mulheres em avaliações em larga escala começaram a emergir nos Estados Unidos na década de 70 e 80, através de testes padronizados amplamente usados, como por exemplo o American College Testing Program (ACT), Graduate Management Admission Test (GMAT) e o mais famoso de todos, o Scholastic Aptitude Test (SAT).

A idade dos homens e mulheres que foram submetidos a esses exames, por serem avaliações de acesso ao ensino superior, foi em média de 18 anos, e o número de candidatos avaliados era considerável. Para se ter uma ideia, o ACT teve em 1970 aproximadamente 12 mil candidatos, e em 1987 aproximadamente 356 mil.

Uma das primeiras grandes pesquisas de meta análise com dados de avaliações americanas em larga escala em nível nacional, estadual ou local, foi realizada por Hyde, Fennema e Lamon (1990) com um banco de dados de 1967 a 1987. O espectro de idade média dos rapazes e das meninas em alguns desses exames foi bem variado, de 6 a 18 anos. Os exames de fim de começo ou fim de ciclo explicam essa grande faixa de idades testadas.

Os resultados dessa pesquisa conduzida por Hyde, Fennema e Lamon corroboram as conclusões de Anastasi, que foram administrados em avaliações com menos respondentes. O desempenho das meninas é essencialmente igual ao dos meninos em matemática até o final do ensino fundamental, com uma leve vantagem feminina em cálculos de operações. No final do ensino médio o desempenho dos rapazes é estatisticamente superior ao das meninas (HYDE; FENEMMA; LAMON, 1990, p.148).

Muito embora essa pesquisa tenha ocorrido em um país grande, com muitas diferenças locais e ainda por cima envolvesse a diferença de desempenho por etnicidade, seria preciso uma avaliação internacional, entre diversos países, para gerar uma visão mais ampla da situação entre homens e mulheres em avaliações de matemática. Com a primeira edição do Pisa em 2000, e o Timss em 2005, essas pesquisas começaram a aparecer em relatórios gerais, e em meta análises posteriormente. As análises, tanto sobre as avaliações americanas quanto às primeiras avaliações em larga escala internacionais, somaram mais elementos na discussão sobre possíveis causas da diferença de desempenho entre homens e mulheres em provas de matemática.

Não é consenso se a diferença, ou similaridade, de desempenho em Matemática entre os sexos masculino e feminino ocorre por motivos sócio culturais. As autoras Else-Quest, Hyde e Linn (2010, p.106, tradução nossa¹⁵) compilaram em um artigo teorias sobre esse assunto, entre eles a hipótese de estratificação de gêneros:

De maneira ampla, a hipótese de estratificação de gêneros propõe que, onde existe mais estratificação social baseada em gênero, e como consequência mais desigualdade de oportunidades, meninas vão ter atitudes menos positivas e uma resposta afetiva mais negativa, e terão uma performance acadêmica pior em exames de matemática de que seus colegas do sexo masculino. Mais ainda, onde há maior igualdade de gêneros, semelhanças entre gêneros em matemática ficarão evidentes.

É preciso ressaltar que a palavra gênero é utilizada pelos autores que estudam desempenhos sobre homens e mulheres se referindo exatamente ao sexo masculino e sexo feminino. A palavra gênero atualmente está mais intimamente ligada ao estudo de gênero, que possui outras consequências e não é nosso escopo nesse trabalho. Por motivos

¹⁵ Broadly, the gender stratification hypothesis proposes that, where there is more societal stratification based on gender, and thus more inequality of opportunity, girls will report less positive attitudes and more negative affect and will perform less well on mathematics achievement tests than will their male peers. Yet, where there is greater gender equity, gender similarities in math will be evident.

históricos, qualquer referência a gênero nesse trabalho deve ser entendida como masculino e feminino, uma vez que as avaliações em larga escala permitem apenas essas duas opções nos questionários socioeconômicos.

Três abordagens dentro da psicologia tentam jogar uma luz nessa hipótese. Duas delas estão no trabalho de Else-Quest, Hyde e Lynn (2010, p.106, tradução nossa) e dizem que “pessoas não se engajam em um desafio, a menos que elas o valorizem e tenham alguma expectativa de sucesso¹⁶”, e “forças socioculturais, como pais e professores, atitudes e expectativas, incluindo estereótipos, também moldam o autoconceito e atitudes frente ao assunto¹⁷”, ou seja, em sociedades em que as meninas não observam mulheres em postos de Ciências Exatas, elas se sentem menos compelidas a aceitar o desafio de se engajar nessas carreiras. Por último existe a teoria da estrutura social, ou do papel social, que diz que a diferença entre os gêneros está enraizada em fatores socioculturais, como por exemplo divisão de trabalho baseada no gênero.

Uma teoria comentada nesse mesmo artigo, mas da autora Hyde, é a hipótese da similaridade entre gêneros (HYDE, 2005, p. 581, tradução nossa¹⁸) que essencialmente diz “que homens e mulheres são similares na maioria, mas não em todas as variáveis psicológicas”. Esse artigo baseia-se no acompanhamento de dois exames aplicados no mundo em vários países, os já mencionados PISA e o TIMSS. Os resultados mostram que em alguns países não há diferença significativa de desempenho entre os gêneros nesses exames.

Foram criadas algumas medidas de igualdade entre gêneros, entre 1990 e 2000 para que a comparação entre países fosse possível. Nas palavras de Kane e Merz (2012, p.19, tradução nossa¹⁹) “igualdade de gênero e outros fatores sócio-culturais, e não a receita dos países, tipo de escola e religião, são os fatores prioritários da performance em matemática em todos os níveis, tanto para meninos quanto para meninas”.

¹⁶ people do not undertake a challenge unless they value it and have some expectation of success

¹⁷ Sociocultural forces such as parents' and teachers' attitudes and expectations, including stereotypes, also shape self-concept and attitudes toward the subject

¹⁸ males and females are similar on most, but not all, psychological variables

¹⁹ gender equity and other sociocultural factors, not national income, school type, or religion per se, are the primary determinants of mathematics performance at all levels for both boys and girls.

Outros fatores que podem contribuir para explicar a diferença no desempenho de mulheres e homens em testes e merecem ser mencionados: as mulheres são mais afetadas por ansiedade em testes do que homens. Esse efeito se chama ansiedade matemática e é “um estado de desconforto causado pela ação de uma tarefa matemática” (DEVINE, 2012, p.2, tradução nossa²⁰), que é mais forte em meninas do que meninos.

Competição: em média, homens são mais competitivos que as mulheres e “os efeitos em ambientes que envolvem ambos os sexos variam de mulheres falhando em ter uma boa performance em competições até elas evitarem ambientes onde elas deveriam competir” (NIDERLE; VESTERLUND, 2010, p.130, tradução nossa²¹).

Autoeficácia: na teoria social cognitiva de Bandura, as capacidades que um indivíduo acredita possuir influenciam no modo em que ele resolve problemas. Em particular, o trabalho de Machado (2014) se dedica em estudar como estão relacionados a autoeficácia de homens e mulheres em itens do Enem. Todas essas teorias acima listadas podem explicar, provavelmente de maneira conjunta, a diferença de desempenho e na atitude com relação a matemática entre homens e mulheres.

Do ponto de vista da TCC Beilock (2008) aponta, em uma pesquisa que envolve resolução de problemas de matemática, que situações estressantes afetam a Memória de Trabalho que estaria disponível para a resolução do problema. Na pesquisa a situação de estresse baseava o fato de que o exame aplicado possuía relevância acadêmica para os alunos. O fato da Memória de Trabalho ser afetada por situações estressantes pode explicar alguns dos efeitos acima mencionados.

Em um trabalho especificamente sobre desempenho de homens e mulheres adultos em problemas básicos e aplicados de matemática, onde buscava-se observar o efeito da ansiedade matemática, Miller e Bichsel (2004) encontraram que, após a ansiedade matemática, as memórias de trabalho verbais (incluindo leitura) e visuais foram significativas como indicadores de desempenho. Os autores inclusive relatam que as pesquisas em ansiedade concluíram que a ansiedade impacta imediatamente a Memória de Trabalho visual, o que pode explicar o efeito prejudicial para as mulheres no que diz respeito aos problemas contextualizados de matemática. Outro achado, que corrobora os

²⁰ a state of discomfort associated with performing mathematical tasks

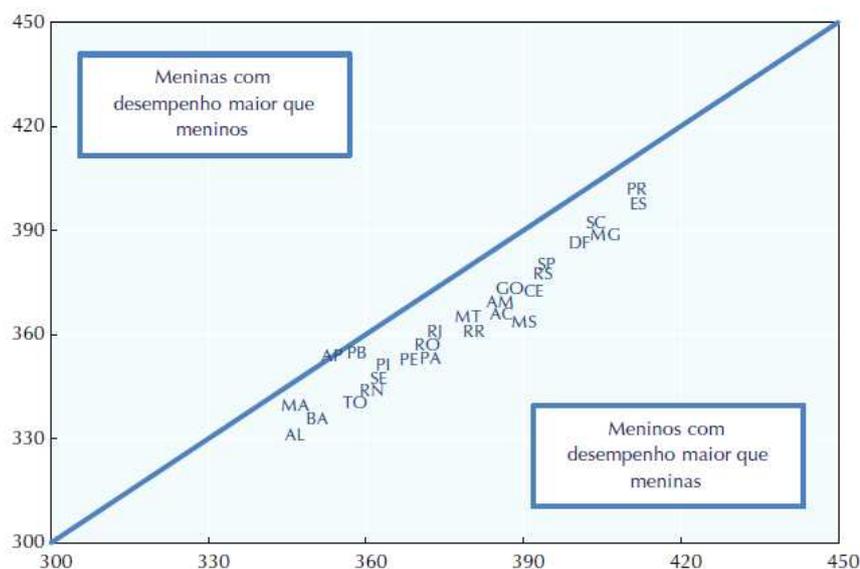
²¹ The effects in mixed-sex settings range from women failing to perform well in competitions to women shying away from environments in which they have to compete

resultados discutidos até aqui, é que problemas aplicados de matemática afetam mais mulheres do que homens.

Buscando uma relação entre desempenho de homens e mulheres, comparando provas de matemática e leitura em diversas séries do ensino fundamental ao médio, Adelson, Dickinson e Cunningham (2015) utilizaram o banco de dados estadual e descobriram que a correlação entre as notas das duas disciplinas era positiva, ou seja, estudantes com notas altas em matemática, possuem notas altas em leitura. Essa correlação só não era forte para estudantes do sexo feminino e também para estudantes não brancos.

De uma maneira sintética, os resultados do PISA são robustos em relação a essa diferença, em 2009 meninos tiveram um desempenho melhor em 35 dos 65 países participantes do PISA. Em 5 países as meninas apresentaram melhor desempenho e em 25 países não houve diferença significativa entre os gêneros (OECD, 2011). Nos ciclos dos anos de 2012 e de 2015, a diferença de desempenho entre os sexos nos países não se modificou significativamente e, embora o número de países que aderiram ao PISA tenha aumentado, o número de países em que os meninos foram melhores que as meninas, e vice versa, se manteve proporcionalmente parecido (OECD 2016).

Figura 5: Escores médios estimados por gênero e unidade da Federação, matemática - PISA 2015

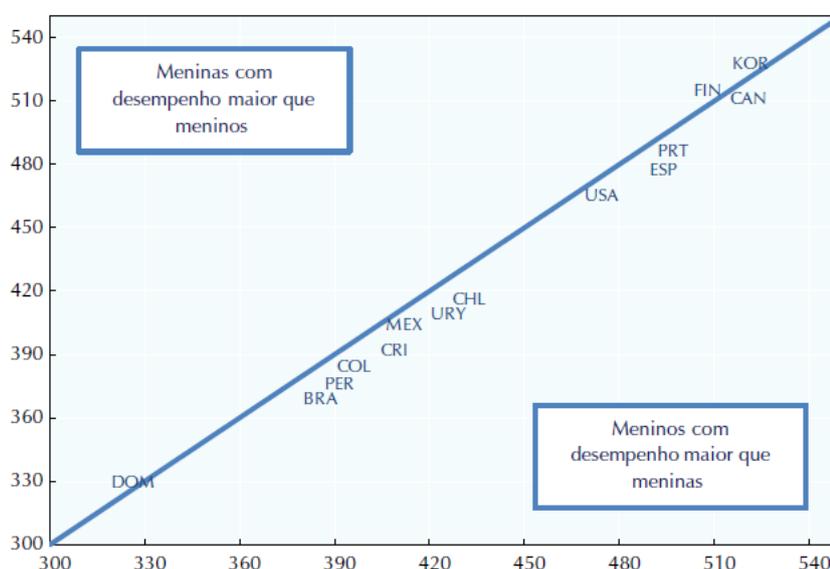


Fonte: OECD, INEP

No ciclo do PISA de 2012 e 2015 o INEP solicitou uma amostragem extra para cada estado brasileiro, o que permitiu comparar os resultados nacionais internamente. Com respeito ao desempenho de matemática entre os sexos, por estado, o desempenho dos homens foi superior em todos os estados brasileiros, com exceção de Amapá e Paraíba, onde ocorreu empate (BRASIL, 2016, p. 176), como mostra a figura 5.

Nesse mesmo ciclo de 2015, o relatório também mostra o escore entre os sexos para uma seleção de países (figura 6), onde verifica-se o efeito de desempenho masculino maior que o feminino, como mencionado anteriormente.

Figura 6: Escores médios estimados por gênero de países selecionados, matemática - PISA 2015

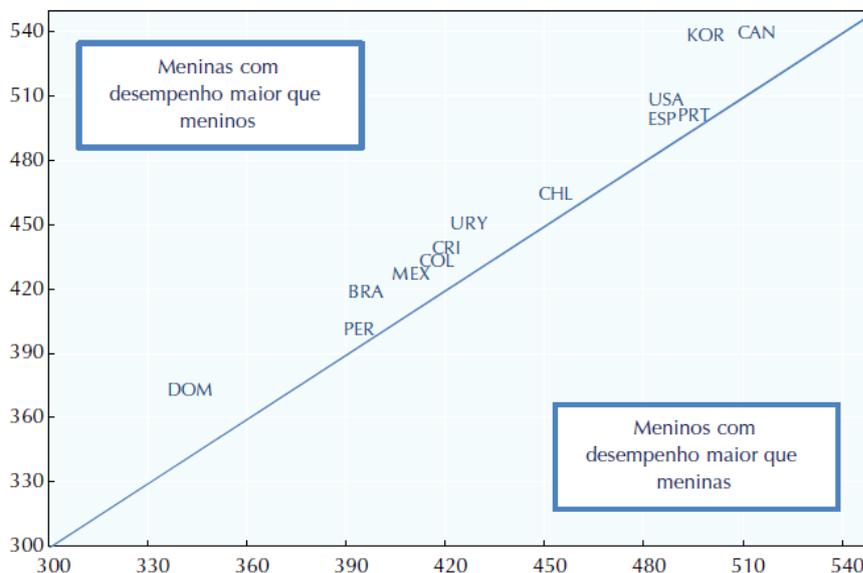


Fonte: OECD, INEP

No espírito de encontrar mecanismos para ajudar o desempenho de matemática em mulheres, podemos olhar aspectos que beneficiam as mulheres em relação aos homens, e existe um assunto em que elas vão consideravelmente melhor que eles: redação, leitura e interpretação. Existem muitos trabalhos sobre essa diferença de desempenho em exames, o que mais chama atenção é o resumo dos dados do PISA de 2009 onde foi detectado que, em habilidades de leitura, mulheres tiveram uma performance melhor que homens em todos os países que participaram do PISA (OECD, 2011, p.16) naquele ciclo, que no caso foram 65 países.

Nos ciclos de 2012 e 2015 o mesmo efeito foi observado, as meninas se sobressaíram em leitura, comparadas aos meninos, em todos os países participantes, muito embora essa diferença de desempenho tenha caído modestamente de maneira geral (OECD, 2016). A figura 7 a seguir apresenta um recorte para alguns países selecionados.

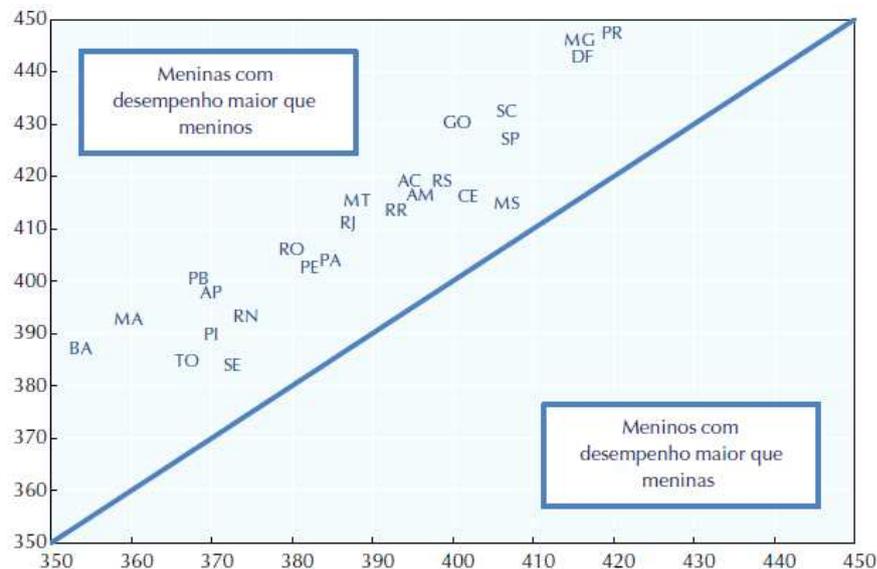
Figura 7: Escores médios estimados por gênero de países selecionados, leitura - PISA 2015



Fonte: OECD, INEP

Os resultados do PISA de 2015 mostram que nas amostras nacionais estratificadas por estado, o desempenho das meninas foi superior ao dos meninos em todas as unidades federativas (BRASIL, 2016, p.135), como mostra a figura 8 a seguir.

Figura 8: Escores médios estimados por gênero e unidade da Federação, leitura - PISA 2015



Fonte: OECD, INEP

Sobre essa questão do desempenho entre homens e mulheres nas provas de leitura do PISA, duas constatações relevantes devem ser mencionadas: mulheres se sobressaem independente do país, da religião, da língua, da igualdade ou desigualdade dos gêneros na sociedade em que elas estão inseridas; o PISA é uma prova, e em

matemática foi discutido brevemente o papel da ansiedade e da competição, e mesmo com esses dois fatores, que particularmente acho secundários frente à estratificação dos gêneros, as mulheres também se sobressaem frente aos homens.

Em diversas avaliações de matemática, internas ou externas, o uso de problemas que envolvem a leitura de texto como um elemento do corpo de sua questão é frequente. No próximo capítulo será discutida a questão do texto nos itens, contextualização em matemática e a TCC.

3. Contextualização e TCC

Segundo a TCC, a maneira como um material ou um item de uma prova é apresentado a um indivíduo, impacta a forma desse indivíduo processar as informações. Em uma avaliação, um item pode exigir o conhecimento de componentes específicos de matemática e também leitura e interpretação da questão, de forma que essa tarefa sobre o texto do item exerce uma carga cognitiva adicional para o estudante.

Os textos em materiais ou itens de matemática podem variar de um enunciado simples, que exige puramente o construto matemático da área de conteúdo em questão, até textos ou gráficos em que o candidato deve extrair uma ou mais informações, as quais podem ou não ser relevantes na resolução do problema em questão. Esse capítulo tem por motivação discutir particularmente o contexto em problemas de matemática, bem como avaliações que se utilizam de questões contextualizadas e o olhar da TCC sobre esse assunto.

Existem algumas correntes e movimentos sobre aprendizado contextual. Em matemática particularmente uma das mais reconhecidas em ensino e aprendizado é a Educação Matemática Realística, que se iniciou na década de 70 e foi fundamentada por Hans Freudenthal.

Freudenthal acreditava na matemática como uma atividade humana, e por isso deveria estar conectada à realidade e às experiências dos alunos. Do ponto de vista de suas crenças, a educação matemática para crianças deve depender da matemática da realidade cotidiana (GRAVEMEIJER; TERWEL, 2000). Este processo pode ser realizado de duas maneiras: Matematização Horizontal que leva do mundo cotidiano e concreto ao mundo dos símbolos e Matematização Vertical que é o processo de reorganização dentro do próprio sistema matemático (GRAVEMEIJER, 1999). Portanto, os designers de instrução devem procurar situações problema que possam funcionar como pontos de partida informais, em que o crescimento cognitivo se torna possível (GRAVEMEIJER; DOORMAN, 1999).

Gravemeijer (1999) define problemas de contexto "como situações problemas que são experiencialmente reais para os alunos". A compreensão final do aluno sobre a matemática formal deve ser ancorada na compreensão desses fenômenos

experiencialmente reais da vida cotidiana (FREUDENTHAL, 1991). Em um momento simultâneo, mudanças são feitas a partir desses problemas contextualizados da vida cotidiana para um enfoque nos conceitos e relações matemáticas, desse modo é possível atingir os dois lados do espectro, situações problemas e rigidez matemática.

Nessa abordagem realística da educação matemática, os problemas de contexto desempenham um papel crucial. Se bem escolhidos, os problemas de contexto oferecem oportunidades para que os alunos desenvolvam conhecimentos informais (cotidianos) e amadurecê-los para símbolos matemáticos. O contexto dá sentido aos conceitos e, em contextos escolares, o conhecimento informal dos alunos pode servir de base para o desenvolvimento da compreensão da matemática (MACK, 1990).

Estudos foram realizados para examinar o efeito de um contexto da vida real para avaliar compreensão de conceitos de matemática, em geral, e as frações em particular (Mack, 1990). Os desfechos das pesquisas mostram que os alunos chegam à escola com um repertório informal de conhecimento sobre as frações que lhes possibilitam resolver os problemas apresentados em um contexto da vida real.

Nesta perspectiva, do ponto de vista da Educação Matemática Realística, tornar os conceitos matemáticos mais cotidianos e informais permite que os alunos estabeleçam relações dos novos conceitos a um esquema existente (MACK, 1990; 1993; 2001).

3.1. Contextualização e Carga Cognitiva Estranha

A familiaridade com o contexto da instrução impacta a carga cognitiva, e movimentos recentes têm estudado como essa redução de carga cognitiva ocorre para facilitar a aprendizagem (CARLSON; CHANDLER; SWELLER, 2003). Uma das sugestões é apresentar aos alunos tarefas realistas e com significados familiares, desse modo ocorrem associações significativas na Memória de Longo Prazo dos participantes (Van MERRIENBOER et al., 2003).

Os autores ancoram toda essa abordagem na teoria do esquema, já mencionada no capítulo anterior. Esses esquemas automatizam e liberam capacidade de Memória de Trabalho (SWELLER, 2003), além de serem responsáveis por resgatar de maneira rápida Memória de Longo Prazo uma vez que um estímulo seja apresentado (COWAN, 1988).

Recordando que os esquemas permitem que muitos elementos sejam tratados como um único elemento, portanto a Memória de Trabalho sofre menos carga. Mais ainda, os especialistas em uma área possuem uma melhor atuação essencialmente por terem seus esquemas à disposição nessa área específica (Van MERRIENBOER et al., 2003).

Estudos recentes, mirando a redução de carga cognitiva intrínseca em materiais complexos de instrução, verificaram que o conhecimento prévio que alunos possuíam, como, por exemplo, treinamento ou familiaridade com o tema resultou em escores maiores se comparados a alunos sem esse tipo de conhecimento (CLARKE et al 2005). Desse modo, é de se esperar que a familiaridade que alguns alunos possuem em alguns contextos se traduz em menor carga cognitiva estranha e, portanto, um desempenho melhor em avaliações, se comparado a seus pares sem essa familiaridade de contexto.

Métodos de redução de carga cognitiva estranha existem na literatura, e em particular para a utilização desses métodos em itens de avaliações de matemática, Gillmore, Poggio e Embretson (2010) fizeram uma revisão dos processos que resultou na Tabela 1 a seguir.

Tabela 1 - Estratégias de redução de carga cognitiva em itens

Método	Descrição	Referência
Tradução	Reduzir a contagem de palavras e simplificar a linguagem	Kettler et al. 2011
Ajuda visual	Uso de diagramas para representar informações espaciais	Clark et al. 2011
Sinalização	Focar atenção em sinais e sugestões	Clark et al. 2011
Extirpação	Aparar o conteúdo até o essencial. Eliminação de efeitos visuais e textos estranhos	Clark et al. 2011
Sequenciamento	Fazer a pergunta antes para guiar o candidato no item, e então colocar informações relevantes.	Clark et al. 2011
Estética	Formatar o item de maneira estética e lógica. Colocar o texto perto das figuras	Miller 2011
Simplificação Numérica	Usar valores menores, arredondados e números familiares quando os números não têm papel no construto	Hitch 1978

Fonte: Gillmore, Poggio e Embretson (2010) adaptado, tradução nossa.

Essa tabela nos permite olhar para itens de matemática de avaliações em larga escala, como o Enem e o Pisa, e ponderar se o item foi construído de maneira que o construto matemático possa ser prejudicado por informações que não são relevantes para

a resolução do problema, causando um ruído nos resultados do teste. Como a tabela mostra, empregar ações como reduzir número de palavras, retirar textos estranhos, ou sinalizar informações parecem intuitivas, no que diz respeito à redução de carga cognitiva estranha.

3.2. Contextualização no Enem e no Pisa

As provas de matemática do Pisa e do novo Enem, que começou a ser aplicado a partir de 2009, possuem algumas características em comum, como por exemplo, uma matriz de áreas do conteúdo consideravelmente próximas, e também o uso de problemas contextualizados, no sentido de situações problemas experientialmente reais para os alunos.

Nas próximas seções serão exploradas as relações entre essas duas avaliações, começando pelo Pisa, cuja prova de matemática exerceu grande influência no novo Enem, em especial com o conceito de letramento matemático, que permeia todas as áreas do conhecimento e a forma como os problemas contextualizados são apresentados.

3.3. Letramento Matemático - panorama

O conceito que permeia o PISA é o letramento, e em matemática, o letramento matemático (mathematical literacy) definido como:

A capacidade de formular, empregar e interpretar a matemática em uma série de contextos, o que inclui raciocinar matematicamente e utilizar conceitos, procedimentos, fatos e ferramentas matemáticos para descrever, explicar e prever fenômenos. Isso ajuda os indivíduos a reconhecer o papel que a matemática desempenha no mundo e faz com que cidadãos construtivos, engajados e reflexivos possam fazer julgamentos bem fundamentados e tomar as decisões necessárias. (OECD, 2016 p. 138)

A ideia por de trás do letramento matemático está na necessidade de entender o que um cidadão deve saber e ser capaz de fazer em situações que envolvem matemática, uma vez que é cada vez maior a porção de problemas e situações encontradas no cotidiano que exigem algum grau de compreensão de matemática, acarretando em boas tomadas de decisão por parte das pessoas. Gráficos no jornal, imposto de renda, análise de juros em empréstimos, tomada de decisões cotidianas para minimizar tempo gasto e compras a

prazo são apenas alguns exemplos da real necessidade de aprender e compreender a matemática, e como sugere a definição de letramento matemático, o pano de fundo para o desenvolvimento da matemática sugerida está no contexto.

A partir da definição de letramento matemático, é possível fazer uma análise utilizando três aspectos correlacionados. (OECD, 2016 p. 139)

- Os processos matemáticos que descrevem o que os indivíduos fazem para conectar o contexto de um problema com a matemática e, assim, resolver o problema, bem como as capacidades subjacentes a esses processos.
- O conteúdo matemático visado para uso nos itens da avaliação.
- O contexto no qual estão situados os itens.

Sob a ótica do PISA e do letramento matemático, os alunos devem encarar três processos ao enfrentar um problema matemático: formular, aplicar e interpretar. Formular matemática se reside na possibilidade de utilizar a matemática para abordar, resolver e compreender problemas ou desafios, utilizar estruturas matemáticas para identificar variáveis, fazer hipóteses e simplificar o problema. Aplicar matemática envolve aplicar o pensamento matemático e utilizar conceitos, procedimentos, fatos e ferramental matemático para derivar uma solução, envolve a parte dos cálculos e manipulações de estruturas algébricas da resolução. Interpretar matemática está associado a reflexão e interpretação do resultado obtido, o confronto do contexto da situação problema com a solução e a plausibilidade desta em relação às condições iniciais apresentadas.

3.3.1. Contextualização no Pisa

As categorias de contexto, em que se apresentam os problemas, são identificadas como as áreas da vida no qual esses problemas emergem. Estes contextos podem ser classificados em quatro categorias, que o PISA apresentou como naturezas. A natureza pessoal está associada a problemas que confrontam o indivíduo, parceiros, grupos ou a família. O contexto de natureza social é aquele que envolve o indivíduo e sua posição, na comunidade local, bairro, ou em escalas mais amplas. Contexto ocupacional está associado ao mundo do trabalho. Por último o contexto de natureza científica é aquele relacionado à aplicação matemática ao mundo científico ou tecnológico.

A caracterização também está associada ao tipo de desafio que está inserido no problema apresentado, são quatro categorias (áreas) de conteúdo que se dispõem a abarcar o espectro amplo dos fenômenos matemáticos. Essas quatro áreas de conteúdo matemático são: quantidade, incerteza e dados, mudanças e relações, e espaço e forma (OECD, 2016. p.146,147).

Quantidade: está presente na quantificação de características de objetos, relações, situações e entidades no mundo, na compreensão de várias representações de quantificações e no julgamento de interpretações e argumentos baseados em quantidades. Para se envolver com a quantificação do mundo, é necessário compreender medidas, contas, grandezas, unidades, indicadores, tamanhos relativos e tendências e padrões numéricos. Aspectos do raciocínio quantitativo, como a percepção dos números, a compreensão da múltipla representação de números, o requinte no cálculo mental e computacional, a estimativa e a avaliação da aceitabilidade de resultados, são a base do letramento matemático no que se refere a quantidade.

Incerteza e dados: está presente nas ciências, em tecnologias e na vida cotidiana. É um fenômeno central na análise matemática de muitas situações-problema, e, para lidar com ela, foram criadas a teoria das probabilidades e a estatística, bem como as técnicas de representação e descrição de dados. A categoria de conteúdo de incerteza e dados inclui identificar o lugar da variação nos processos, com a percepção da quantificação dessa variação, do reconhecimento da incerteza, do erro na medição e do conhecimento das probabilidades. Isso também envolve formular, interpretar e avaliar conclusões tiradas em situações nas quais a incerteza é um aspecto central. A apresentação e a interpretação dos dados são conceitos-chave nessa categoria.

Mudanças e relações: o mundo natural e o mundo projetado pelo homem têm uma série de relações temporárias e permanentes entre objetos e circunstâncias, nos quais acontecem mudanças nos sistemas de objetos inter-relacionados ou em circunstâncias em que os elementos se influenciam mutuamente. Em muitos casos, as mudanças ocorrem ao longo do tempo; em outros, mudanças em um objeto ou circunstância estão relacionadas a mudanças em outros. Algumas dessas situações envolvem mudanças discretas; outras mudam continuamente. Certas relações são de natureza permanente ou invariante. Ser mais letrado sobre mudanças e relações envolve compreender os tipos de mudanças fundamentais e reconhecer quando elas ocorrem a fim de utilizar modelos matemáticos

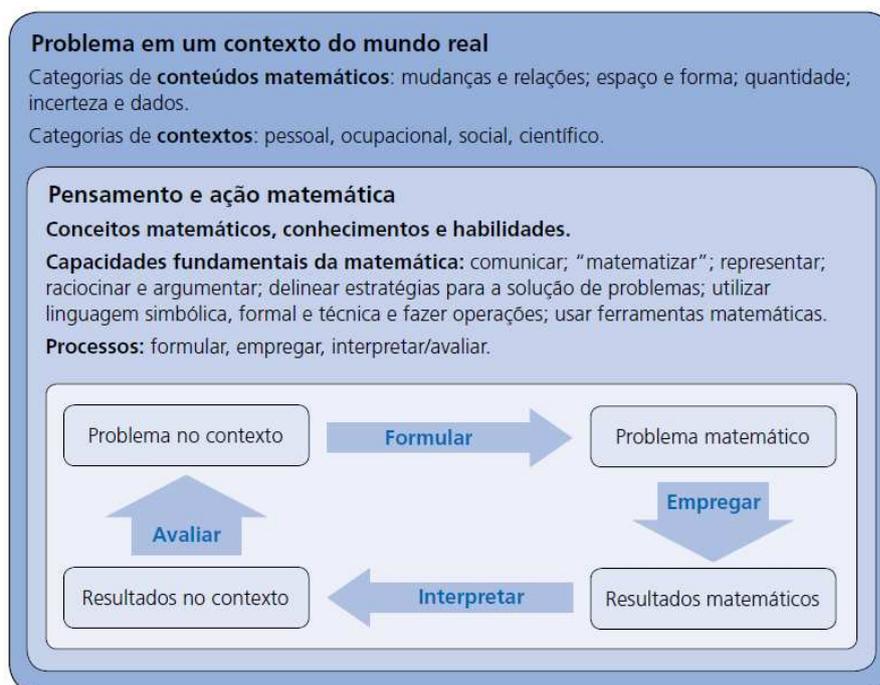
adequados para descrevê-las e predizê-las. Matematicamente, isso significa modelar essas mudanças e relações com funções e equações apropriadas, bem como criar, interpretar e traduzir as diversas representações de relações gráficas e simbólicas.

Espaço e forma: Essa categoria abrange uma ampla gama de fenômenos encontrados em vários lugares no mundo físico e visual: padrões; propriedades dos objetos; posições e orientações; representação dos objetos; codificação e decodificação de informação visual; navegação e interação dinâmica com formas reais e com suas representações. A geometria pode ser considerada uma área-base para o espaço e a forma, mas essa categoria vai além da geometria tradicional no que se refere a conteúdo, significado e método, utilizando-se de recursos de outras áreas matemáticas, como visualização espacial, medida e álgebra. O letramento matemático nessa área envolve uma série de atividades, como o entendimento de perspectiva (por exemplo, na pintura), a criação e leitura de mapas, a transformação de formas com e sem uso de tecnologias, a interpretação de paisagens de cenas tridimensionais de várias perspectivas e a construção de representações de formas.

Uma primeira ideia, de um indivíduo externo à visão do letramento matemático, seria tentar classificar de maneira única cada uma das áreas de conteúdo, como uma bijeção entre conteúdos escolares usuais com as propostas do PISA. Por exemplo, espaço e forma seria a conhecida geometria, incerteza e dados poderia ser colocado como estatística e análise combinatória, e assim por diante. Muito embora pareça natural fazer essa associação, temos que tomar cuidado com regiões de interseção, inclusive mencionadas nas definições de cada uma dessas áreas de conteúdo. **Espaço e forma** contempla elementos de medida e álgebra, que poderiam ser pensadas em pertencer exclusivamente a mudanças e relações.

De uma maneira sintética, as ideias apresentadas anteriormente se inter-relacionam como mostrado no seguinte modelo de letramento matemático na prática.

Figura 9: Modelo de Letramento no PISA



Fonte: OECD, 2016 p. 140

Quanto aos conteúdos de matemática inseridos e avaliados pelo PISA, temos a seguinte lista com uma breve descrição (OECD, 2016):

Funções: O conceito de função enfatiza (mas não se limita a) funções lineares, suas propriedades e uma série de descrições e suas representações. As representações mais comuns são as verbais, simbólicas, tabulares e gráficas.

Expressões algébricas: Interpretação verbal ou manipulação de expressões algébricas, envolvendo números, símbolos, operações aritméticas, potências e raízes simples.

Equações e inequações: Equações lineares e relacionadas, inequações, equações simples de segundo grau e métodos de solução analíticos e não analíticos.

Sistemas coordenados: Representação e descrição de dados, posições e relações.

Relações com e entre objetos geométricos em duas ou três dimensões: Relações estáticas como conexões algébricas entre elementos de figuras (por exemplo: teorema de Pitágoras como definidor de relações entre os comprimentos dos lados de um triângulo retângulo), posições relativas, semelhança e congruência, relações dinâmicas

envolvendo transformações e movimentação de objetos, e correspondência entre objetos bi e tridimensionais.

Medidas: Quantificação das características de formas e objetos, como medidas de ângulo, distância, comprimento, perímetro, circunferência, área e volume.

Números e unidades: Conceitos, representações de números e sistemas numéricos, incluindo propriedades de números inteiros e racionais, aspectos relevantes dos números irracionais, bem como quantidades e unidades referentes a fenômenos como tempo, dinheiro, peso, massa, temperatura, distância, área e volume, e quantidades derivadas com sua descrição numérica correspondente.

Operações aritméticas: Natureza e propriedades dessas operações e respectiva notação convencional.

Porcentagem, razão e proporção: Descrição numérica de magnitude relativa e aplicação de proporções e raciocínios com proporcionalidade para resolver problemas.

Princípios de contagem: Combinações e permutas simples.

Estimativas: Aproximação de quantidades e de expressões numéricas com uma finalidade, incluindo dígitos significativos e arredondamentos.

Coleta, representação e interpretação de dados: Natureza, origem e coleta de vários tipos de dados, e as diferentes formas de representá-los e interpretá-los.

Variabilidade e descrição de dados: Conceitos como variabilidade, distribuição e tendência central de conjunto de dados, e formas de descrevê-los e interpretá-los em termos quantitativos.

Amostras e amostragem: Conceitos de amostra e amostragem de dados populacionais, incluindo inferência simples baseada nas propriedades da amostra.

Risco e probabilidade: Noção de eventos randômicos, variação randômica e sua representação, risco e frequência de eventos, e aspectos básicos do conceito de probabilidade.

Lembrando que o PISA tenta ser uma avaliação não baseada em currículo, mas que estipula que alunos em idade escolar entre 15 anos e meio e 16 anos e meio já tenham tido contato com esses conteúdos matemáticos até essa idade.

3.3.2. Contextualização no Enem

Como mencionado no capítulo anterior, a mudança no Enem fez com que Matemática e suas tecnologias tivessem 45 itens a partir de 2009. Na matriz de referência (**Anexo III**) é possível encontrar, além da matriz de competências e habilidades, os objetos de conhecimento associados às matrizes de referência. No caso da matemática são cinco objetos de conhecimento: numéricos, geométricos, estatística e probabilidade, algébricos e algébricos/geométricos. A descrição dos conhecimentos específicos para cada um desses objetos de conhecimento são as seguintes:

Conhecimentos numéricos: operações em conjuntos numéricos (naturais, inteiros, racionais e reais), desigualdades, divisibilidade, fatoração, razões e proporções, porcentagem e juros, relações de dependência entre grandezas, sequências e progressões, princípios de contagem.

Conhecimentos geométricos: características das figuras geométricas planas e espaciais; grandezas, unidades de medida e escalas; comprimentos, áreas e volumes; ângulos; posições de retas; simetrias de figuras planas ou espaciais; congruência e semelhança de triângulos; teorema de Tales; relações métricas nos triângulos; circunferências; trigonometria do ângulo agudo.

Conhecimentos de estatística e probabilidade: representação e análise de dados; medidas de tendência central (medias, moda e mediana); desvios e variância; noções de probabilidade.

Conhecimentos algébricos: gráficos e funções; funções algébricas do 1º e do 2º grau, polinomiais, racionais, exponenciais e logarítmicas; equações e inequações; relações no ciclo trigonométrico e funções trigonométricas.

Conhecimentos algébricos/geométricos: plano cartesiano; retas; circunferências; paralelismo e perpendicularidade, sistemas de equações.

Observe que uma avaliação nacional do Ensino Médio que apresenta uma matriz de referência, com competências e habilidades, juntamente com um largo espectro de áreas de conteúdo deve ser orientada por documentos mais abrangentes na área de educação. Envolvendo o Enem existem ainda os PCNs, e é de se esperar que os documentos educacionais estejam em conformidade com a avaliação. Sobre os PCNs vale ressaltar no documento o destaque da abrangência que a matemática possui e como ela é transdisciplinar, e como competências e habilidades o documento lista a representação e comunicação, além de investigação e compreensão e, finalmente, a contextualização sócio cultural (BRASIL, 1998).

Uma grande dúvida em relação a ambição do Enem está nessa direção, será que essa avaliação de grandes proporções consegue contemplar essas competências e habilidades, especificamente investigação, compreensão e contextualização? Outra dúvida pertinente está em relação ao conteúdo da prova, uma vez que ela possui cinco objetos do conhecimento, recheados de conteúdos específicos.

Em uma análise sobre os quatro primeiros anos (2009 a 2012) de aplicação desse Enem reformulado, Rodrigues (2013) explorou, como pergunta de pesquisa, se as questões da prova de matemática estavam "distribuídas em conformidade com a Matriz de Referência contida nos aportes metodológicos do Novo Enem". No mesmo trabalho foram exploradas outras perguntas pertinentes, como a quantificação da frequência dos itens com relação a competências, habilidades, contextualização, interdisciplinaridade e áreas de conteúdo. Para tal propósito o pesquisador utilizou análise de conteúdo.

Posteriormente o mesmo autor, juntamente com outros colegas, realizou uma pesquisa atualizada, de modo que contemplasse todas as provas de matemática do Enem até 2016, utilizando as mesmas perguntas de pesquisa (NASCIMENTO et al, 2017). Na pesquisa, foi considerado que um item é contextualizado se ele "representar uma situação que pode estar presente no cotidiano dos alunos e é possível interpretar, verificar, analisar para resolver priorizando o raciocínio lógico e não as fórmulas envolvendo conteúdos matemáticos" e um item que possui uma situação problema de matemática foi categorizado se "mesmo que possuindo um contexto, este não relata uma situação do dia-a-dia, sendo o foco o conteúdo matemático em si mesmo, exigindo assim a utilização de fórmulas e cálculos matemáticos para a resolução da questão".

De 2009 a 2016, os autores classificaram 360 itens de matemática com relação a sua área de conhecimento, e dentro dessas categorias distinguiram se o item era contextualizado ou se era situação problema.

Tabela 2: Formato dos Itens do ENEM por Objeto de Conhecimento

Objeto de Conhecimento	Incidência (%)	Contextualização	Situação Problema
Conhecimentos Numéricos	37,3	123	11
Conhecimento de Estatística e Probabilidade	21,6	78	70
Conhecimentos Geométricos	21,1	66	10
Conhecimentos Algébricos	16,7	44	16
Conhecimento Algébrico/geométrico	3,3	6	6
Total	100	309	51

Fonte: NASCIMENTO et al, 2017

Os itens do Enem reformulado estão majoritariamente distribuídos nos objetos de conhecimento Numérico, Estatística e Probabilidade e Geométricos, além de serem em sua maioria problemas contextualizados.

Em relação às competências presentes na matriz de competências (**Anexo III**), os pesquisadores também efetuaram uma categorização, juntamente com a classificação do tipo de problema: contextualizado ou situação problema.

Tabela 3: Competências da Matriz de Referência do Novo ENEM no Período de 2009 a 2016

Competência do Enem	Contextualização	Situação Problema	Total
Competência 1	36	4	40
Competência 3	70	6	76
Competência 3	39	11	50
Competência 4	44	4	48
Competência 5	50	18	68
Competência 6	50	3	53
Competência 7	20	5	25
Total	309	51	360

Fonte: NASCIMENTO et al, 2017

O Enem contempla todas as competências ao longo desses anos e, na maior parte das vezes, em problemas contextualizados, em conformidade com as competências e habilidades dos PCNs, particularmente no quesito de contextualização e conexão com a realidade do aluno.

Interdisciplinaridade também foi objeto de pesquisa para os autores, e eles concluíram que 38% dos itens no período possuíam conexão com outra disciplina, as três

disciplinas mais envolvidas, que juntas somaram mais da metade dos itens interdisciplinares foram Geografia, Biologia e Física.

A proposta em matemática do Enem reformulado, com itens envolvendo competências e habilidades associadas a contextualização com pitadas de interdisciplinaridade, está bem alinhada com a proposta do de matemática do PISA, que como já mencionamos, é o conceito de letramento matemático (OECD, 2003).

[...] que está preocupado com a capacidade dos estudantes analisarem, raciocinarem e comunicarem efetivamente enquanto levantam, resolvem e interpretam problemas matemáticos em uma variedade de situações envolvendo conceitos quantitativos, espaciais, probabilísticos e outros.

A já mencionada matriz de referência do Enem apresenta um conjunto de eixos cognitivos similares, associados a dominar linguagens; compreender fenômenos; enfrentar situações-problema; construir argumentação e elaborar propostas.

Em um trabalho sobre o letramento matemático e o Enem, Ortigão e Aguiar citam duas dissertações de mestrado que concluem que

A matriz conceitual de avaliação do PISA, que tem como referência principal a articulação entre o conceito de educação básica e o de cidadania, está em perfeita sintonia com os objetivos educacionais do Ensino Médio, propostos pelo Ministério da Educação – MEC [...]. Ela está presente também no Enem, cujos preceitos demonstram seu caráter transdisciplinar e o entendimento do perfil terminativo desse nível de ensino, complementando o aprendizado iniciado no Ensino Fundamental. (ORTIGÃO; AGUIAR, 2012, p.3)

As semelhanças entre Enem e PISA ainda transcendem para o plano de objetos do conhecimento, que para a avaliação internacional se chama áreas de conteúdo como já mencionado no capítulo sobre o PISA. Com exceção do objeto de conhecimento algébrico/geométrico, que essencialmente é geometria analítica, é possível fazer uma bijeção entre os objetos de conhecimento e áreas de conteúdo. **Conhecimentos numéricos** está associado a área de **Quantidade, Estatística e Probabilidade** se relaciona com **Incerteza**, **Conhecimentos Geométricos** para o PISA é a área de **Espaço e Forma** e finalmente **Conhecimentos Algébricos** está ligado a **Mudanças e Relações**.

Sendo razoável a comparação dos objetivos educacionais entre essas duas provas, Enem e PISA, as pesquisas que envolvem comparação do desempenho entre homens e

mulheres na prova de matemática do PISA também se apresentam como um norte para o mesmo tipo de análise nas provas de matemática do Enem.

3.3.3. Contexto em Testes de Múltipla Escolha e Carga Cognitiva Estranha

Uma vez que já discutimos o assunto de contextualização em problemas de matemática, e também foi apresentada uma tabela que consiste em mecanismos de redução de carga cognitiva estranha em materiais a serem processados, é possível tomar itens de matemática e submetê-los à esse processo.

Como exemplo, tomamos dois itens do artigo de Gillmore, Poggio e Embretson (2010) para observar como o trabalho de redução de carga cognitiva foi realizado e quais os métodos que foram utilizados, daqueles presentes na Tabela 1. Esses dois itens tiveram diferença estatística significativa de desempenho entre os grupos que fizeram a prova com itens originais (grupo controle) e itens com carga cognitiva reduzida (grupo experimental). Além disso, esses dois exemplos são adequados para questionar a construção do item original em relação ao construto que se pretende avaliar.

Quadro 1 - Item 2 do artigo em sua forma tradicional e submetido à redução de carga cognitiva

<u>Item tradicional (grupo controle)</u>	<u>Item de carga reduzida (grupo experimental)</u>
Alice, Brad, Cory e Derek tiveram, cada um, pizza para almoçar. Alice comeu $\frac{3}{10}$ de sua pizza; Brad comeu 42%, Cory comeu $\frac{2}{5}$ de sua pizza, e Derek comeu 45%. Quem deixou a maior quantidade de pizza sobrando após o almoço?	Qual valor abaixo possui o menor valor de pizza?
A. Alice	A. $\frac{3}{10}$ de pizza
B. Brad	B. 42% de pizza
C. Cory	C. $\frac{2}{5}$ de pizza
D. Derek	D. 45% de pizza

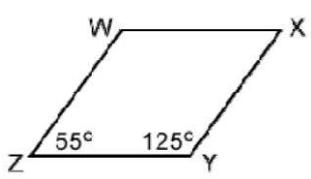
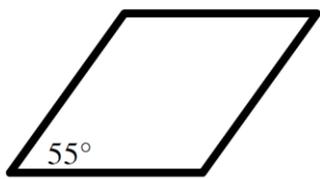
Fonte - Gillmore, Poggio e Embretson (2010) - Tradução nossa.

No item 2 da prova dos pesquisadores, apresentado no quadro 1, o construto proposto pelo avaliador se refere a comparação de ordem (maior ou menor) entre números escritos na forma de porcentagem e fração. Dois fatores aqui podem prejudicar os respondentes, o primeiro está no comando do item, uma vez que não se pede diretamente qual é a maior fração, e sim o complementar, quem deixou mais pizza. O segundo fator

está na contextualização, uma vez que para se medir o construto, foram invocados quatro personagens. O caminho do aluno avaliado não é imediato, ele tem que associar um personagem a um número, perceber que o comando está associado ao complementar desse número e finalmente comparar os quatro valores.

O *designer* que reduziu a carga cognitiva do item 2 assumiu que o construto é de fato a comparação numérica de ordem entre fração e porcentagem, por isso excluiu a presença dos personagens e da contextualização, ainda deixou o passo dos alunos calcularem o valor complementar, muito embora tenha enfatizado isso ao sublinhar essa informação. Nesse item os autores utilizaram as técnicas de: sinalização, ao sublinhar a palavra "menor"; tradução; estética, uma vez que o texto e as alternativas ficaram bem próximas; e extirpação.

Quadro 2 - Item 13 do artigo em sua forma tradicional e submetido à redução de carga cognitiva

<u>Item tradicional (grupo controle)</u>	<u>Item de carga reduzida (grupo experimental)</u>
<p>As propriedades do quadrilátero WXYZ estão descritas abaixo.</p>  <p>O lado XY é paralelo e congruente ao lado WZ. O lado WX é paralelo e congruente ao lado ZY. $\angle WZY = 55^\circ$. $\angle XYZ = 125^\circ$.</p> <p>Qual termo descreve o quadrilátero WXYZ?</p> <p>A. Paralelogramo B. Retângulo C. Quadrado D. Trapézio</p>	<p>Qual termo descreve esse formato? Lados opostos são paralelos e possuem o mesmo comprimento.</p>  <p>A. Paralelogramo B. Retângulo C. Quadrado D. Trapézio</p>

Fonte - Gillmore, Poggio e Embretson (2010) - Tradução nossa.

O quadro 2 apresenta o item 13 da prova, o construto em questão está nas propriedades geométricas de quadriláteros, mais especificamente que um paralelogramo é figura com dois pares de lados paralelos. Uma consequência desse construto, em particular, é que um paralelogramo possui lados opostos congruentes, essa informação não era necessária para acertar o item. Em sua versão original, o corpo do item definiu o quadrilátero por meio de seus vértices, associou pares de vértices a lados para mostrar

casos de mesma medida, e ainda inseriu dois ângulos. Aqui muitas informações poderiam afastar alunos com alguma proficiência em geometria da resposta certa: a definição do quadrilátero pelos vértices, os lados pelos vértices e os ângulos eram desnecessários para medir o construto associado ao paralelogramo.

Ao submeter esse item à redução de carga cognitiva, os *designers* foram muito felizes na escolha de remover todas essas informações, exceto um dos ângulos, que indica não ser uma visão em perspectiva de um retângulo ou de um quadrado. É possível identificar as técnicas de redução de carga cognitiva utilizadas: sequenciamento, uma vez que a pergunta está no começo; tradução; e extirpação, uma vez que tanto na figura quanto no texto, informações desnecessárias foram removidas. É importante atentar para o fato de que em nenhum dos dois itens modificados houve informação adicional para ajudar os estudantes na resolução, a transformação do item foi motivada pela inconsistência entre o construto que pretendia ser avaliado e o *design* do item, seu corpo, comando e alternativas.

3.3.4. Aplicações em Itens do Enem

A expertise compartilhada por Gillmore, Poggio e Embretson (2010) nos itens mostrados, nos permite aplicar os métodos de redução de carga cognitiva estranha em itens de matemática de avaliações em larga escala. É muito relevante destacar que esse processo não é único e depende da escolha do *designer*. Uma análise do item, no que diz respeito ao construto pretendido pelo item, é crucial para poder entender que tipo de carga cognitiva se torna estranha no item em questão.

Como exemplo de aplicação dessa redução de carga cognitiva estranha, baseada nos trabalhos de Gillmore, Poggio e Embretson (2010), tomamos alguns itens de matemática do Enem reformulado que utilizamos no nosso trabalho. Nos capítulos seguintes outros itens que se destacaram nas análises serão mencionados, bem como suas características e o processo de redução de carga cognitiva estranha. Colocamos a versão Original dos itens na esquerda, da forma que foi apresentada no exame; já a versão Reduzida, na qual aplicamos o processo de redução de carga cognitiva estranha, está apresentada à direita.

Quadro 3 - Item 8 da Prova e sua respectiva versão Reduzida

<u>Item 8 - Original</u> (Enem 2011 - 171)	<u>Item 8 - Reduzido</u>
<p>Nos últimos cinco anos, 32 mil mulheres de 20 a 24 anos foram internadas nos hospitais do SUS por causa de AVC. Entre os homens da mesma faixa etária, houve 28 mil internações pelo mesmo motivo.</p> <p>Época. 26 abr. 2010 (adaptado).</p> <p>Suponha que, nos próximos cinco anos, haja um acréscimo de 8 mil internações de mulheres e que o acréscimo de internações de homens por AVC ocorra na mesma proporção. De acordo com as informações dadas, o número de homens que seriam internados por AVC, nos próximos cinco anos, corresponderia a</p> <p>a) 4 mil b) 9 mil c) 21 mil d) 35 mil e) 39 mil</p>	<p>Em uma cidade, 32 mil mulheres e 28 mil homens foram hospitalizados no ano de 2017. Suponha que em 2018, ocorra um acréscimo de 8 mil mulheres hospitalizadas.</p> <p>Considerando que ocorra um acréscimo de homens hospitalizados na mesma proporção, qual o total de homens hospitalizados em 2018?</p> <p>a) 4 mil b) 9 mil c) 21 mil d) 35 mil e) 39 mil</p>

O item 8 está apresentado, em suas duas versões, no quadro 3 acima. O construto do item afere sobre razão e proporção aplicadas a um contexto. No item original, o enunciado apresenta um texto com informações relevantes à resolução do problema, mas nesse texto existem muitas informações que não são necessárias para aferir o construto pretendido, como as siglas AVC e SUS, que devem exercer carga cognitiva estranha aos respondentes.

Para o item reduzido, essas siglas foram removidas do corpo do item, mas o mesmo contexto foi mantido. Foram utilizadas as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que eliminamos muitas palavras e também simplificamos a linguagem; e Extirpação, uma vez que parte do texto associado ao contexto foi removido.

O item 17, em suas duas versões, está localizado no quadro 4 a seguir. Sobre o construto do item, entendemos que é obter o valor de uma operação aritmética. Para o item reduzido ainda deixamos um contexto para os respondentes, mas removemos a parte do contexto associada à motivação para se utilizar um vaso sanitário ecológico, mencionando inclusive a ABNT, o que aumenta a carga cognitiva estranha.

Quadro 4 - Item 17 da Prova e sua respectiva versão Reduzida

<u>Item 17 - Original</u> (Enem 2012 - 169)	<u>Item 17 - Reduzido</u>
<p>Há, em virtude da demanda crescente de economia de água, equipamentos e utensílios como, por exemplo, as bacias sanitárias ecológicas, que utilizam 6 litros de água por descarga em vez dos 15 litros utilizados por bacias sanitárias não ecológicas, conforme dados da Associação Brasileira de Normas Técnicas (ABNT).</p>	<p>O vaso sanitário ecológico utiliza 6 litros de água por descarga, já o vaso sanitário convencional utiliza 15 litros de água por descarga. Considere que uma família com o vaso sanitário convencional utiliza 60 litros por dia com a descarga.</p>
<p>Qual será a economia diária de água obtida por meio da substituição de uma bacia sanitária não ecológica, que gasta cerca de 60 litros por dia com a descarga, por uma bacia sanitária ecológica?</p>	<p>Se essa família tivesse um vaso sanitário ecológico ela economizaria quanta água por dia?</p>
<p>a) 24 litros b) 36 litros c) 40 litros d) 42 litros e) 50 litros</p>	<p>a) 24 litros b) 36 litros c) 40 litros d) 42 litros e) 50 litros</p>

Outra mudança substancial para o item reduzido está no comando, uma vez que o item original ainda possui dados no comando e decidimos na versão reduzida dispor esse dado no próprio corpo. As estratégias de redução de carga cognitiva utilizadas foram: Tradução, uma vez que reduzimos um pouco a quantidade de palavras; Extirpação, já que retiramos parte do texto, reduzindo o problema à essência do que ele quer saber do estudante; e Estética, pois entendemos que o comando não estava muito claro na versão original, e as informações foram colocadas de maneira mais natural para a resolução dos estudantes.

Como mencionado, o processo de redução de carga cognitiva estranha depende do *designer*, e pode ser feito de diversas maneiras. Alguns itens podem se mostrar mais fáceis de aplicar esse processo, mas o grande tema é perceber que apenas aplicando um instrumento de avaliação é possível saber se um item transformado dessa maneira efetivamente mudou a postura dos respondentes.

4. Procedimentos metodológicos

Os objetivos desse trabalho, assim como a pergunta de pesquisa, estão relacionados com o desempenho dos alunos na prova de matemática do Enem. As análises envolvem as relações entre o sexo do candidato, escolaridade dos pais e desempenho em itens de matemática que apresentam carga cognitiva estranha e que sejam passíveis de redução desse tipo de carga cognitiva. Analisaremos também as inter-relações dessas variáveis no impacto do desempenho dos candidatos.

Para tais análises poderíamos partir de dados já existentes, que contemplassem os objetivos da pesquisa, ou ainda confeccionar uma prova e um questionário a serem aplicados a um grande número de respondentes. Mesmo no segundo caso, ao se elaborar uma nova prova, é extremamente desejável ter resultados prévios que envolvam alguns elementos dos objetivos do trabalho para balizar o formato do questionário socioeconômico a ser aplicado e os itens da prova de matemática a serem escolhidos.

O INEP disponibiliza em sua página eletrônica²² os microdados de todas as avaliações do Enem, desde seu início. Esses microdados geram arquivos grandes, em ordem de grandeza de gigabytes, e ficam disponíveis, entre um ou dois anos, após a aplicação de cada edição do Enem. Nesses arquivos são encontradas informações sobre o perfil socioeconômico dos candidatos, uma vez que cada indivíduo que presta o Enem preenche um questionário socioeconômico. No conjunto de perguntas desse questionário estão informações importantes para nossa pesquisa como o sexo do candidato, escolaridade dos pais e renda familiar, além de idade, situação de conclusão do ensino médio e tipo de colégio.

Os microdados fornecem também todas as respostas dos candidatos do Enem nas cinco provas, ou seja, as alternativas dos 180 itens de múltipla escolha e também a nota total de redação, com suas notas discriminadas por cada uma das cinco competências de redação. Os microdados foram tratados com o objetivo de unificar em um único arquivo os quatro modelos (cores) de prova aplicados.

Uma vez que os objetivos e a pergunta de pesquisa exigiam um espectro bem amplo e diverso de sujeitos de pesquisa, o primeiro desafio foi encontrar um conjunto de estudantes que contemplassem essa condição inicial. A pesquisa contou com uma grande

²² portal.inep.gov.br/microdados

colaboração de colegas professores de diversas cidades, de modo que pudesse contemplar localidades e experiências escolares distintas. Os autores agradecem a todos os colegas que propiciaram e colaboraram com a presente pesquisa.

4.1. Sujeitos de Pesquisa

O Instituto Federal de São Paulo (IFSP) possui 37 campus em todo o estado, e com suas políticas de inclusão e ações afirmativas nos últimos anos, possibilitaram que alunos oriundos exclusivamente de ensino público no Ensino Fundamental, alunos com poucas condições financeiras, alunos pretos, pardos e indígenas, e também alunos com deficiência, pudessem ingressar no Ensino Médio técnico integrado através de cotas. Metade das vagas em cada curso técnico são distribuídas entre as cotas e a outra metade são vagas de ampla concorrência. O ingresso até 2017 ocorreu por meio de processo de avaliação de entrada, conhecido como Vestibulinho.

Foram escolhidos como sujeitos de pesquisa, estudantes de sete campus do IFSP e também do Colégio Técnico da Unicamp (COTUCA). Esses alunos eram do segundo e terceiro ano do Ensino Médio e fizeram a prova no segundo semestre de 2017, uma ou duas semanas antes do Enem daquele ano. Para motivar a participação dos estudantes, eles foram lembrados da importância que tinha um simulado com problemas no estilo do Enem antes de fazer a prova, e também para ampliar ainda mais o espectro de alunos, diversos professores bonificaram alunos presentes no simulado com pontuação extra na nota do quarto bimestre, o que possibilitou uma grande adesão e uma amostra bem variada de sujeitos de pesquisa. A tabela 4 a seguir mostra a distribuição dos participantes de nossa pesquisa, por localização e sexo.

Tabela 4 - Distribuição dos respondentes em cada localidade pelo sexo

Localidade	Nº de Homens	Nº de Mulheres	Total
Campinas (COTUCA)	36	22	58
Bragança	51	32	83
Avaré	51	79	130
Capivari	26	31	57
Cubatão	8	7	15
Hortolândia	16	16	32
São João da Boa Vista	26	19	45
São Paulo	164	94	258
Total	378	310	688

Com essas duas estratégias, de simulado e pontuação para os presentes, somado a grande variedade regional e políticas de ação inclusiva, a população estudada se mostrou bem diversificada e adequada ao nosso estudo.

No total 691 estudantes fizeram o simulado, três deles não marcaram opção de sexo, masculino ou feminino, por isso esses três indivíduos ficaram fora das análises. A distribuição dos respondentes, com respeito à escolaridade do pai ou da mãe, apresentou relativamente poucos participantes que marcaram a opção "não sei" para a escolaridade de pelo menos um dos pais, 56 no total. Não é muito claro o que essa opção efetivamente significa, ou seja, o desconhecimento sobre a escolaridade de um dos pais ou a ausência dessa pessoa na vida do aluno.

Tabela 5 - Distribuição dos respondentes pela escolaridade do pai e da mãe

		Mãe						Total	
		NS	FI	FC	MI	MC	SC		PG
Pai	NS	9	3	1	4	16	6	3	42
	FI	0	24	11	7	21	9	1	73
	FC	0	3	3	1	19	5	1	32
	MI	0	3	5	4	21	13	3	49
	MC	3	9	15	24	138	62	16	267
	SC	1	6	7	2	37	89	25	167
	PG	1	0	0	2	10	17	20	50
	Total	14	48	42	44	262	201	69	680

Na tabela 5, acima, está a distribuição dos respondentes em termos das escolaridades dos pais: ensino fundamental incompleto (FI), ensino fundamental completo (FC), ensino médio incompleto (MI), ensino médio completo (MC), ensino superior completo (SC) e pós-graduação (PG). A resposta "não sei" está como NS.

Essa distribuição mostra uma grande variedade de escolaridades dos pais. Aproximadamente metade dos sujeitos de pesquisa (48%) possuem pai e mãe com ensino médio completo e/ou ensino superior completo, um número grande (45%) de estudantes possuem pais com no máximo ensino médio completo, e outra parcela dos respondentes possui um número considerável de pais com ensino superior completo e/ou pós graduação (22%). A moda dessa distribuição são pais e mães exclusivamente com ensino médio completo (20%).

Em relação a rendimento familiar total, o número de sujeitos que responderam "não sei" (NS) foi alto (23%), e novamente não é possível saber o real motivo do aluno ter marcado essa opção, se é por desconhecimento da renda ou se, por exemplo, é o temor de expor uma renda familiar grande. A maioria dos alunos está concentrada em famílias cuja renda familiar é de 2 a 6 salários mínimos, como mostra a tabela 6 abaixo.

Tabela 6 - Distribuição dos respondentes pela renda familiar total

Rendimento da família em Salários Mínimos						
NS	0 a 2	2 a 4	4 a 6	6 a 8	8 a 10	>10
157	97	171	139	65	21	31

Os sujeitos de pesquisa corresponderam a nossa previsão de contemplar uma boa diversidade do espectro socioeconômico, a grande quantidade de participantes da pesquisa (688) possibilitou esse efeito. Além disso, não ocorreu uma distorção muito grande entre o número de homens (55%) e mulheres (45%), muito embora os cursos de ensino técnico do IFSP e do COTUCA tenham cunho predominantemente masculino.

4.2. Instrumento de Avaliação

Para os propósitos desse trabalho, selecionamos inicialmente 28 itens de matemática da prova do Enem de 2010 a 2015 que apresentassem duas características importantes para o nosso estudo: fossem passíveis de modificação no enunciado com a finalidade de reduzir a carga cognitiva estranha do problema; e que os itens tivessem uma taxa de acerto considerável, pelo menos 30%. Os itens de matemática da prova do Enem de 2009 não foram selecionados apenas por considerarmos que a taxa de acerto desses itens em média foram baixas, e a partir dessa primeira aplicação do Enem reestruturado, os responsáveis pela prova de matemática talvez tenham utilizado essa expertise nas provas dos anos consequentes para criar itens mais adequados, ou seja, em todo o espectro de dificuldade.

A utilização dos microdados do Enem possibilita verificar resultados importantes da Teoria Clássica dos Testes (TCT), particularmente a dificuldade do item, que nada mais é do que a taxa de acerto dos candidatos para aquele item em questão. No Anexo III se encontram algumas medidas e discussões sobre a TCT, incluindo o Alfa de Crombach. Com essa informação em mãos, o número de itens candidatos à nossa pesquisa nas seis provas de matemática (270 itens) foi reduzido drasticamente. Foram pré

selecionados itens cuja dificuldade variou entre 0,33 e 0,65 pelos motivos já apontados na primeira subseção desse capítulo. Itens com uma dificuldade muito alta ou muito baixa são ruins (RABELO, 2013) no sentido de não abrir o espectro de notas (discriminação baixa). O Enem possui pouquíssimos itens com taxa de acerto maior que 0,5, mesmo com um recorte bem particular de candidatos que participaram dos dois dias de prova e que também não zeraram a redação. Após uma inspeção minuciosa nesses itens, foram selecionados 28 itens que contemplavam as características desejadas.

Utilizamos as mesmas hipóteses que o do trabalho de Gillmore, Poggio e Embretson (2015), de que a redução de carga cognitiva estranha deve possuir impacto no desempenho dos candidatos em cada item e, portanto, no teste como um todo e também na validade do instrumento de avaliação. Utilizamos as mesmas estratégias de redução de carga cognitiva dos autores, que está presente na Tabela 8 do capítulo anterior.

Em cada item, para a tentativa da redução de carga cognitiva estranha, pelo menos uma das estratégias apresentadas foi utilizada nos itens de nosso instrumento de avaliação. A única estratégia que não utilizamos foi a de Sequenciamento, muito embora em um item mudamos a ordem da frase no comando, mas ele continuou como a última parte do item antes das alternativas. Alguns itens permitiram a aplicação de mais de uma estratégia. Segundo o trabalho de Gillmore, Poggio e Embretson, as técnicas de Sinalização, Estética e Extirpar foram as que se mostraram mais eficazes na melhoria do desempenho dos sujeitos de pesquisa, nas condições de pesquisa desses autores.

Consideramos, assim como os autores, que é muito difícil afirmar que uma técnica é universalmente melhor que outra, já que cada item é um conjunto de diversas variáveis, como o corpo da questão, que pode conter ou não contextualização e em diversos níveis, um comando que pode exigir uma ou duas informações presentes em lugares próximos ou distantes no corpo, e geralmente um construto a ser avaliado. No trabalho dos autores, dos 15 itens, oito (8) apresentaram uma distância de Cohen não desprezível (pelo menos pequena) em favor daqueles itens com carga cognitiva estranha reduzida enquanto dois (2) itens tiveram uma distância de Cohen não desprezível em favor dos itens originais. Ainda no trabalho de Gillmore (2010), com respeito às duas provas inteiras, a constituída apenas por itens originais e a outra que apresenta apenas itens com carga cognitiva reduzida, os alunos que fizeram a prova com itens modificadas foram melhor, com uma distância de Cohen de 0,37, que é um valor considerável (moderado).

Utilizando como referência alguns exemplos de redução de carga cognitiva estranha no trabalho de Gillmore, Poggio e Embretson (2010), fizemos as modificações nos 28 itens que foram selecionados, e após algumas considerações sobre pertinência e adequação dos itens acabamos por utilizar 24 itens. Finalmente, para saber se a modificação da carga cognitiva estranha dos itens não mudou a estrutura e o construto avaliado, dois alunos do quinto semestre da licenciatura em matemática do Instituto Federal de São Paulo, campus de São Paulo, de maneira independente, resolveram esses 24 itens modificados, marcando as alternativas que consideravam corretas e anotando eventuais comentários sobre o corpo da questão, o comando, valores numéricos, alternativas ou qualquer imprecisão que poderia afetar o desempenho dos estudantes.

Após os ajustes, os 24 itens modificados do Enem, com redução de carga cognitiva, foram considerados adequados. Todos os itens escolhidos e suas versões reduzidas podem ser encontradas nesse trabalho, dois itens no capítulo anterior, alguns no Capítulo 6 e o restante no Apêndice VII. As discussões sobre o tipo de construto pretendido em cada item, as técnicas utilizadas para redução de carga cognitiva estranha e os resultados da aplicação do item aos respondentes também são encontrados junto de cada item.

Para verificar o quanto o desempenho dos alunos difere nos itens originais do Enem e nos mesmos itens com redução de carga cognitiva pensamos em construir um instrumento de avaliação com os 24 itens originais (teste controle) e outro com os 24 itens de carga cognitiva reduzida (teste experimento). Desistimos dessa alternativa por considerarmos que uma avaliação contendo apenas os itens originais seria desgastante para os respondentes se comparado com a avaliação contendo apenas os itens modificados. Por esse motivo optamos por fazer dois extremos de prova, uma com $2/3$ de itens originais do Enem e $1/3$ de itens modificados, que chamamos de prova 3366A, e outra com $1/3$ de itens originais do Enem e $2/3$ de itens modificados, que chamamos de 6633B. Por um erro de construção, a prova 3366A ficou com sete itens modificados ao invés do correto, que seria oito.

Para tentarmos analisar o desempenho dos alunos em cada item, construímos duas provas espelho, 5050A e 5050B, onde cada uma tinha 50% de itens originais do Enem e 50% de itens com carga cognitiva reduzida. Cada item original do Enem na prova 5050A era reduzido na prova 5050B, e cada item reduzido na prova 5050A era um item original

do enem na prova 5050B. Os itens foram realocados nessas provas, de modo que nenhum aluno fez mais do que três itens originais ou modificados em sequência.

As denominações dessas provas estavam presentes em cada página da avaliação e do questionário que cada estudante recebeu. Para a apresentação e análise dos resultados, acreditamos que essa nomenclatura de cada prova não seja adequada, e por esse motivo daqui para frente denominaremos de ">Original" a prova 3366A, por possuir uma quantidade maior de itens originais e ">Reduzida" a prova 6633B, por possuir uma quantidade maior de itens reduzidos. As provas 5050A e 5050B terão a nomenclatura alterada para Igual_A e Igual_B.

A ordem dos itens nas quatro provas (>Orig, Igual_A, Igual_B e >Red) foi a mesma, o que diferencia cada prova é a classificação do item como original (do Enem) ou reduzido (carga cognitiva reduzida). Foi também tomado o cuidado para que os itens não estivessem desbalanceados com respeito à taxa de acerto (dificuldade), de modo que de quatro em quatro itens o aluno encontrasse pelo menos um item com taxa de acerto maior ou igual a 0,45 (média) e um item com taxa de acerto menor ou igual a 0,35 (difícil). Lembrando que essa é a taxa de acerto do item utilizando os microdados do Enem para o recorte de alunos concluintes do Ensino Médio, que compareceram nos dois dias de prova e não zeraram a redação. A média da taxa de acerto nos 24 itens do nosso instrumento de avaliação é de 0,42.

Para que os alunos estivessem em condições parecidas com as da prova do Enem, o tempo médio que utilizamos por item foi de 3 minutos, e, portanto, para a prova toda foi estipulado um tempo de 72 minutos. Adicionamos 3 minutos para o preenchimento do questionário socioeconômico, que se encontra no **Apêndice II**, e no total os alunos tiveram então 75 minutos (1h15min) para resolver a prova e preencher o questionário.

Como esse trabalho trata do experimento com pessoas, particularmente com muitos estudantes menores de 18 anos, o procedimento de submeter o projeto a um comitê de ética foi executado, e os termos de consentimento e assentimento (**Apêndice I**) foram preenchidos pelos próprios alunos e pelos seus responsáveis, quando estes eram menores de idade.

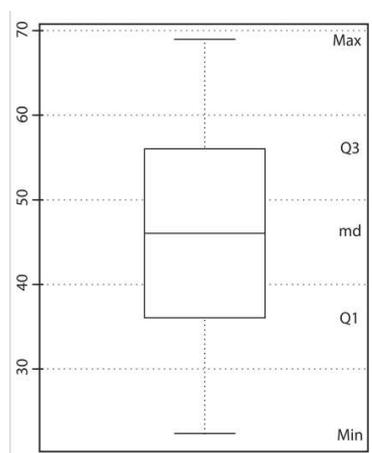
4.3. Métodos Estatísticos - panorama

Para esse trabalho, a parte quantitativa da pesquisa assumiu um papel importante de apontar os caminhos, corretos e incorretos, frente às nossas hipóteses. O uso e tratamento de dados, de modo a tomar as melhores decisões frente às indagações e perguntas de pesquisa, são o cerne da estatística. Para Guimarães (2008), a estatística é

[...] um conjunto de técnicas de análise de dados, cientificamente formuladas, aplicáveis a quase todas as áreas do conhecimento que nos auxiliam no processo de tomada de decisão. É a Ciência que estuda os processos de coleta, organização, análise e interpretação de dados relevantes e referentes a uma área particular de investigação.

A estatística descritiva é a parte da estatística que mais está presente no cotidiano da sociedade, sua tarefa é "resumir as principais características de um conjunto de dados por meio de tabelas, gráficos e resumos numéricos" (GUIMARÃES, 2008, p. 12). Como os dados brutos em geral se tornam de difícil compreensão, a tarefa de agrupar, categorizar e apresentar de maneira mais compreensiva, se torna muito importante. Utilizaremos predominantemente os conceitos numéricos de média e desvio padrão nesse trabalho.

Figura 10 - Exemplo de Boxplot



Fonte: Guimarães, 2008

Para a apresentação de dados em forma gráfica utilizaremos histogramas e o diagrama de caixa, também conhecido por Box-Plot, que é "uma representação gráfica de valores, conhecidos como resumo de 5 números (...): menor valor (Min), primeiro quartil (Q1), a Mediana (Md), o terceiro quartil (Q3) e o maior valor (Max)." (GUIMARÃES, 2008, p. 39).

Os quartis, Q1, Q2 (ou md) e Q3 são os valores de corte, ao se colocar os dados em ordem crescente, e que dividem a distribuição em quatro partes. O Box-Plot possui uma vantagem de possibilitar a visualização de medidas de tendência central, variabilidade e valores extremos de maneira conjunta.

Outro aspecto importante da nossa pesquisa está na utilização da chamada estatística inferencial, que essencialmente nos oferece um conjunto de técnicas e tomadas de decisão em situações onde não podemos ou não conseguimos trabalhar com toda a população. A estatística inferencial pode ser descrita como um "conjunto de técnicas e procedimentos que permitem dar ao pesquisador um grau de confiabilidade nas afirmações que faz para a população, baseadas nos resultados das amostras" (GUIMARÃES, 2008, p. 13).

Os testes de hipótese entram nessa classe de métodos estatísticos, onde em geral escolhemos uma hipótese, chamada de hipótese nula, e sua complementar, chamada de hipótese alternativa. Escolhemos um chamado nível de significância, que é um erro intrínseco que estamos dispostos a aceitar, e qualquer valor (p-valor) menor que esse nível de significância, nos faz aceitar a hipótese nula.

Para trabalharmos com um número grande de respondentes em um teste, há ferramentas estatísticas que auxiliam a medir e qualificar diferenças de desempenho entre dois grupos. Neste trabalho, estamos interessados em medir e classificar as diferenças de desempenho entre rapazes e moças. Para tal propósito utilizamos a distância estatística de Cohen, que pode ser determinada pela expressão:

$$d_c = \frac{m_A - m_B}{dp_{AB}}$$

onde m_A e m_B são as médias dos grupos A e B, respectivamente e dp_{AB} o desvio padrão do grupo formado pela união dos subgrupos A e B. Com base no valor da distância estatística, podemos qualificar a relevância da diferença de desempenho entre os grupos de acordo com a seguinte escala: $d_c \leq 0,10$ trivial; $0,10 < d_c \leq 0,35$ pequena; $0,35 \leq d_c < 0,65$ moderada e $d_c \geq 0,65$ grandes (COHEN, 1988).

Essa expressão para distância de Cohen tem a mesma estrutura de testes de hipótese para a média, e essencialmente tenta medir a distância entre dois conjuntos de dados baseados em número de desvio padrão. Por exemplo, um valor de $d_c = 0,2$ significa

que as distribuições são 0,2 desvio padrão distantes entre si. Um valor positivo significa que a média do grupo A é maior que a do grupo B, um valor negativo significa o oposto.

Por último, para tratar estatisticamente possíveis desigualdades ou igualdades entre classes, utilizaremos a Razão de Chance (RC). A razão de chance é um indicador da possibilidade de sucesso do grupo focal em relação ao grupo de referência.

Tabela 7 - Grupo focal e grupo de referência com candidatos

	Candidatos com sucesso	Candidatos sem sucesso
Grupo focal	A	B
Grupo de referência	C	D

A partir dessa tabela anterior, podemos construir uma tabela das chances de sucesso ou de insucesso para cada um dos grupos, conforme nos mostra a tabela abaixo.

Tabela 8 - Grupo focal e grupo de referência com a chance

	Chance de sucesso	Chance de insucesso
Grupo focal	$A/(A+B)$	$B/(A+B)$
Grupo de referência	$C/(C+D)$	$D/(C+D)$

Na tabela anterior, as chances de sucesso ou de insucesso são calculadas a partir do número de candidatos com sucesso ou sem sucesso e o número total de candidatos em cada grupo. A partir das chances de sucesso, calculamos a razão entre a chance de sucesso e a de insucesso para cada grupo, dividindo o grupo focal pelo grupo de referência. Daí o nome “razão de chances”, ou do inglês, *odds ratio*.

$$RC = \frac{\frac{A/(A+B)}{B/(A+B)}}{\frac{C/(C+D)}{D/(C+D)}} = \frac{\frac{A}{B}}{\frac{C}{D}} = \frac{A \cdot D}{B \cdot C}$$

Uma razão de chance igual a um (1) implica na ausência de diferença de desempenho entre os grupos, enquanto um valor positivo indica uma maior chance de melhor desempenho do grupo focal em relação ao grupo de referência. As categorias de referência geralmente são as seguintes:

- RC entre 0,8 e 1,25 - considerada trivial, sem importância;
- RC entre 0,5 e 0,8 ou entre 1,25 e 2,0 - baixo impacto;
- RC entre 0,33 e 0,5 ou entre 2,0 e 3,0 - médio impacto;
- RC abaixo de 0,33 ou acima de 3,0 - alto impacto

Esses valores de referência, juntamente com o número de elementos na amostra, podem acarretar uma diferença significativa entre os grupos, essa análise é feita através de testes de hipótese específicos para razão de chance.

5. Resultados

Nesse capítulo serão apresentados os resultados do trabalho envolvendo os quatro modelos de provas e os itens apresentados no capítulo anterior. As primeiras análises serão divididas em três seções: Estatística Descritiva e Teoria Clássica dos Testes, que possibilita o retorno de um olhar abrangente sobre a prova e os itens com dados de interesse em relação às variáveis estudadas; Distância de Cohen, como uma primeira medida largamente utilizada em avaliações para quantizar a diferença entre dois conjuntos de dados; e Testes de Hipótese, onde tomaremos todos os resultados de interesse levantados pelas medidas estatísticas anteriores e analisaremos com rigor estatístico a pergunta de pesquisa as questões direcionadas aos objetivos desse trabalho.

Lembramos que estamos particularmente interessados em analisar como a redução de carga cognitiva de itens do Enem impacta no desempenho dos respondentes. É feita uma análise sobre o desempenho dos estudantes em cada uma das provas e também nos itens individualmente, sejam eles Originais ou Reduzidos. Uma próxima análise contempla o impacto da redução de carga cognitiva no desempenho por sexo, tanto nas provas quanto nos itens. Por último, analisaremos as outras variáveis socioeconômicas que temos acesso: a escolaridade dos pais e renda familiar.

Na quarta seção olharemos especificamente os itens. Considerando todas as análises estatísticas feitas, alguns itens apresentam particularidades com respeito a redução de carga cognitiva que estão associados à maneira como o item foi modificado, em todos os itens lançamos um olhar qualitativo sobre a teoria da carga cognitiva e as técnicas de redução de carga cognitiva estranha.

5.1. Estatística Descritiva e Teoria Clássica dos Testes

Aqui serão expostos os resultados das notas, das provas e dos itens, juntamente com os efeitos das variáveis socioeconômicas de interesse. Consideramos que a inclusão dos alunos do segundo ano do ensino médio nos sujeitos de pesquisa não deve ter influência no desempenho desses alunos, uma vez que os construtos matemáticos

presentes nesses 24 itens não exigem conteúdos avançados e foram contemplados durante o primeiro ano do ensino médio e os três bimestres do segundo ano do ensino médio.

Na tabela 9 a seguir são apresentados a média e o desvio padrão do desempenho dos estudantes em cada item das quatro provas, >Orig, Igual_A, Igual_B e >Red, além de todas as taxas de acerto de cada item original do Enem, no recorte de alunos presentes nos dois dias de prova e que não zeraram a redação. Os dados das taxas de acerto foram obtidos através dos microdados do Enem. As células laranjas correspondem aos itens originais e as células verdes correspondem aos itens reduzidos. O número de respondentes em cada prova (N) foi próximo de 170.

Tabela 9 - Taxa de acerto e desvio padrão em cada item por tipo de prova

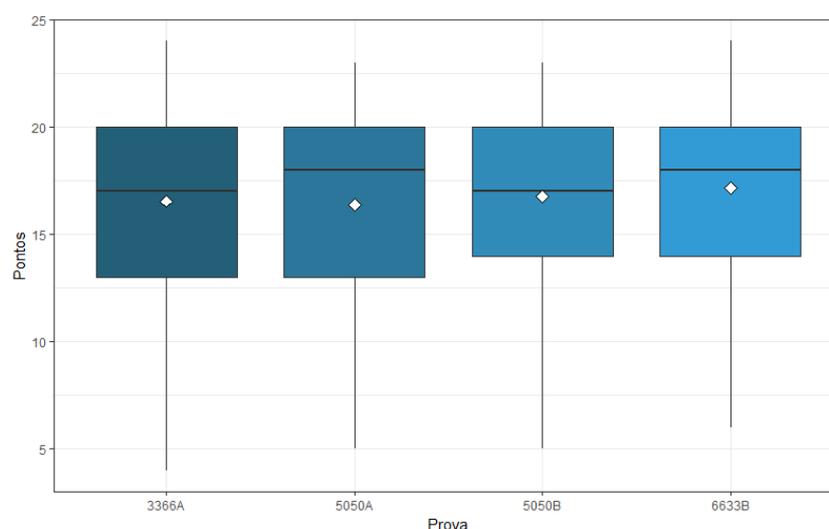
	Taxa de Acerto ENEM	>Orig N = 172		Igual_A N = 165		Igual_B N = 171		>Red N = 170	
		Média	Dp	Média	Dp	Média	Dp	Média	Dp
Q1	0,58	0,84	0,37	0,81	0,39	0,78	0,41	0,85	0,36
Q2	0,65	0,89	0,31	0,93	0,25	0,94	0,24	0,95	0,21
Q3	0,33	0,61	0,49	0,63	0,48	0,61	0,49	0,61	0,49
Q4	0,35	0,51	0,50	0,59	0,49	0,56	0,50	0,64	0,48
Q5	0,48	0,91	0,29	0,79	0,41	0,95	0,21	0,91	0,28
Q6	0,36	0,70	0,46	0,66	0,47	0,77	0,42	0,67	0,47
Q7	0,33	0,65	0,48	0,69	0,46	0,65	0,48	0,68	0,47
Q8	0,45	0,79	0,41	0,80	0,40	0,79	0,41	0,78	0,41
Q9	0,59	0,91	0,29	0,87	0,34	0,91	0,28	0,94	0,25
Q10	0,35	0,72	0,45	0,86	0,35	0,69	0,46	0,82	0,39
Q11	0,37	0,88	0,32	0,79	0,41	0,90	0,30	0,89	0,31
Q12	0,52	0,76	0,43	0,77	0,42	0,83	0,38	0,81	0,39
Q13	0,35	0,41	0,49	0,39	0,49	0,40	0,49	0,40	0,49
Q14	0,33	0,52	0,50	0,46	0,50	0,51	0,50	0,57	0,50
Q15	0,40	0,77	0,42	0,73	0,44	0,76	0,43	0,80	0,40
Q16	0,43	0,67	0,47	0,64	0,48	0,66	0,47	0,63	0,48
Q17	0,43	0,57	0,50	0,61	0,49	0,52	0,50	0,67	0,47
Q18	0,35	0,66	0,47	0,47	0,50	0,57	0,50	0,49	0,50
Q19	0,34	0,58	0,49	0,62	0,49	0,65	0,48	0,62	0,49
Q20	0,38	0,42	0,49	0,55	0,50	0,46	0,50	0,56	0,50
Q21	0,43	0,72	0,45	0,72	0,45	0,71	0,45	0,76	0,43
Q22	0,33	0,69	0,46	0,66	0,47	0,73	0,44	0,69	0,46
Q23	0,33	0,59	0,49	0,49	0,50	0,57	0,50	0,61	0,49
Q24	0,63	0,80	0,40	0,86	0,35	0,82	0,38	0,82	0,38
Provas		16,57	2,15	16,39	2,17	16,74	2,13	17,17	2,11

O desempenho médio dos estudantes em cada prova pode ser considerado bom, oscilando de 16,57 (69% de acerto) na prova >Orig, que possui mais itens originais, a 17,17 (72% de acerto) na prova >Red, que possui mais itens reduzidos. Os sujeitos da

pesquisa foram consideravelmente melhor que os respondentes do Enem nesses itens, comparando com os resultados dos microdados. O desvio padrão do acerto de cada item oscilou entre 0,3 e 0,5, evidenciando a heterogeneidade dos respondentes. A sensação é de que as notas das provas Igual_A e Igual_B ficaram próximas, como era de se esperar, mas as notas das provas >Red e >Orig não ficaram tão distantes. Medidas estatísticas quantitativas serão aplicadas nas próximas seções para verificar esse efeito.

O gráfico de boxplot, na figura 11 a seguir, nos dá esse panorama. Análises mais profundas com teste de hipótese vão explorar o desempenho geral entre cada prova.

Figura 11 - Boxplot do escore em cada uma das provas



Esse escore das provas perto de 70% juntamente com essa variabilidade considerável de cada item traz consequências para algumas medidas da teoria clássica dos testes. Estamos particularmente interessados em dois valores, o Alfa de Cronbach e a discriminação do item em cada uma das quatro provas. O Alfa é mostrado na tabela 10 a seguir.

Tabela 10 - Alfa de Cronbach em cada uma das provas

	>Orig	Igual_A	Igual_B	>Red
Alfa de Cronbach	0,818	0,821	0,782	0,746

Esses valores de Alfa oscilam perto de 0,8, com exceção da prova >Red, que segundo a tabela de valores compilados de Alfa em Peterson (1994), possuem uma fiabilidade moderada a elevada. Já com relação a discriminação dos itens em cada uma das provas, a maioria desses itens possui uma discriminação maior que 0,3, que segundo

a literatura (EBEL; FRISBIE, 1986), possui boa qualidade. Todos esses dados estão na Estatística Clássica de Testes do **Apêndice IV**.

Para ter uma visão geral no desempenho de cada item original ou reduzido, foi feita a média de acerto de cada um dos itens com essas características e também a sua diferença, como mostra a tabela 11 a seguir.

Tabela 11 - Taxa de acerto médio em itens originais e reduzidos

	Reduzida	Original	Diferença				
Q1	0,81	0,82	-0,01	Q13	0,40	0,40	0,00
Q2	0,94	0,91	0,03	Q14	0,53	0,46	0,07
Q3	0,62	0,61	0,01	Q15	0,76	0,77	0,00
Q4	0,53	0,61	-0,08	Q16	0,64	0,65	-0,01
Q5	0,93	0,79	0,14	Q17	0,64	0,54	0,10
Q6	0,77	0,68	0,09	Q18	0,48	0,61	-0,14
Q7	0,69	0,66	0,04	Q19	0,61	0,62	-0,01
Q8	0,79	0,79	0,00	Q20	0,55	0,44	0,11
Q9	0,92	0,89	0,04	Q21	0,74	0,71	0,03
Q10	0,84	0,71	0,13	Q22	0,71	0,68	0,03
Q11	0,89	0,79	0,10	Q23	0,49	0,59	-0,10
Q12	0,78	0,83	-0,05	Q24	0,84	0,81	0,03
				Média	16,93	16,37	

Nessa primeira inspeção, em 14 itens os candidatos tiveram uma média de desempenho melhor nas versões reduzidas, enquanto em 7 ocorreu o inverso, os estudantes tiveram uma média melhor nas versões originais. A média de desempenho nos itens reduzidos (16,93) foi ligeiramente maior que a média de desempenho nos itens originais (16,37).

A análise do desempenho por sexo pode ser feita por prova e também por item, as tabelas de resultado geral se localizam no **Apêndice IV**, em que o F e M antes de cada prova denominam exclusivamente os candidatos do sexo feminino e masculino, respectivamente, que fizeram aquela prova. A notação de células laranjas para itens originais e células verdes para itens reduzidos continua a ser aplicada. O número N de elementos de amostra de cada sexo, por provas, está na tabela 12 a seguir.

Tabela 12 - Número de respondentes por sexo em cada prova

	>Orig	Igual_A	Igual_B	>Red
N	172	166	172	171
N feminino	74	74	82	70
N masculino	98	92	90	101

O resultado geral de desempenho, por sexo, em cada uma das quatro provas está colocado na tabela 13 a seguir.

Tabela 13 - Escore de homens e mulheres em cada prova

	>Orig	Igual_A	Igual_B	>Red
Média dos Homens	17,66	17,30	17,76	18,03
Média das Mulheres	15,07	15,30	15,60	15,93

Olhando os valores das médias das provas >Orig e >Red, que são as provas mais distintas, uma vez que a primeira possui 16 itens originais e 8 itens reduzidos e a segunda possui 16 itens reduzidos e 8 itens originais, temos a sensação de que uma prova com mais itens reduzidos beneficia mais as mulheres. A diferença de pontuação dos homens da prova >Orig para a prova >Red é de 0,37 enquanto para as mulheres essa diferença foi de 0,86. Esse efeito será mais explorado nas análises das próximas seções.

Tabela 14 - Diferença de desempenho em cada item categorizado por sexo e versão do item

	Reduzida F	Original F	Diferença	Reduzida M	Original M	Diferença
Q1	0,78	0,75	0,04	0,83	0,88	-0,06
Q2	0,93	0,87	0,06	0,96	0,94	0,02
Q3	0,54	0,54	0,00	0,68	0,67	0,01
Q4	0,44	0,54	-0,10	0,61	0,67	-0,05
Q5	0,91	0,78	0,13	0,93	0,79	0,14
Q6	0,78	0,60	0,18	0,77	0,74	0,02
Q7	0,65	0,55	0,09	0,73	0,74	-0,01
Q8	0,79	0,80	0,00	0,79	0,79	0,00
Q9	0,89	0,85	0,04	0,95	0,92	0,04
Q10	0,85	0,62	0,22	0,84	0,78	0,06
Q11	0,89	0,74	0,14	0,90	0,83	0,07
Q12	0,71	0,82	-0,11	0,83	0,83	0,00
Q13	0,37	0,47	-0,09	0,42	0,35	0,07
Q14	0,50	0,35	0,15	0,55	0,55	0,00
Q15	0,65	0,67	-0,02	0,87	0,84	0,03
Q16	0,62	0,61	0,02	0,65	0,69	-0,04
Q17	0,52	0,47	0,05	0,73	0,60	0,13
Q18	0,47	0,54	-0,07	0,49	0,67	-0,19
Q19	0,55	0,58	-0,03	0,66	0,65	0,01
Q20	0,51	0,42	0,10	0,59	0,46	0,12
Q21	0,70	0,64	0,06	0,77	0,77	0,00
Q22	0,65	0,55	0,09	0,76	0,77	-0,01
Q23	0,39	0,50	-0,10	0,57	0,66	-0,10
Q24	0,78	0,76	0,02	0,89	0,85	0,04

Esse questionamento de que os itens reduzidos podem ter mais impacto no desempenho das mulheres pode ser verificado se tomarmos as respostas de todos os itens

reduzidos e as respostas de todos os itens originais, e olharmos o desempenho por sexo, juntamente com a diferença de pontuação de cada item, entre reduzidos e originais, como na tabela 12.

Aqui a diferença positiva significa que no item os estudantes, em média, acertaram mais a versão reduzida do que o item original, ou seja, a redução de carga cognitiva beneficiou o desempenho médio dos respondentes naquele item. Olhando individualmente os sexos notamos que majoritariamente as diferenças são positivas, ou seja, a redução de carga cognitiva está melhorando o desempenho de ambos os sexos na prova inteira.

Comparando as diferenças entre os itens reduzidos e originais, entre os dois sexos, temos a sensação de que a redução de carga cognitiva possa ter acarretado alguns efeitos. Em alguns itens particulares as mulheres tiveram um salto de desempenho maior que os homens, como são os casos da Q6, Q7, Q10, Q11 e Q14. Em dois itens, Q12 e Q13, o desempenho das mulheres piorou, se comparado ao desempenho dos homens, entre os itens originais e reduzidos. Em três itens, Q1, Q7 e Q22, a redução de carga cognitiva melhorou o desempenho das mulheres enquanto que o desempenho dos homens se manteve estável ou diminuiu. Essa primeira inspeção será confrontada com testes de hipótese, para ver se esses efeitos são significativos.

Nessa análise preliminar existe uma sensação de que a redução de carga cognitiva estranha produz um impacto maior na melhoria do desempenho das mulheres do que no desempenho dos homens. Esses efeitos serão melhor explorados nas próximas seções, juntamente com uma análise sobre esses itens que possuem maior discrepância entre a versão original e reduzida.

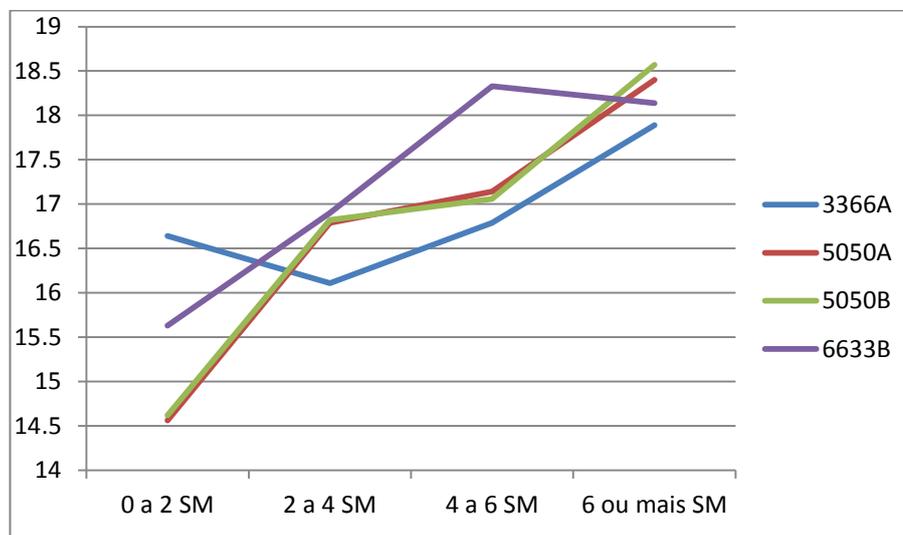
Como vimos na tabela de número de estudantes por renda, muitos estudantes (23%) não declararam a renda familiar, e o número de pessoas cuja renda familiar é maior ou igual a 8 salários mínimos (SM) é muito pequeno, por esse motivo, para analisar o desempenho dos alunos em cada prova, agrupamos as categorias 6 a 8 SM, 8 a 10 SM e maior que 10 SM como uma só: 6 ou mais SM. Os resultados da tabela 15 a seguir mostram o desempenho dos estudantes em cada uma das provas nas categorias de renda familiar.

Tabela 15 - Escore médio em cada prova por renda familiar

	0 a 2 SM	2 a 4 SM	4 a 6 SM	6 ou mais SM
>Orig	16,64	16,11	16,79	17,89
N	14	37	39	37
Igual_A	14,56	16,79	17,14	18,4
N	27	47	35	15
Igual_B	14,62	16,82	17,06	18,57
N	29	39	35	28
>Red	15,63	16,9	18,33	18,14
N	27	48	30	36

Em média vemos que quanto maior a renda da família do estudante, maior é sua pontuação média. O desempenho dos alunos entre os extremos desse espectro econômico é grande em todas as quatro provas. As provas Igual_A e Igual_B possuem resultados bem parecidos para todas as categorias, enquanto a prova >Orig, com mais itens originais, teve uma média de acertos consideravelmente menor para famílias com renda entre 2 a 6 SM, se comparados a prova >Red, que possui mais itens reduzidos. Para a categoria de estudantes cuja família tem uma renda mensal superior a 6 SM a diferença entre essas duas provas foi pequena. A figura 11 a seguir sumariza os desempenhos por prova.

Figura 11 - Escore médio em cada prova por renda familiar

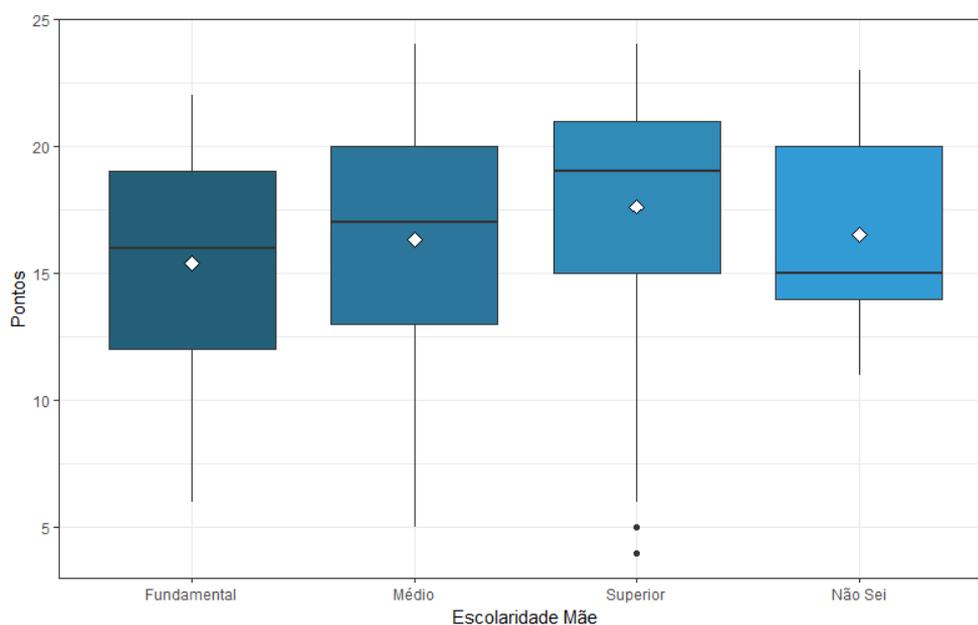


A quantidade de alunos em cada um dos grupos é próxima, muito embora não seja tão grande, a questão da confiabilidade dos resultados passa pelo número pequeno de elementos por grupo e pela possibilidade do aluno não saber ou colocar outro valor na renda mensal familiar. Por esse motivo, não será feita uma análise sobre o desempenho dos estudantes em itens reduzidos ou originais, quando variamos a renda familiar.

Uma situação muito parecida ocorre quando exploramos o desempenho dos respondentes em termos das escolaridades dos pais. Existem muitas categorias, como mostrado na tabela da seção anterior, e é possível separar a categoria escolaridade em três: ensino fundamental incompleto e completo (EF); ensino médio incompleto e completo (EM); e superior e pós-graduação (SUP). Aqui pouco estudantes assinalaram a opção "não sei", 9 (1%) marcaram tanto na escolaridade do pai quanto para a da mãe, 33 (5%) assinalaram para a escolaridade do pai exclusivamente, enquanto apenas 5 marcaram exclusivamente para a escolaridade da mãe. Avaliamos que os estudantes marcam para essa variável uma informação bem mais confiável do que a renda média mensal, uma vez que os alunos e alunas possuem mais convivência com sua família do que com a finança familiar.

As figuras 12 e 13, e a Tabela 14 a seguir mostram o desempenho dos estudantes estratificados exclusivamente pela escolaridade da mãe ou do pai em cada uma das quatro provas, além do número N de elementos na amostra.

Figura 12 - Score médio dos alunos em função da escolaridade da mãe



Alguns valores da Tabela 16 podem evidenciar o quanto a variável escolaridade individual, de um dos pais, necessita uma análise cuidadosa. Tomemos por exemplo o desempenho dos alunos na prova >Red cujos pais possuem no máximo ensino fundamental. O valor médio de desempenho, que é 17,26, dá a sensação de ser alto se comparado à mesma estratificação de escolaridade nas outras três provas, ou até comparando com pais ou mães com ensino médio completo, no máximo, em todas as

quatro provas. Essa informação não é conflitante, dois efeitos são responsáveis por esse valor: o primeiro é que o número de elementos nesse grupo é pequeno (23) e calha que entre esses 23 estudantes cujo pai possui no máximo ensino fundamental completo, a mãe possui pelo menos o ensino médio completo para 11 estudantes.

Figura 13 - Escore médio dos alunos em função da escolaridade do pai.

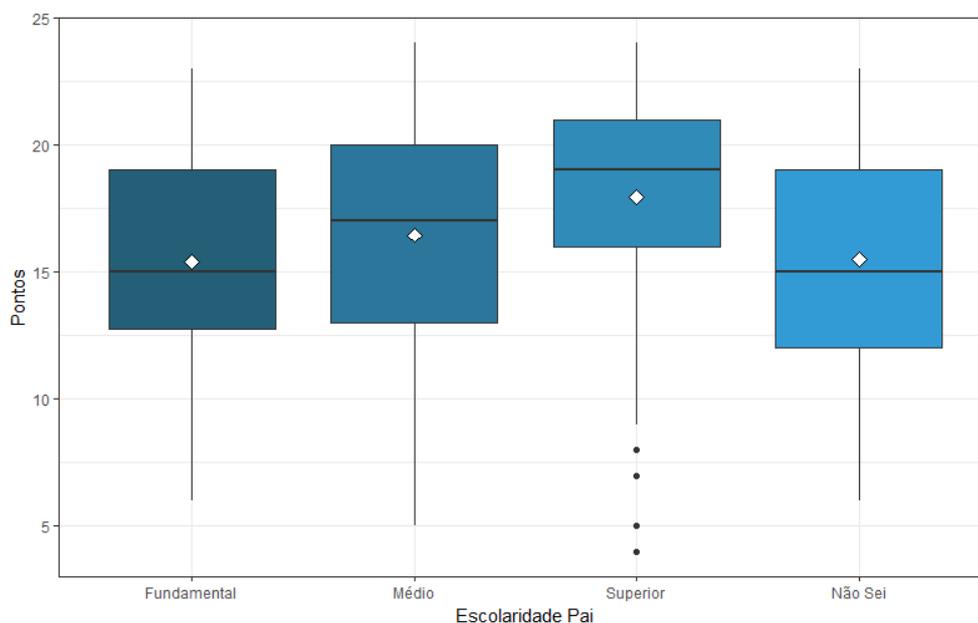


Tabela 16 - Escore médio em cada prova por escolaridade da mãe e do pai

	>Orig	Igual_ A	Igual_ B	>Red		>Orig	Igual_ A	Igual_ B	>Red
Mãe SUP	17,71	16,8	17,06	18,74	Pai SUP	17,68	17,31	18,63	18,38
N	80	59	70	62	N	62	58	56	42
Mãe EM	15,65	16,04	16,76	16,65	Pai EM	16,68	16,11	16,09	16,59
N	68	79	78	80	N	80	72	66	97
Mãe EF	14,75	16,48	15,52	14,96	Pai EF	13,55	15,36	15,63	17,26
N	20	25	21	24	N	22	25	35	23

Podemos continuar olhando a escolaridade dos pais e ver como ela impacta no desempenho dos respondentes, mas para que o número N de elementos da amostra não seja pequeno foram consideradas apenas duas classes: pai ou mãe com pelo menos ensino superior (Sup); e pai e mãe com no máximo ensino médio (Med).

Tabela 17 - Escore médio em cada prova por categoria de escolaridade familiar

	>Orig	Igual_ A	Igual_ B	>Red
Sup	17,29	17,10	17,53	18,51
N	97	80	88	72
Med	15,50	15,77	15,88	16,19
N	74	86	84	99

Como mostra a tabela 17, o desempenho dos alunos, tanto do grupo Sup quanto do Med, foi parecido entre as provas Igual_A e Igual_B. As provas opostas, >Orig que possui mais itens originais e >Red, que possui mais itens reduzidos, tiveram médias diferentes, com os estudantes apresentando um melhor desempenho na prova >Red. Os estudantes cujos pais estão na categoria Sup apresentaram uma variação média de desempenho entre as provas >Orig e >Red maior do que os estudantes da categoria Med. Em um primeiro instante podemos imaginar que a redução de carga cognitiva na prova está beneficiando mais alunos e alunas cujos pais possuem uma melhor formação escolar.

As variáveis escolaridade dos pais e renda estão intimamente ligadas, uma vez que o rendimento médio de uma pessoa possui correlação com seus anos de estudo e titulação. Análises envolvendo desempenho dos estudantes através da escolaridade do pai, da mãe, e da renda média familiar, não são tão imediatas, uma vez que essas variáveis possuem uma interdependência grande.

Separar os alunos e alunas em grupos cada vez mais particulares pode ser uma solução para uma análise mais detalhada, por exemplo, desempenho de estudantes do sexo feminino cuja mãe possui uma escolaridade maior que a do pai e possuam renda familiar igual ou superior a 4 salários mínimos. Esse procedimento esbarra em números de elemento de amostra que podem ser zero ou bem próximos disso, que provavelmente não permitem a convergência de testes estatísticos mais precisos.

As análises estatísticas mais elaboradas sobre o desempenho em itens originais e em itens reduzidos serão realizadas sobre a variável sexo, e não com as demais variáveis socioeconômicas. As distâncias de Cohen para algumas variáveis socioeconômicas serão estimadas na busca por quantificar a distância entre grupos, e desse modo começamos a compreender a dimensão da distância entre essas variáveis.

5.2. Distância de Cohen

Após as análises estatísticas descritivas, que permitem uma visualização de características de interesse para uma próxima análise mais profunda, será calculada a distância de Cohen entre alguns grupos para procurarmos quantificar as diferenças, sob o ponto de vista da significância estatística e da sua intensidade.

As primeiras medidas de distância de Cohen (d_c) foram realizadas entre as provas, como fizemos na estatística descritiva: entre as provas >Orig e >Red; entre as provas espelho com mesmo número de itens originais e reduzidos; e finalmente entre as provas constituídas apenas por itens Reduzidos e itens Originais. A tabela 18 a seguir mostra os resultados.

Tabela 18 - Distância de Cohen entre os tipos de prova

	<u>Igual_A - Igual_B</u>	<u>>Red - >Orig</u>	<u>Reduzida-Original</u>
d_c	0,07	0,14	0,19

Era esperado que as provas Igual_A e Igual_B não apresentassem diferenças estatísticas significativas, como a estatística descritiva já nos tinha indicado. Uma distância de Cohen com valor até 0,1 indica uma variação sem significância estatística, com uma variação irrelevante entre os escores. Para as provas >Red e >Orig, a d_c positiva significa que os estudantes acertaram mais itens na prova com maior número de itens reduzidos, e essa distância de Cohen ficou em 0,14 que é uma diferença pequena, mas estatisticamente significativa. A diferença entre as provas Reduzida e Original apresentou uma d_c igual a 0,18 que também é uma diferença pequena, em favor dos itens reduzidos.

Esses valores encontrados estão confirmando as análises estatísticas iniciais, com a vantagem de que a distância de Cohen conseguiu quantificar diferenças entre grupos. Os itens reduzidos possuem, de fato, um papel importante em um melhor desempenho dos estudantes.

A próxima análise se refere à análise do efeito dos itens reduzidos nos respondentes do sexo feminino. Serão calculadas as distâncias de Cohen entre os grupos masculino e feminino para cada uma das provas. Valores positivos de d_c significam escore masculino maior que escore feminino. Os dados estão na Tabela 19 a seguir.

Tabela 19 - Distância de Cohen entre os sexos para cada prova

<u>Prova</u>	<u>>Orig</u>	<u>Igual_A</u>	<u>Igual_B</u>	<u>>Red</u>
d_c	0,56	0,44	0,51	0,53

Uma d_c entre 0,35 e 0,65 é considerada moderada, mas é importante apontar que os valores estão na casa de 0,5 e que esse valor nos indica que os homens em média estão

meio desvio padrão na frente das mulheres nessas provas, que corresponde a uma grande vantagem, por exemplo, em processos seletivos que utilizem esse ranqueamento.

Para as provas com itens exclusivamente Reduzidos e Originais não é possível encontrar o desvio padrão da prova em si, uma vez que essas provas são reconstituídas, elas não existem enquanto uma prova única. O que é possível é estimar o desvio padrão com base nos desvios padrões das outras provas, cujo menor é 4,0 e o maior é 4,7. Nessas circunstâncias, para a prova Reduzida, a distância de Cohen entre homens e mulheres está entre 0,37 e 0,43, e para a prova Original a distância de Cohen entre homens e mulheres está entre 0,52 e 0,61. Percebe-se que para os itens Originais, não só o desempenho das mulheres foi bem inferior ao dos homens, como mostrado na seção anterior, mas agora a distância entre homens e mulheres é bem grande se comparado aos itens Reduzidos, cuja distância de Cohen é a menor de todas, comparando entre todas as outras provas, muito embora seja um valor superior a 0,35, o que nos diz que é um valor ainda bem grande.

Para verificar se a redução de carga cognitiva nos itens beneficia mais mulheres do que homens, podemos fazer uma comparação entre as provas >Red e >Orig, entre cada um dos grupos compostos exclusivamente por Homens e Mulheres. A seguir temos, na Tabela 20, as distâncias de Cohen para essas categorias. Uma comparação entre os grupos Igual_A e Igual_B também está contemplada.

Tabela 20 - Distância de Cohen entre os sexos entre as provas

	Homens	Mulheres
>Red - >Orig	0,10	0,19
Igual_A - Igual_B	-0,11	-0,07

A distância de Cohen entre as Mulheres que fizeram a prova >Red, que possui mais itens reduzidos, e as que fizeram a prova >Orig, que possui mais itens originais, é de fato maior que a distância de Cohen entre os Homens para as mesmas provas. Na seção anterior vimos que a nota média das mulheres subiu consideravelmente, na comparação entre as provas >Red e >Orig. Essas duas informações combinadas nos fornecem indícios de as mulheres podem se beneficiar mais do que os homens em itens submetidos a redução de carga cognitiva, essa hipótese será testada na próxima seção.

Na questão socioeconômica, como visto na seção anterior, a questão de separar os alunos em muitas classes distintas pode refletir em resultados pouco confiáveis, por esse

motivo parece razoável efetuar análises com a separação de apenas duas classes. Com respeito à escolaridade, novamente faremos a análise de duas classes: pai ou mãe com pelo menos ensino superior; e pai e mãe com no máximo ensino médio. A distância de Cohen (d_c) para cada uma das provas entre esses dois grupos está na Tabela 21 abaixo.

Tabela 21 - Distância de Cohen entre os estratos escolares dos pais para cada prova

	>Orig	Igual_A	Igual_B	>Red
d_c	0,39	0,29	0,40	0,62

A prova >Orig possui mais itens originais do que reduzidos, e a distância entre esses dois grupos é de $d_c = 0,39$, que é moderada. Já para a prova >Red, que possui uma quantidade maior de itens reduzidos, entre as classes temos $d_c = 0,62$, que embora seja consideravelmente maior que 0,39, ainda é moderada. Essa primeira inspeção nos coloca uma hipótese de que a redução de carga cognitiva nos itens de matemática abre mais o abismo de desempenho entre estudantes cujos pais possuem mais instrução ou menos instrução.

Um contraponto a essa análise é perceber que para as provas Igual_A e Igual_B a distância de Cohen deveria ser parecida, mas obtivemos $d_c = 0,29$ para a prova Igual_A, que é uma diferença pequena, e $d_c = 0,40$ que é uma diferença moderada. Na mesma análise para a diferença de desempenho entre os sexos para essas mesmas provas, foram obtidos valores mais próximos.

Olhamos a Tabela 22, que mostra a distância de Cohen entre essas duas provas, para as classes de pais com no máximo ensino médio (med) e com pelo menos um pai com o ensino superior (sup).

Tabela 22 - Distância de Cohen entre os estratos escolares dos pais entre as provas

	med	sup
>Red - >Orig	0,16	0,30
Igual_A - Igual_B	0,03	0,10

Esses dados nos sugerem a mesma hipótese de que a redução de carga cognitiva beneficia mais estudantes cujos pais possuem nível de instrução mais elevado. Entre as provas Igual_A e Igual_B é esperado que não haja diferença significativa, assim como na distância de Cohen da tabela anterior.

Ainda na questão socioeconômica, podemos separar os estudantes em outros dois grupos, um cuja renda familiar é de no máximo quatro salários mínimos (inf) e o outro cuja renda familiar é maior do que quatro salários mínimos (sup). Esses grupos foram escolhidos por possuírem número de elementos considerável e próximo entre si. A Tabela 23 abaixo possui a média (M), o desvio padrão (DP) e o número de elementos do grupo (N) para cada um desses grupos, e a distância de Cohen (d_c) para cada uma das provas, entre esses dois grupos de renda.

Tabela 23 - Distância de Cohen entre os estratos socioeconômicos para cada prova

	>Orig	Igual_A	Igual_B	>Red
M sup	17,33	17,52	17,73	18,23
DP sup	4,34	4,25	4,07	3,28
N	76	50	63	66
M inf	16,25	15,97	15,88	16,44
DP inf	4,62	4,88	4,11	4,13
N	51	74	68	75
d_c	0,24	0,33	0,45	0,48

A distância de Cohen dentro dos grupos aumenta, de $d_c = 0,24$ para a prova >Orig, que possui mais itens originais, para $d_c = 0,48$ para a prova >Red, que possui mais itens reduzidos. Nas provas Igual_A e Igual_B temos $d_c = 0,33$ e $d_c = 0,45$, respectivamente. Esse resultado nos propõe que os alunos provenientes de família com renda maior se beneficiam mais dos itens reduzidos do que alunos cuja família apresente uma renda familiar menor. Para esse olhar calcularemos a distância de Cohen entre as provas >Orig e >Red, e também entre as provas Igual_A e Igual_B, como mostra a Tabela 24 a seguir.

Tabela 24 - Distância de Cohen entre os estratos socioeconômicos entre as provas

	inf	sup
>Red - >Orig	0,04	0,23
Igual_A - Igual_B	0,02	0,05

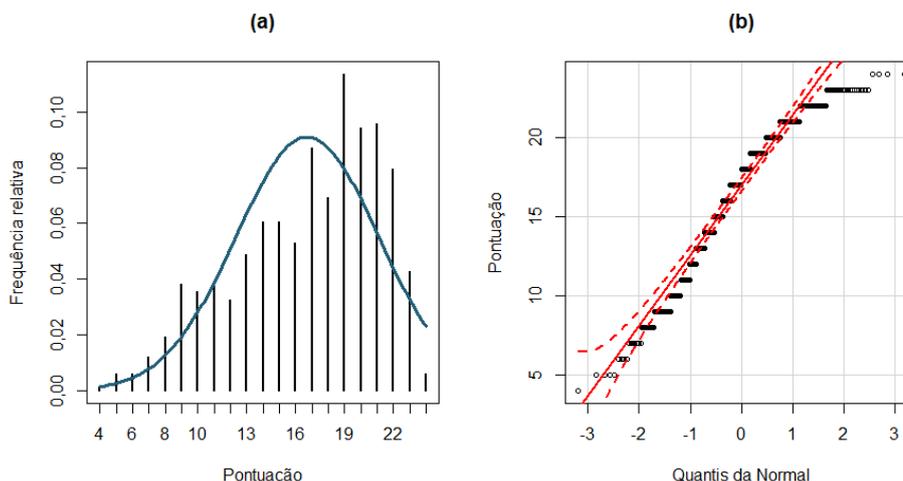
Como visto na outra tabela, a média do grupo sup cresce consideravelmente entre as provas >Orig e >Red, já no grupo inf a média fica estável, a distância de Cohen suporta essa hipótese, uma vez que para o grupo inf temos $d_c = 0,04$, que é irrelevante, enquanto que para o grupo sup temos $d_c = 0,23$, que é pequeno, mas considerável. Nas provas Igual_A e Igual_B se esperam resultados parecidos, e foi o que ocorreu na média, como mostra a distância de Cohen menor que 0,1.

5.3. Testes de Hipótese

As análises estatísticas feitas até esse momento não exigiram ferramentas computacionais ou teóricas muito sofisticadas, mas nessa seção de estatística inferencial a relação do nosso conjunto de dados com as perguntas de pesquisa não é evidente, e por esse motivo as decisões e os resultados obtidos foram realizadas com apoio de uma assessoria estatística de alunos e professores do Instituto de Matemática e Computação Científica da Unicamp.

Primeiramente, para decisão de qual metodologia utilizar, verifica-se a condição de normalidade dos dados, através do gráfico disposto a seguir: na Figura 14 (a), as colunas não seguem o comportamento esperado indicado pela curva azul, ou seja, os dados não parecem seguir uma distribuição normal. Da mesma maneira, no gráfico quantil-quantil (b), as observações não estão em sua maioria dentro da banda de confiança indicada em vermelho. Além disso, o teste de normalidade de Shapiro-Wilks rejeita a normalidade das observações com p -valor < 0.001 . Portanto, nas análises estatísticas, não é possível aplicar as técnicas convencionais utilizadas, que consideram uma população normal.

Figura 14 - Distribuição dos escores por frequência e aderência à distribuição normal



Para uma análise geral dos dados, testando os efeitos de cada variável (Renda, Escolaridade dos pais e gênero) na pontuação dos alunos, utiliza-se o teste não paramétrico de Kruskal-Wallis (1952), análogo a comumente utilizada Análise de Variância (ANOVA) para dados normalmente distribuídos (teste F), pois em testes não paramétricos não existem suposições quanto a distribuição das observações.

No teste de Kruskal-Wallis, avalia-se se a média de diferentes populações são iguais umas às outras. De outra forma, seja, por exemplo, μ_1 , μ_2 , μ_3 e μ_4 as médias da pontuação para cada tipo de resposta na variável renda (respectivamente “Baixa”, “Média”, “Alta”, “Não Sei”), a hipótese nula é $\mu_1 = \mu_2 = \mu_3 = \mu_4$. Rejeita-se a hipótese se pelo menos uma das médias for diferente das outras. De maneira mais formal, seja $F_i(x + \tau_i)$ a função de distribuição acumulada da pontuação do grupo i , $i = 1, 2, \dots, k$. $H_0 : \tau_1 = \tau_2 = \dots = \tau_k = 0$, $H_1 : \tau_j \neq 0$ para pelo menos um j de 1 a k .

- Com relação ao tipo de prova o teste teve p-valor = 0,743
- Com relação ao sexo o teste teve p-valor ≤ 0.001
- Com relação a renda o teste teve p-valor ≤ 0.001
- Com relação a escolaridade do pai o teste teve p-valor ≤ 0.001
- Com relação a escolaridade da mãe o teste teve p-valor ≤ 0.001

Note que não há evidência estatística para detectar diferença entre as provas, como suspeitado na análise inicial. Para as outras variáveis, há pelo menos um grupo estatisticamente diferente dos demais. Em especial, na variável Sexo, que possui apenas dois grupos, pode-se afirmar que um difere do outro, sendo os homens aqueles com maior desempenho.

Como exposto nas análises das seções anteriores, a questão da Renda e Escolaridade dos pais não nos interessa para fazer uma análise aprofundada. Será feita uma análise em relação ao tipo de item, se é Original ou Reduzido, e seu efeito sobre o Sexo do estudante. No entanto, como não havia pessoas que fizeram só itens reduzidos ou só originais, houve a necessidade de analisar o efeito da redução por item separadamente, já verificando também seu efeito de acordo com o sexo do respondente.

Para a análise por tipo de item, foi realizado um teste de Qui-Quadrado de Pearson para quantificar a associação entre tipo de item (original e reduzida) e desempenho dos alunos (acerto e erro). Utilizaremos um nível de significância de 0,05, isso significa que, se o p-valor para algum item for menor que a significância, então há indícios de associação entre as variáveis tipo de item (original ou reduzido) e desempenho dos alunos, para esse item.

Para a análise por tipo de item e sexo, testou-se separadamente, para original e reduzida, a independência entre acerto e sexo utilizando o teste de Pearson, também com uma significância de 0,05. A seguir estão os resultados estatísticos deste teste de hipótese para cada um dos itens utilizando a razão de chance (RC) para esses tipos de variável:

Tipo de Item vs. Desempenho testa se a chance dos estudantes errarem os itens originais é estatisticamente igual a chance dos estudantes errarem os itens reduzidos. Aqui, RC maior do que 1 significa que a chance dos alunos errarem o item original é maior que a chance deles errarem o item reduzido. Como mostrado nas seções anteriores, esperamos que, para a maioria dos itens, essa RC seja maior do que 1, e além disso, para alguns itens seja estatisticamente significativo, com p-valor $< 0,05$.

Acerto vs. Sexo - Original e Acerto x Sexo - Reduzida testa o desempenho de homens e mulheres em cada tipo de item, Original ou Reduzido. Aqui, se RC maior do que 1 significa que a chance de um aluno do sexo masculino acertar é maior que a chance de uma aluna do sexo feminino acertar, discriminando para tipo de item, Original ou Reduzido. Nas seções anteriores vimos que devemos ter RC maior do que 1 na maioria dos itens, tanto Original quanto Reduzido, essa diferença só será significativa para p-valor $< 0,05$.

Estamos em busca de duas informações com esses testes de hipótese. A primeira é se efetivamente a redução de carga cognitiva melhorou, ou piorou, o desempenho dos alunos em cada item, independente do sexo do estudante. Para verificar esse efeito olhamos apenas o teste **Tipo de Item vs. Desempenho**, se o valor da RC for maior do que um para um p-valor $< 0,05$ então quer dizer que o item reduzido possui uma taxa de acerto efetivamente maior que o item original, se o valor for menor do que um para um p-valor $< 0,05$ então isso quer dizer que o item reduzido possui uma taxa de acerto efetivamente menor que o item original.

A segunda informação diz respeito à redução de carga cognitiva do item e o acerto por sexo, será que o item reduzido beneficia o desempenho das mulheres? Para testar esse efeito vamos olhar os itens no **Acerto vs. Sexo - Original** e procurar aqueles em que os homens possuem um desempenho estatisticamente melhor que as mulheres, ou seja, RC maior do que 1 com um p-valor $< 0,05$, e no mesmo item vamos olhar se **Acerto x Sexo - Reduzida** possui um RC tal que o p-valor $> 0,05$. Se isso ocorrer, significa que a taxa

de acerto do item original era maior para os homens antes da redução, e com a redução de carga cognitiva o desempenho de homens e mulheres se torna indistinguível.

Na seção a seguir, cada item da prova, tanto na versão original quanto na versão reduzida, será avaliado conforme o construto pretendido e técnicas de redução de carga cognitiva estranha. Será feita a discussão sobre o desempenho dos estudantes, e as variáveis sexo e tipo de item através dos testes de hipótese descritos.

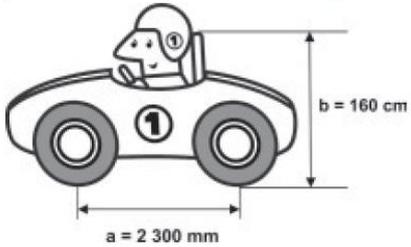
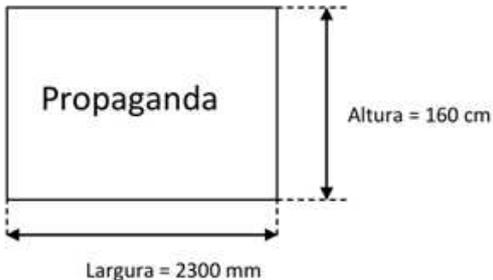
6. O Efeito da Redução de Carga Cognitiva nos Itens

Será feita a seguir uma análise detalhada sobre os efeitos da redução de carga cognitiva estranha em oito itens selecionados. Serão colocados em quadros cada item, na versão Original e Reduzida, lado a lado, e também para cada item os valores dos testes de Razão de Chance (RC) com seus respectivos p-valores para cada um dos três testes de hipótese já mencionados na seção anterior, em quadros distintos. Além disso, nos mesmos quadros onde se localizam as razões de chance, há o histograma de desempenho, com as proporções de acerto (azul) e erro (cinza) para cada um dos sexos, feminino (F) e masculino (M), possibilitando uma visão geral dos efeitos da redução de carga cognitiva estranha. No **Apêndice VI** se encontram todos os valores brutos que geraram as análises à seguir e no **Apêndice VII** se encontram todos os outros itens com suas respectivas análises.

Os valores de relevância estão em negrito e em cores. Para Tipo de Item x Desempenho as cores são: verde, se o item reduzido possui uma taxa de acerto consideravelmente maior que o item original; e amarelo se o item original possui uma taxa de acerto consideravelmente maior que o item reduzido. Para Acerto x Sexo as cores são: azul se nesse item as mulheres passaram a ter um desempenho indistinguível com os homens no item reduzido, mas no original elas tinham um desempenho pior que os homens; e cinza caso nesse item as mulheres tinham um desempenho indistinguível com os homens no item original e passaram a ter um desempenho pior que os homens nos itens reduzidos.

Os itens selecionados a seguir contemplam todos os casos relevantes encontrados, e como as análises dos itens em alguns casos acabaram ficando muito redundantes, decidimos apresentar a maioria deles no apêndice. Ao submeter os itens ao processo de redução de carga cognitiva encontramos casos em que essa redução: beneficiou de maneira geral os estudantes, mas não algum sexo; melhorou o desempenho geral dos estudantes e igualou o desempenho das mulheres com o dos homens, onde no item original isso não acontecia; prejudicou o desempenho geral dos estudantes, mas não algum sexo; piorou o desempenho geral dos estudantes, e fez com que o desempenho das mulheres ficasse pior que o dos homens, onde no item original isso não ocorria. Alguns itens possuíam viés masculino no item original.

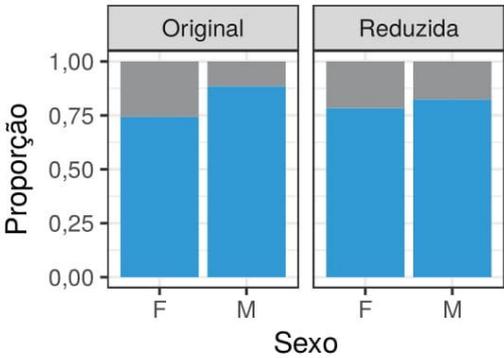
Quadro 5 - Item 1 da Prova e sua respectiva versão Reduzida

Item 1 - Original (Enem 2011 - 136)	Item 1 - Reduzido
<p>Um mecânico de uma equipe de corrida necessita que as seguintes medidas realizadas em um carro sejam obtidas em metros:</p> <p>a) distância a entre os eixos dianteiro e traseiro; b) altura b entre o solo e o encosto do piloto.</p>  <p>Ao optar pelas medidas a e b em metros, obtêm-se, respectivamente,</p> <p>a) 0,23 e 0,16 b) 2,3 e 1,6 c) 23 e 16 d) 230 e 160 e) 2300 e 1600</p>	<p>Considere a seguinte placa de propaganda e suas dimensões</p>  <p>Ao optar pelas medidas da largura e altura em metros, obtêm-se, respectivamente,</p> <p>a) 0,23 e 0,16 b) 2,3 e 1,6 c) 23 e 16 d) 230 e 160 e) 2300 e 1600</p>

O item 1 em suas duas versões está localizado no quadro 5 acima. A redução de carga cognitiva estranha no item 1 apresentou o efeito de igualar o desempenho das mulheres com o dos homens. Como mostra o quadro 6 a seguir, o item original apresenta RC maior que 2,6 para o acerto dos homens em relação ao das mulheres, e de maneira significativa. Já o item reduzido possui RC próximo de 1,3 sem diferença estatística. É preciso mencionar também que a RC entre o item original e reduzido é menor do que 1 mas sem significância estatística, portanto o desempenho entre os dois tipos de item foi indistinguível.

Quadro 6 - Valores de Razão de Chance e p-valor para o Item 1 por tipo de item e sexo, e taxa de acerto por sexo.

Tipo de Item x Desempenho	RC	p-valor
Acerto x Sexo - Original	2,657	<0,001
Acerto x Sexo - Reduzida	1,293	0,649



Na versão original do problema, a taxa de acerto das mulheres (0,75) era consideravelmente menor que a dos homens (0,88) e na versão reduzida as mulheres (0,78) empataram estatisticamente com o desempenho dos homens (0,83). O efeito interessante aqui é que as mulheres tiveram um leve aumento em seu desempenho, e os homens tiveram uma leve queda no seu desempenho, na comparação entre o item original e o reduzido, em geral, como já foi mencionado, os estudantes acertaram mais o original do que o reduzido.

Esse efeito pode ter ocorrido por causa do contexto da questão. O item original possui um viés mais masculino, uma vez que apresenta na imagem um carro e no corpo da questão palavras como mecânico, equipe de corrida, eixos dianteiro e traseiro e encosto do piloto. Avaliamos que o construto desse item é retirar da figura de uma situação problema algumas medidas e fazer transformação de unidades.

Foram efetuadas duas mudanças mais estruturais no item, a primeira foi remover o viés masculino do contexto, e não o contexto em si. O construto pretendido pelo item pode ser acessado através das medidas em uma placa de publicidade. A segunda mudança foi efetivamente na carga cognitiva estranha, o item original explicava o que eram as distâncias a e b de maneira verbal, mostrando depois na figura, e o comando não exigia a compreensão da explicação das distâncias a e b.

Para esse item, utilizamos as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que eliminamos muitas palavras e também simplificamos a linguagem; Extirpação, uma vez que parte do texto associado ao contexto foi removido; e Estética, já que removemos a parte do item que envolvia a explicação das distâncias a e b.

A redução de carga cognitiva estranha no item 3, que possui suas duas versões no quadro 7 a seguir, não teve efeito com relação à melhoria do desempenho das mulheres em relação aos homens na versão reduzida, bem como essa redução de carga cognitiva estranha não acarretou impacto na taxa global de acerto. Como mostra o quadro 8, o item original apresenta RC perto de 1,8 para o acerto dos homens em relação ao das mulheres, já o item reduzido também possui RC próximo de 1,8, ambos com diferença estatística significativa. A RC entre a questão original e reduzida é de aproximadamente 0,9, em outras palavras, o efeito da redução de carga cognitiva piorou modestamente o desempenho global dos estudantes.

Quadro 7 - Item 3 da Prova e sua respectiva versão Reduzida

<u>Item 3 - Original</u> (Enem 2011 - 141)	<u>Item 3 - Reduzido</u>
<p>Em 2010, um caos aéreo afetou o continente europeu, devido à quantidade de fumaça expelida por um vulcão na Islândia, o que levou ao cancelamento de inúmeros voos.</p> <p>Cinco dias após o início desse caos, todo o espaço aéreo europeu acima de 6000 metros estava liberado, com exceção do espaço aéreo da Finlândia. Lá, apenas voos internacionais acima de 31 mil pés estavam liberados.</p> <p>Disponível em: http://www1.folha.uol.com.br. Acesso em: 21 abr. 2010 (adaptado).</p> <p>Considere que 1 metro equivale a aproximadamente 3,3 pés.</p> <p>Qual a diferença, em pés, entre as altitudes liberadas na Finlândia e no restante do continente europeu cinco dias após o início do caos?</p> <p>a) 3390 pés b) 9390 pés c) 11200 pés d) 19800 pés e) 50800 pés</p>	<p>Um avião voa a uma altitude de 6000 metros. Em uma região onde ocorreu a erupção de um vulcão, esse avião teve que subir a 31 mil pés. Considere que 1 m equivale a aproximadamente 3,3 pés.</p> <p>Qual a diferença, em pés, entre as altitudes nessas duas situações?</p> <p>a) 3390 pés b) 9390 pés c) 11200 pés d) 19800 pés e) 50800 pés</p>

Quadro 8 - Valores de Razão de Chance e p-valor para o Item 3 por tipo de item e sexo, e taxa de acerto por sexo.

	RC	p-valor
Tipo de Item x Desempenho	1,037	0,879
Acerto x Sexo - Original	1,77	0,014
Acerto x Sexo - Reduzida	1,803	0,013

Sexo	Original (Proporção)	Reduzida (Proporção)
F	0,55	0,55
M	0,67	0,67

Não é explícito se o item original possui viés em relação ao sexo, uma vez que o contexto versa sobre a altitude de voo de um avião. O contexto parece ligeiramente mais atrativo para estudantes do sexo masculino, mesmo após a redução de carga cognitiva estranha, a essência do contexto prevaleceu, o que significa que se o item possui algum viés em relação ao sexo, ele está presente nas duas versões. Para a versão original e reduzida do problema, a taxa de acerto das mulheres foi igual tanto para os homens (0,67) quanto para as mulheres (0,55).

O construto do item afere sobre aritmética e transformação de unidades em um contexto. No item original existe uma explicação que motiva a existência do item, nesse texto existem alguns dados que são relevantes para a resolução do item, mas no geral o texto inicial possui uma carga cognitiva estranha muito alta. Para o item reduzido mantivemos o contexto, apenas o texto inicial que motiva a situação do problema foi

retirado. Utilizamos as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que eliminamos muitas palavras e também simplificamos a linguagem; e Extirpação, uma vez que parte do texto associado ao contexto foi removido. Porém, cabe aqui constatar que não foi alterada a visão de gênero masculino do item.

Quadro 9 - Item 5 da Prova e sua respectiva versão Reduzida

<u>Item 5 - Original</u> (Enem 2011 - 161)	<u>Item 5 - Reduzido</u>
O número mensal de passagens de uma determinada empresa aérea aumentou no ano passado nas seguintes condições: em janeiro foram vendidas 33000 passagens; em fevereiro, 34500; em março, 36000. Esse padrão de crescimento se mantém para os meses subsequentes. Quantas passagens foram vendidas por essa empresa em julho do ano passado?	Considere uma sequência onde o primeiro termo é 330, o segundo é 345, o terceiro é 360, onde o padrão de crescimento se mantém para os próximos termos. O sétimo termo dessa sequência é:
a) 38000	a) 380
b) 40500	b) 405
c) 41000	c) 410
d) 42000	d) 420
e) 48000	e) 480

O item 5, cujas versões original e reduzida estão no quadro 9 acima, apresentou um resultado diferente dos outros itens. Não houve discrepância entre taxa de acerto de homens e mulheres, tanto na versão original quanto na versão reduzida, mas os estudantes em geral acertaram significativamente mais o item reduzido em comparação ao item original, com uma RC próxima de 3,2, como mostra o quadro 10 a seguir. A RC para o acerto x sexo é bem próxima de um para os dois tipos de item. Em outras palavras, a redução de carga cognitiva beneficiou homens e mulheres de maneira indistinguível.

Quadro 10 - Valores de Razão de Chance e p-valor para o Item 5 por tipo de item e sexo, e taxa de acerto por sexo.

	RC	p-valor
Tipo de Item x Desempenho	3,154	<0,001
Acerto x Sexo - Original	1,119	0,922
Acerto x Sexo - Reduzida	1,225	0,658

O construto do item foi avaliado em encontrar um termo específico de uma sequência numérica, no caso uma progressão aritmética, em uma situação problema. Na versão original do problema, a taxa de acerto das mulheres (0,78) foi estatisticamente igual ao dos homens (0,79), e na versão reduzida as mulheres (0,91) também tiveram o mesmo desempenho que os homens (0,93).

O item original evoca uma empresa cujo número de passagens tem aumentado de mês em mês. Muito embora esse pequeno texto, em teoria, não devesse exercer uma carga cognitiva estranha tão grande, resolvemos removê-lo e o item reduzido ficou, portanto, livre de contexto. Como o construto do item pode ser verificado com números menores, tomamos a iniciativa de remover a carga cognitiva estranha relativa às contas com números grandes.

Nesse item utilizamos as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que a contagem de palavras foi drasticamente reduzida, e além disso no contexto a pergunta foi feita relativa ao mês, e traduzimos apenas para um determinado termo; Extirpação, já que reduzimos o problema na essência do que ele quer saber do estudante; e Simplificação Numérica, uma vez que foram colocados números menores para os respondentes.

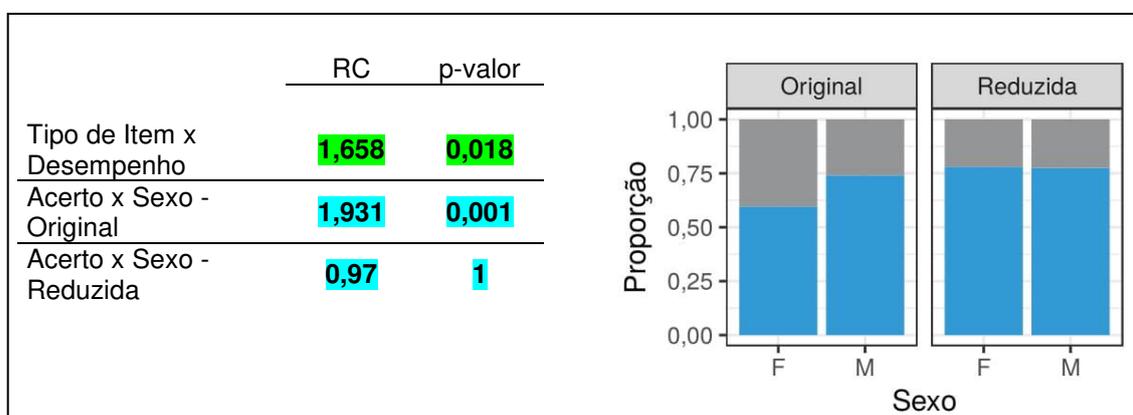
O item 6 está apresentado no quadro 11 a seguir. Esse item apresentou dois efeitos com relação a redução de carga cognitiva estranha: o de igualar o desempenho das mulheres ao dos homens; e fazer com que a taxa de acerto da versão reduzida fosse significativamente maior que a taxa de acerto do item original. Como mostra o quadro 12 a seguir, o item original apresenta RC maior que 1,9 para o acerto dos homens em relação ao das mulheres, de modo significativo. Já o item reduzido possui RC próximo de 1 sem diferença estatística. Apenas entre a versão original e reduzida temos um RC maior que 1,7, estatisticamente significativo.

Esse item, em especial em sua versão original, nos parece privilegiar mais estudantes do sexo masculino, uma vez que o contexto aborda o consumo energético de um chuveiro elétrico. Termos como: energia elétrica e potência, que estão no item original e não no reduzido, podem ter impacto no desempenho dos sexos.

Quadro 11 - Item 6 da Prova e sua respectiva versão Reduzida

<u>Item 6 - Original</u> (Enem 2011 - 163)	<u>Item 6 – Reduzido</u>
Muitas medidas podem ser tomadas em nossas casas visando à utilização racional de energia elétrica. Isso deve ser uma atitude diária de cidadania. Uma delas pode ser a redução do tempo no banho. Um chuveiro com potência de 4800 W consome 4,8 kWh. Uma pessoa que toma dois banhos diariamente, de 10 minutos cada, consumirá, em sete dias, quantos kWh?	Considere que um chuveiro consome 4,8kW por hora. Uma pessoa que toma dois banhos diariamente, de 10 minutos cada, consumirá, em sete dias, quantos kW?
a) 0,8	a) 0,8
b) 1,6	b) 1,6
c) 5,6	c) 5,6
d) 11,2	d) 11,2
e) 33,6	e) 33,6

Quadro 12 - Valores de Razão de Chance e p-valor para o Item 6 por tipo de item e sexo, e taxa de acerto por sexo

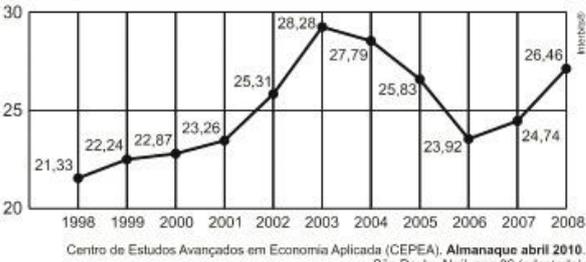
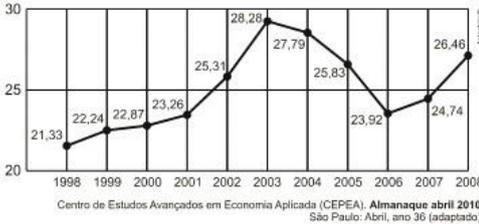


Sobre o construto do item, entendemos que é obter, a partir de um contexto, o valor de uma operação aritmética. Chamamos a atenção sobre dois resultados relevantes referentes à redução da carga cognitiva estranha nesse item. O primeiro é que a na versão original do problema, a taxa de acerto das mulheres (0,6) era consideravelmente menor que a dos homens (0,74) e na versão reduzida as mulheres (0,78) ultrapassaram o desempenho dos homens (0,77). O segundo é que globalmente os estudantes foram melhores na versão reduzida do item como mostrou o teste de hipótese, e isso ocorreu quase exclusivamente por causa das mulheres, uma vez que o desempenho dos homens do item original para o reduzido não foi muito melhor.

Nesse item decidimos manter um contexto em sua versão reduzida, apenas retiramos a primeira parte do corpo, que julgamos exercer carga cognitiva estranha para os respondentes. Até a menção sobre a redução do tempo de banho, nenhuma informação

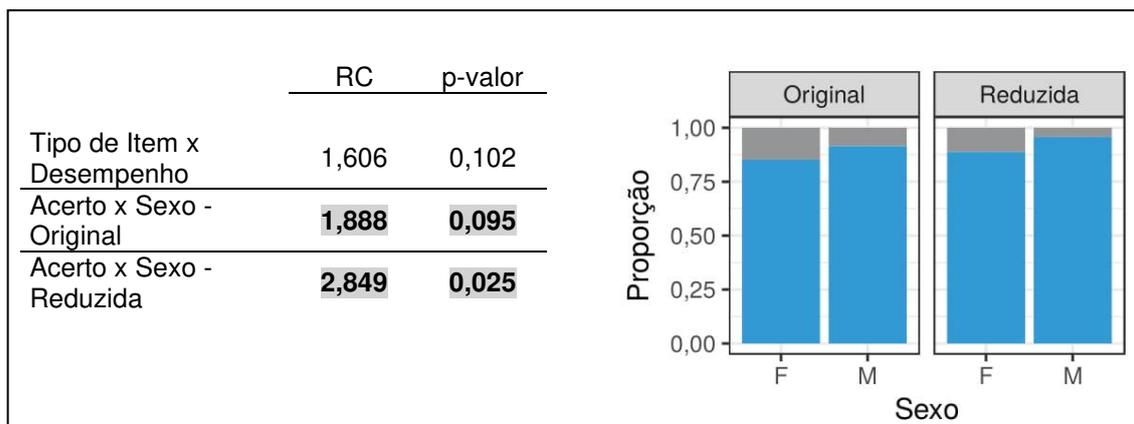
relevante para a resolução do item foi apresentada. As técnicas de redução de carga cognitiva estranha que utilizamos para esse item foram: Tradução, uma vez que a contagem de palavras foi reduzida, e Extirpação, já que extraímos do corpo do item um pedaço do corpo, que provavelmente exerce carga cognitiva estranha nos respondentes, reduzindo assim o problema na essência do que ele quer saber do estudante.

Quadro 13 - Item 9 da Prova e sua respectiva versão Reduzida

<p>Item 9 - Original (Enem 2011 - 176)</p> <p>O termo agronegócio não se refere apenas à agricultura e à pecuária, pois as atividades ligadas a essa produção incluem fornecedores de equipamentos, serviços para a zona rural, industrialização e comercialização dos produtos.</p> <p>O gráfico seguinte mostra a participação percentual do agronegócio no PIB brasileiro:</p>  <p>Centro de Estudos Avançados em Economia Aplicada (CEPEA). Almanaque abril 2010. São Paulo: Abril, ano 36 (adaptado).</p> <p>Esse gráfico foi usado em uma palestra na qual o orador ressaltou uma queda da participação do agronegócio no PIB brasileiro e a posterior recuperação dessa participação, em termos percentuais. Segundo o gráfico, o período de queda ocorreu entre os anos de</p> <ol style="list-style-type: none"> 1998 e 2001 2001 e 2003 2003 e 2006 2003 e 2007 2003 e 2008 	<p>Item 9 – Reduzido</p> <p>O gráfico a seguir mostra a participação percentual do agronegócio no PIB brasileiro:</p>  <p>Centro de Estudos Avançados em Economia Aplicada (CEPEA). Almanaque abril 2010. São Paulo: Abril, ano 36 (adaptado).</p> <p>Segundo o gráfico, o período de queda ocorreu entre os anos de:</p> <ol style="list-style-type: none"> 1998 e 2001 2001 e 2003 2003 e 2006 2003 e 2007 2003 e 2008
---	---

O item 9 em suas duas versões está localizado no quadro 13 acima. O desempenho das mulheres nesse item foi pior que o dos homens quando comparamos o item original com a sua versão reduzida. Como mostra o quadro 14 a seguir, o item original apresenta RC maior que 1,8 para o acerto dos homens em relação ao das mulheres, já o item reduzido possui RC próximo de 2,8, em favor dos homens, com um p-valor bem abaixo da significância. A RC entre a questão original e reduzida é de aproximadamente 1,6 com um p-valor próximo de 0,1, portanto não podemos afirmar que a versão reduzida beneficiou globalmente os estudantes.

Quadro 14 - Valores de Razão de Chance e p-valor para o Item 9 por tipo de item e sexo, e taxa de acerto por sexo



A redução de carga cognitiva estranha em si não prejudicou as mulheres na versão reduzida, mas beneficiou muito mais os homens. Os índices de acerto nos dois tipos de item foram bem altos, na versão original do problema, a taxa de acerto das mulheres (0,85) era consideravelmente menor que a dos homens (0,92) e na versão reduzida as mulheres (0,89) não tiveram um aumento tão grande enquanto que os homens quase gabaritaram (0,95).

O construto pretendido pelo item, ao nosso ver, é extrair informações de um gráfico a partir de um contexto. Esse contexto pode impactar o desempenho entre os sexos, uma vez que o tema é o agronegócio que pode ser mais atrativo para o sexo masculino. A taxa de acerto desse item, utilizando os microdados do Enem, no recorte de candidatos presentes nos dois dias de prova, e que não zeraram a redação, já era consideravelmente alta. Por esse motivo tentamos reduzir bastante a carga cognitiva estranha do problema, mas um contexto mínimo ainda permanece.

Para esse item utilizamos as estratégias de: Tradução, uma vez que a contagem de palavras foi bem reduzida; e Extirpação, já que muitos termos que coletivamente podiam afetar a Memória de Trabalho dos alunos, como fornecedores, industrialização, comercialização e orador, foram removidos.

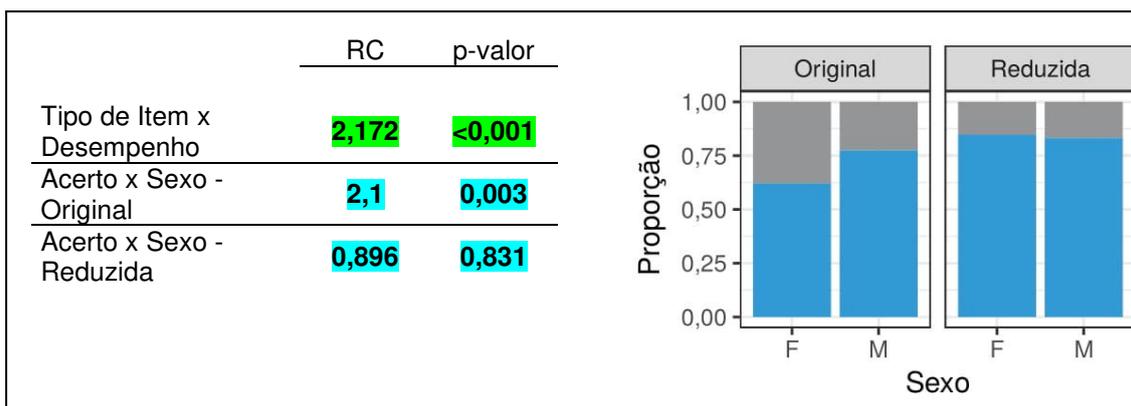
O item 10, em suas duas versões, está localizado no quadro 15 a seguir. Assim como ocorreu no item 6, o item 10 apresentou dois efeitos com relação a redução de carga cognitiva estranha: o de igualar o desempenho das mulheres ao dos homens; e fazer com que a taxa de acerto da versão reduzida fosse significativamente maior que a taxa de acerto do item original. Como mostra o quadro 16 a seguir, o item original apresenta RC

maior que 2,1 para o acerto dos homens em relação ao das mulheres de maneira significativa, já o item reduzido possui RC próximo de 0,9 sem diferença estatística. Entre a versão original e reduzida temos um RC maior que 2,1, com um p-valor bem abaixo da significância.

Quadro 15 - Item 10 da Prova e sua respectiva versão Reduzida

Item 10 - Original (Enem 2012 - 155)	Item 10 - Reduzido
<p>As curvas de oferta e de demanda de um produto representam, respectivamente, as quantidades que vendedores e consumidores estão dispostos a comercializar em função do preço do produto. Em alguns casos, essas curvas podem ser representadas por retas. Suponha que as quantidades de oferta e de demanda de um produto sejam, respectivamente, representadas pelas equações:</p> $Q_O = -20 + 4P$ $Q_D = 46 - 2P$ <p>Em que Q_O é quantidade de oferta, Q_D é a quantidade de demanda e P é o preço do produto. A partir dessas equações, de oferta e de demanda, os economistas encontram o preço de equilíbrio de mercado, ou seja, quando Q_O e Q_D se igualam. Para a situação descrita, qual o valor do preço de equilíbrio?</p> <p>a) 5 b) 11 c) 13 d) 23 e) 33</p>	<p>Seja $y_1 = -20 + 4x$ e $y_2 = 46 - 2x$. Para que y_1 seja igual a y_2, devemos ter que o valor de x é:</p> <p>a) 5 b) 11 c) 13 d) 23 e) 33</p>

Quadro 16 - Valores de Razão de Chance e p-valor para o Item 10 por tipo de item e sexo, e taxa de acerto por sexo



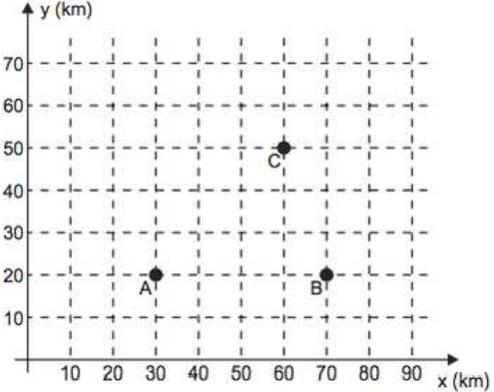
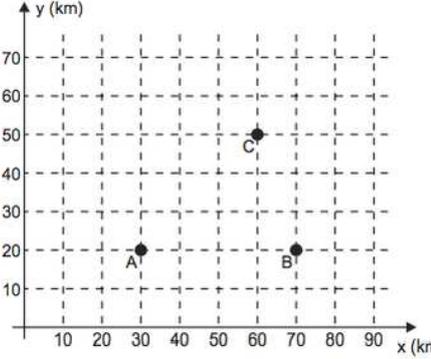
É necessário mencionar dois resultados relevantes referentes a esses números apresentados no quadro 16. O primeiro é que na versão original do problema, a taxa de acerto das mulheres (0,62) era consideravelmente menor que a dos homens (0,78) e na versão reduzida as mulheres (0,85) ultrapassaram o desempenho dos homens (0,84). O segundo é que globalmente os estudantes foram melhores na versão reduzida do item

como mostrou o teste de hipótese, tanto homens quanto mulheres tiveram um aumento considerável no desempenho.

Sobre o construto do item, acreditamos que é resolver um sistema de equações a partir de um contexto que, para nós não, apresenta viés em relação ao sexo, uma vez que o item versa sobre curva de oferta e demanda. Para esse item tomamos a decisão de verificar o efeito sobre o desempenho dos estudantes se retirássemos todo o contexto.

Para esse item utilizamos as estratégias de redução de carga cognitiva de: Tradução, já que as equações presentes no item original apresentavam variáveis Q e P, e transformamos para y e x; e Extirpação, uma vez que todo contexto associado ao item foi removido.

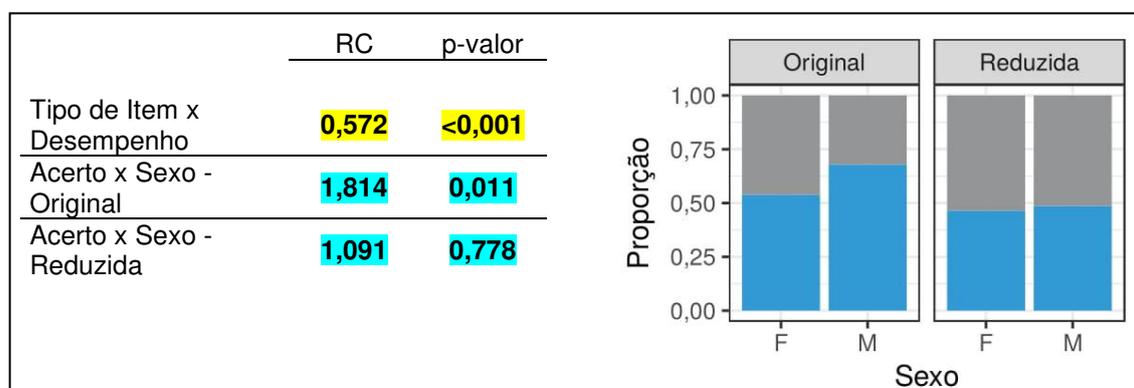
Quadro 17 - Item 18 da Prova e sua respectiva versão Reduzida

Item 18 - Original (Enem 2013 - 168)	Item 18 – Reduzido
<p>Nos últimos anos, a televisão tem passado por uma verdadeira revolução, em termos de qualidade de imagem, som e interatividade com o telespectador. Essa transformação se deve à conversão do sinal analógico para o sinal digital. Entretanto, muitas cidades ainda não contam com essa nova tecnologia. Buscando levar esses benefícios a três cidades, uma emissora de televisão pretende construir uma nova torre de transmissão, que envie sinal às antenas A, B e C, já existentes nessas cidades. As localizações das antenas estão representadas no plano cartesiano:</p>  <p>A torre deve estar situada em um local equidistante das três antenas. O local adequado para a construção dessa torre corresponde ao ponto de coordenadas</p> <p>a) (65;35) b) (53;30) c) (45;35) d) (50;20) e) (50;30)</p>	<p>Considere os pontos A, B e C representados no plano cartesiano</p>  <p>O ponto equidistante dos pontos A, B e C possui coordenadas:</p> <p>a) (65;35) b) (53;30) c) (45;35) d) (50;20) e) (50;30)</p>

O item 19, em suas duas versões, está localizado no quadro 17. A redução de carga cognitiva estranha proporcionou a igualdade de desempenho entre homens e mulheres nesse item, entre as versões original e reduzida. Em contrapartida a redução de carga cognitiva estranha fez com que o escore global fosse menor na versão reduzida. O contexto era relevante e aparentemente ajudou os respondentes.

Como está no quadro 18 a seguir, a RC para de acerto entre homens e mulheres na versão original é próxima de 1,8, com vantagem significativa para os homens. Na versão reduzida do item, a RC ficou próxima de 1,1, logo o desempenho entre os sexos ficou indistinguível. As mulheres não igualaram o desempenho com os homens, nesse item o desempenho geral dos estudantes no item reduzido, em comparação ao item original, foi significativamente menor, com uma RC próxima de 0,6. Na versão original do problema, a taxa de acerto das mulheres (0,54) foi consideravelmente menor do que a dos homens (0,67) e na versão reduzida as mulheres (0,47) tiveram um desempenho estatisticamente igual aos homens (0,49).

Quadro 18 - Valores de Razão de Chance e p-valor para o Item 18 por tipo de item e sexo, e taxa de acerto por sexo



Sobre o construto do item, entendemos que é utilizar geometria analítica, ou geometria, em um contexto. Não verificamos viés de sexo uma vez que o problema aborda um contexto neutro sobre torres de sinal de televisão. Para o item reduzido removemos todo o contexto do problema, desse modo esperávamos que a carga cognitiva estranha associada à motivação da conexão do problema com a geometria analítica fosse suprimida. Pela redução do escore na versão reduzida desse item, é possível que o contexto, mesmo

trazendo carga cognitiva estranha, traga mais benefícios do que prejudique os respondentes.

Para esse item, as estratégias de redução de carga cognitiva utilizadas foram: Tradução, uma vez que reduzimos substancialmente a quantidade de palavras; e Extirpação, já que reduzimos o problema na essência do que ele quer saber do estudante.

De maneira global, em seis itens (5, 6, 10, 11, 17 e 20) a redução de carga cognitiva aumentou a chance dos alunos acertarem o item reduzido, muito embora em dois itens (18 e 23) a redução de carga cognitiva impactou negativamente o acerto dos alunos no item reduzido. Esse efeito também ocorreu no artigo de Gillmore, Poggio e Embretson (2015), onde em uma prova com 15 itens, a redução de carga cognitiva teve escores maiores em seis itens para um p-valor $< 0,05$, mas em um item o desempenho ficou pior em comparação com o item original, também com um p-valor $< 0,05$.

Esses resultados de teste de hipótese, juntamente com os dados de distância de Cohen e estatística descritiva, corroboram a tese de que a redução de carga cognitiva estranha nos itens de matemática tende a melhorar o desempenho dos estudantes. A distância de Cohen entre as provas >Red e >Orig foi de 0,14, enquanto que essa distância na prova constituída apenas por itens Reduzidos e itens Originais foi de 0,19. De um modo geral, isso afeta o escore total de uma prova, que contém uma quantidade maior, ou menor, questões reduzidas (ou originais).

Com respeito ao efeito sexo no desempenho entre itens originais e reduzidos, em nove itens (1, 2, 6, 7, 10, 14, 16, 18 e 21) ocorreu efeito positivo para as mulheres. Esses itens em sua versão original proporcionavam uma taxa de acerto maior para os homens em comparação com as mulheres com um p-valor $< 0,05$, mas na versão reduzida desses itens não houve algum sexo com desempenho melhor. Em outras palavras, o desempenho dos estudantes se tornou indistinguível em relação ao sexo após a redução de carga cognitiva estranha.

Em contrapartida em quatro itens (9, 12, 19 e 24), as versões originais tinham taxa de acerto indistinguível em relação ao sexo, mas na versão reduzida desses itens os homens acertam mais do que as mulheres com um p-valor $< 0,05$. Em outras palavras, o desempenho feminino foi prejudicado.

A distância de Cohen para os sexos também reforça a nossa hipótese de que a redução de carga cognitiva estranha beneficia mais as mulheres do que os homens. Essa distância, entre os homens que fizeram a prova >Red e >Orig, ficou em 0,1 enquanto para as mulheres, também comparando as provas >Red e >Orig, a distância foi de 0,19. Comparando as provas constituídas apenas por itens Originais e Reduzidos, o escore feminino passou de 15 para 15,9, enquanto os homens passaram de 17,5 para 17,8.

Como foi descrito no capítulo sobre distância de Cohen, pudemos apenas estimar o valor dessa distância entre homens e mulheres para essas duas provas, Original e Reduzida. Enquanto para a Original essa distância está entre 0,52 e 0,61, para a Reduzida ela se encontra entre 0,37 e 0,43. Os valores encontrados nas seções anteriores indicam que a redução de carga cognitiva estranha nos itens dessa prova impacta mais o desempenho das mulheres, melhorando o escore feminino, do que o desempenho dos homens.

Sobre o viés de sexo nos itens dessa prova, verificamos que esse efeito pode ter ocorrido em quatro desses itens. Em três deles, com a redução de carga cognitiva, o desempenho das mulheres melhorou, em comparação com os homens, e em um o desempenho feminino piorou frente ao masculino.

A remoção completa do contexto ocorreu em sete itens, em quatro itens o item modificado (sem contexto) proporcionou às mulheres um desempenho estatisticamente igual ao dos homens, onde no item original o desempenho masculino era superior ao feminino. Em dois itens, essa remoção de contexto não alterou estatisticamente a diferença de desempenho entre homens e mulheres, particularmente nesses dois itens, tanto a versão original quanto a modificada, os homens tiveram um escore maior. Em um item, o desempenho das mulheres piorou consideravelmente, em comparação com os homens, ao retirar o contexto do item.

7. Considerações Finais

As avaliações em larga escala, mesmo com todas as suas implicações e discussões educacionais, políticas ou ideológicas, são um marco fixo e perene para a nossa sociedade do século XXI. Mesmo em universos onde a avaliação pode existir em sua interface micro e macro, interna ou externa, as teorias aplicadas a avaliações educacionais podem trazer informações úteis, e quando o conjunto de respondentes se torna consideravelmente grande, a avaliação em larga escala é um instrumento viável e que pode retornar boas devolutivas, tanto do teste quanto dos questionários socioeconômicos.

A sensação que tive durante esse trabalho é que nunca compreendemos exatamente um teste que aplicamos, tanto com relação à finalidade do teste quanto ao construto pretendido por cada componente do teste. Existe uma aparente função, que envolve o esforço que o *designer* do teste empenha na construção de cada item com a confiabilidade e a validade do teste. Uma vez que devemos considerar que os respondentes não saltaram de paraquedas para fazer um instrumento de avaliação, temos que respeitar o tempo e dedicação, tanto de quem confeccionou o teste e também dos respondentes que se prepararam para ele.

Tendo escrito isso, e com essa ideia em mente, tomemos o Enem. Esse exame em larga escala, que outrora pressupostamente serviu para balizar o Ensino Médio, desde 2009 é o maior vestibular do Brasil, em que milhares de vagas em instituições federais de ensino superior (IFES) são preenchidas a cada semestre por candidatos que sofreram um calvário de nove horas de exame divididos em dois dias. Algumas carreiras acessíveis pelo Enem possuem uma concorrência brutal, de modo que o erro em um ou dois testes, ou a transcrição equivocada de uma resposta para o gabarito, podem significar o fim do sonho de um estudante em entrar em um curso muito disputado.

O que exatamente se pretende medir sobre o conhecimento em matemática de um jovem, quando esse estudante começa a resolver a prova de matemática do Enem após algumas horas investidas em uma redação e 45 itens de linguagens? Após enfrentar diversos textos e interpretações, o estudante precisa fazê-lo novamente em todos os itens de matemática?

Além das vagas em IFES, é preciso lembrar que durante muito tempo o Enem também foi utilizado para certificação do Ensino Médio, que é um documento muito forte do ponto de vista pessoal e profissional. Até hoje um bom desempenho no Enem também possibilita o acesso a bolsas, parciais ou integrais, em instituições particulares de ensino superior para alunos carentes, por meio do Prouni. O futuro de muita gente depende dessa avaliação, e meditar sobre toda essa cascata educacional é um dever de professores e pesquisadores.

Em matemática, como discutido nos primeiros capítulos desse trabalho, o Enem abrange o conceito de letramento matemático, a contextualização do saber, a aplicação do que o aluno vê em sala de aula para seu universo. Como professor há mais de dez anos em todos os níveis do ensino, essa ideia me parece fundamental. O aluno deve sair do Ensino Médio com um conjunto de conteúdos, em todas as disciplinas, para poder exercer sua cidadania na sociedade. Em matemática, o conjunto de competências e habilidades adquiridas pelos estudantes pode ser utilizado de diversas maneiras no dia a dia, desde fazer uma compra de maneira sábia até estimar quanto tempo demora para chegar a pé em determinado lugar.

O caso do contexto em matemática sempre me deixou com a impressão de que ele proporciona um benefício maior para alunas do que para alunos. As primeiras inspeções que fizemos nos microdados do Enem eram na tentativa de encontrar alguma correlação entre as notas das mulheres nas provas de redação, particularmente das competências 3 e 4, e linguagens, com as questões contextualizadas que exigiam leitura atenta, uma vez que acreditávamos que os homens tinham uma correlação menor nesse sentido.

Encontramos que as mulheres tinham de fato uma média maior, em comparação aos homens, em redação e em linguagens, e menor em matemática, mas a correlação que desejávamos, entre as notas da redação (ou itens de linguagem) e questões de matemática que envolviam mais leitura e interpretação, não existia. O trabalho de Gillmore, Poggio e Embretson (2015) abriu uma porta, que não foi evidente e nem imediata, para tentar quantificar o desempenho na contextualização em questões de matemática.

Ao perceber que poderíamos utilizar itens com o mesmo construto, em duas versões diferentes, decidimos produzir e aplicar um instrumento de avaliação e um questionário socioeconômico. Os resultados dos microdados do Enem nos ajudaram a selecionar bons itens, com índice de discriminação adequado de modo que o Alfa de

Cronbach (confiabilidade) fosse alto, e as técnicas e exemplos de redução de carga cognitiva do trabalho de Gillmore, Poggio e Embretson nos guiaram para executar o mesmo com os itens do Enem.

O conjunto de respondentes foi surpreendentemente diversificado do ponto de vista de desempenho e também do questionário socioeconômico. Com esse grande número de dados fomos capazes de transcender a discussão item original x item reduzido, ou em uma outra leitura, item contextualizado x item menos contextualizado, em algum grau de nível de contextualização. A comparação de desempenho por sexo, por renda, ou por escolaridade dos pais nessa discussão de item original, ou reduzido, tornou-se possível.

O resultado dos testes mostrou que o desempenho médio feminino é menor que o masculino em todas as provas, mas de maneira robusta podemos afirmar que a redução de carga cognitiva nos itens foi benéfica no desempenho das mulheres. O fato de problemas algébricos contextualizados que possuem mais informação, ou até identificar que faltam informações, prejudicarem mais mulheres do que homens já foi percebido por Low e Over (1993).

Em um trabalho extenso, com um banco de dados de 20 anos, Walkington, Clinton e Shivraj (2018) mostraram que em problemas contextualizados de matemática status socioeconômico reflete de maneira robusta o desempenho dos estudantes: fatores textuais mais rebuscados estão negativamente correlacionados com o desempenho de grupos socialmente mais frágeis, enquanto que em problemas com linguagem mais simples parecem melhorar o desempenho desse mesmo grupo. Como apontado por Beilock (2008), uma situação estressante afeta a Memória de Trabalho nas avaliações, no nosso caso parece ser o gatilho para que essa falta de Memória de Trabalho tenha como efeito nas alunas de incorporar ou se ocupar mais com informações irrelevantes.

Do ponto de vista da TCC temos uma hipótese para explicar esse efeito. Grupos socialmente mais fortes, ou seja, com renda maior ou pais com nível de escolaridade mais altos, entregam a seus filhos e filhas ambientes (escola ou cursos por exemplo) em que diversas informações são processadas pelo seu sistema cognitivo, e onde a carga cognitiva pertinente possa ser efetivamente trabalhada, aumentando o repositório na Memória de Longa Duração. Esse estofo intelectual adquirido é utilizado também nos problemas de

leitura, e estudantes de grupos sociais mais frágeis, com menos contato com os ambientes intelectuais mais privilegiados possuem um repositório mais limitado.

Em particular, o tipo de contexto pode mudar completamente o desempenho dos alunos, se existe uma pequena possibilidade de viés, vimos dois casos onde a neutralidade do contexto equiparou as notas médias dos homens e das mulheres na versão reduzida do item. A redução de carga cognitiva estranha, assim como a forma que um item é montado, é uma escolha pessoal do *designer* da questão. O propósito dessa modificação era tentar atingir de maneira mais contundente o construto pretendido pelo item. Acreditamos que esse efeito passou pela contextualização mais enxuta de cada item, em algum nível de intensidade.

Agora voltamos a refletir sobre o começo desse capítulo de considerações finais, em especial na parte sobre a relevância do Enem para os candidatos a vagas, via Sisu nas IFES ou no Prouni. Se o desempenho das mulheres pode ser afetado fortemente pelo tipo de contextualização, e algumas carreiras muito concorridas demandam escores altíssimos, então o jogo pode deixar em desigualdade um conjunto bem grande de candidatos. Mais ainda, muito embora os testes de hipótese não tenham sido utilizados nos estratos de renda e escolaridade dos pais, os resultados indicam que a redução de carga cognitiva beneficia mais candidatos com mais renda e de pais com mais escolaridade.

A balança não se equilibra, mas o designer do item agora precisa saber das implicações sobre a escolha do contexto da questão, das palavras e definições que ele pode colocar lá para aumentar, ou não, a carga cognitiva estranha, da figura que ele for utilizar, e assim por diante.

Não sabemos o motivo por trás desse efeito sobre o desempenho das mulheres nas questões originais e reduzidas. Uma hipótese é a seguinte: como as mulheres possuem um desempenho melhor em redação e linguagens em relação aos homens, acreditamos que isso possa ocorrer pelo fato das mulheres lerem atentamente todo o item. Em uma prova com diversos itens contextualizados de matemática essa atenção pode ser prejudicial, uma vez que a carga cognitiva estranha embutida em informações inócuas do item afetam mais quem mais presta atenção a todos os detalhes, sobrecarregando a Memória de Trabalho. Em especial, a maioria dos itens de matemática do Enem podem ser acertadas por meio de uma leitura diagonal no corpo e uma leitura atenta no comando da questão.

Outra hipótese, menos provável, para explicar o efeito do melhor desempenho feminino nos itens reduzidos, é a possível associação entre a redução de carga cognitiva estranha e outros efeitos, como ansiedade matemática ou crenças de autoeficácia, onde as mulheres possuem maior sensibilidade do que os homens. Muito embora o instrumento de avaliação não tenha tido impacto na vida dos respondentes, já que foi apenas um simulado do Enem, e alguma intensidade desses efeitos pode ter ocorrido.

Assim como no trabalho de Gillmore, Poggio e Embretson (2015), o mesmo efeito surpreendente ocorreu em relação ao Alfa de Cronbach. Quando se tenta medir melhor o construto de cada item, espera-se que o Alfa da prova aumente, uma vez que a avaliação deve possuir mais consistência interna, e como os dados mostram, entre as provas >Orig e >Red, a avaliação com mais itens originais possui um Alfa de Cronbach maior que a avaliação com mais itens reduzidos. Uma explicação possível é a que, nas avaliações onde as mulheres em média vão pior que os homens, a posição delas está mais consistente do que estaria se por acaso elas tivessem em média desempenhos melhores, nesse caso elas possuem desempenhos parecidos com o dos homens, e a posição delas fica em média mais embaralhada com a dos homens. Em outras palavras, a redução de carga cognitiva não abriu ainda mais o leque de notas, acabou condensando mais.

Da mesma forma que a TCC pode explicar o motivo de um estudante errar um problema, através da sobrecarga da Memória de Trabalho, a TCC também pode dar indícios de como trabalhar esse efeito prejudicial, para as mulheres, de resolver itens de matemática com muita informação. Independente dos efeitos, que em média são mais prejudiciais para as mulheres, como ansiedade, autoeficácia ou competição, esse trabalho mostrou que ao nos fixarmos nos construtos de matemática, o desempenho das mulheres não ficou discrepante em relação aos homens.

Talvez trabalhar itens de matemática em seus dois formatos com as turmas possa trazer essa percepção para todos os alunos, de discernir o construto matemático das outras informações. A TCC também ensina que a Memória de Trabalho menos sobrecarregada ajuda a carga cognitiva pertinente a guardar e aprender informações relevantes de modo permanente, a continuidade desse processo leva a esquemas, portanto o sistema cognitivo processa menos elementos a serem contabilizados na Memória de Trabalho. Para exercer a cidadania, essa habilidade é muito desejável, uma vez que somos bombardeados por

dados e contextos distintos diariamente, reter as informações relevantes é sempre desafiador.

Em suma, o trabalho mostrou-se gratificante. Os resultados se mostraram valiosos, para professores, setores de educação e *designers* de itens. Avaliações podem ter um valor imensurável para alguns candidatos, e como mencionado no começo desse último capítulo, se existe uma função que envolve o esforço e empenho para a construção de um item, ou de um conjunto de itens, para acessar de uma melhor maneira o construto pretendido, essa deve ser tangenciada com a intenção de equilibrar essa balança tênue que é o desempenho de homens e mulheres em matemática.

8. Referências Bibliográficas

ADELSON, Jill L.; DICKINSON, Emily R.; CUNNINGHAM, Brittany C. Differences in the reading–mathematics relationship: A multi-grade, multi-year statewide examination. **Learning and Individual Differences**, v. 43, p. 118-123, 2015.

DA SILVA AGUIAR, Glauco; ORTIGÃO, Maria Isabel Ramalho. Letramento em Matemática: um estudo a partir dos dados do PISA2003. **Boletim de Educação Matemática**, v. 26, n. 42 A, p. 1-21, 2012.

AERA (AMERICAN EDUCATIONAL RESEARCH ASSOCIATION) et al. **Standards for educational and psychological testing**. Amer Educational Research Assn, 1999.

ANASTASI, Anne. **Differential psychology** (3rded.). New York: Macmillan, 1956.

AYRES, Paul. Systematic Mathematical Errors and Cognitive Load. **Contemporary Educational Psychology**, v. 26, n. 2, p. 227-248, 2001.

BADDELEY, Alan. Domains of recollection. **Psychological Review**, v. 89, n. 6, p. 708, 1982.

BADDELEY, Alan. Exploring the central executive. **The Quarterly Journal of Experimental Psychology Section A**, v. 49, n. 1, p. 5-28, 1996.

BADDELEY, Alan. Recent developments in working memory. **Current opinion in neurobiology**, v. 8, n. 2, p. 234-238, 1998

BEILOCK, Sian L. Math performance in stressful situations. **Current Directions in Psychological Science**, v. 17, n. 5, p. 339-343, 2008.

BRASIL, Ministério da Educação e do Desporto, Secretaria de Educação Fundamental. **Parâmetros curriculares nacionais**. Brasília: MEC/SEF, 1998.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Portaria no 109, de 27 de maio de 2009. Diário Oficial da União, Brasília, n. 100, p. 56-63, maio 2009.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais. **Exame Nacional do Ensino Médio – Enem : relatório final 1998**. Brasília, INEP, 79 p. 1999.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais. **Brasil no PISA 2015: Análises e reflexões sobre o desempenho dos estudantes brasileiros**. Disponível em: http://download.inep.gov.br/acoes_internacionais/pisa/resultados/2015/pisa2015_completo_final_baixa.pdf. Acesso em: 22 mar 2018.

BROWN, William. Some experimental results in the correlation of mental abilities 1. **British Journal of Psychology**, 1904-1920, v. 3, n. 3, p. 296-322, 1910.

CARLSON, Robert; CHANDLER, Paul; SWELLER, John. Learning and understanding science instructional material. **Journal of educational psychology**, v. 95, n. 3, p. 629, 2003.

CHANDLER, Paul; SWELLER, John. Cognitive load theory and the format of instruction. **Cognition and instruction**, v. 8, n. 4, p. 293-332, 1991.

CHANDLER, Paul; SWELLER, John. Cognitive load while learning to use a computer program. **Applied cognitive psychology**, v. 10, n. 2, p. 151-170, 1996.

CLARKE, Tracey; AYRES, Paul; SWELLER, John. The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. **Educational Technology Research and Development**, v. 53, n. 3, p. 15-24, 2005.

CLARK, Ruth C.; NGUYEN, Frank; SWELLER, John. **Efficiency in learning: Evidence-based guidelines to manage cognitive load**. John Wiley & Sons, 2011.

COHEN, J. Statistical power analysis for the behavioral sciences (2ª ed.). **Hillsdale, NJ: Lawrence Erlbaum Associates**, 1988.

CORTI, Ana Paula. As diversas faces do ENEM: análise do perfil dos participantes (1999-2007). **Estudos em Avaliação Educacional**, v. 24, n. 55, p. 198-221, 2013.

COWAN, Nelson. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. **Psychological bulletin**, v. 104, n. 2, p. 163, 1988.

COWAN, Nelson. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. **Behavioral and brain sciences**, v. 24, n. 1, p. 87-114, 2001.

COWAN, Nelson. What are the differences between long-term, short-term, and working memory? **Progress in brain research**, v. 169, p. 323-338, 2008.

CRONBACH, Lee J. Coefficient alpha and the internal structure of tests. **Psychometrika**, v. 16, n. 3, p. 297-334, 1951.

DEPRESBITERIS, Léa. Avaliação na educação básica: ampliando a discussão. **Estudos em Avaliação Educacional**, n. 24, p. 137-146, 2001.

DEPRESBITERIS, Léa. Avaliando competências na escola de alguns ou na escola de todos? **Boletim Técnico do Senac**. São Paulo, v. 27, n. 3, p. 36-48, 2001.

DEVINE Amy et al, Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety **Behavioral and Brain Functions**, Londres, v.8, n.33, p.1 - 9, 2012.

EBEL, Robert L.; FRISBIE, David A. **Essentials of education measurement**. Englewood Cliffs, NJ: Prentice-Hall. 1986.

ELSE-QUEST, Nicole M.; HYDE, Janet Shibley; LINN, Marcia C. Cross-national patterns of gender differences in mathematics: a meta-analysis. **Psychological bulletin**, v. 136, n. 1, p. 103, 2010.

FREUDENTHAL, Hans. **Revisiting mathematics education**. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1991

GILLMORE, Susan C.; POGGIO, John; EMBRETSON, Susan. Effects of reducing the cognitive load of mathematics test items on student performance. **Numeracy**, v. 8, n. 1, p. 4, 2015.

GATTI, Bernardete. A. Avaliação de sistemas educacionais no Brasil. **Revista de Ciências da Educação**, v.9, p.7-18, 2009.

GRAVEMEIJER, Koeno. How emergent models may foster the constitution of formal mathematics. **Mathematical thinking and learning**, v. 1, n. 2, p. 155-177, 1999.

GRAVEMEIJER, Koen; DOORMAN, Michiel. Context problems in realistic mathematics education: A calculus course as an example. **Educational studies in mathematics**, v. 39, n. 1-3, p. 111-129, 1999.

GRAVEMEIJER, Koen; TERWEL, Jan. Hans Freudenthal: a mathematician on didactics and curriculum theory. **Journal of curriculum studies**, v. 32, n. 6, p. 777-796, 2000.

GUIMARÃES, Paulo Ricardo Bittencourt. **Métodos quantitativos estatísticos**. Curitiba: IESDE Brasil S.A. 2008.

HALADYNA, Thomas M. **Developing and validating multiple-choice test items**. Routledge, 2004.

HYDE, Janet Shibley. The gender similarities hypothesis. **American psychologist**, v. 60, n. 6, p. 581, 2005.

HYDE, Janet S.; FENNEMA, Elizabeth; LAMON, Susan J. Gender differences in mathematics performance: A meta-analysis. **Psychological bulletin**, v. 107, n. 2, p. 139, 1990.

INEP. **Exame Nacional do Ensino Médio – ENEM : documento básico 2000**. Instituto Nacional de Estudos e Pesquisas Educacionais. – Brasília: O Instituto, 1999.

INEP. **Exame Nacional do Ensino Médio (Enem): fundamentação teórico-metodológica**. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. – Brasília : O Instituto, 2005.

INEP. **Relatório pedagógico: Enem 2011-2012**. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Brasília, DF : Inep, 2015.

INEP. **Brasil no Pisa 2015 – Sumário Executivo**, Daeb, Inep, 2016. Disponível em: http://download.inep.gov.br/acoes_internacionais/pisa/documentos/2016/pisa_brasil_2015_sumario_e_xecutivo.pdf

INEP. **Brasil no Pisa 2012 – Sumário Executivo**, Daeb, Inep, 2014. Disponível em: http://download.inep.gov.br/acoes_internacionais/pisa/documentos/2014/relatorio_nacional_pisa_2012_resultados_brasileiros.pdf

KANE, Jonathan M.; MERTZ, Janet E. Debunking myths about gender and mathematics performance. **Notices of the AMS**, v. 59, n. 1, p. 10-21, 2012.

KATHEEB, Majeda. **Cognitive Load Theory and Mathematics Education**. Tese de Doutorado. Universidade de Nova Gales do Sul. Austrália. p.312. 2008.

KIRSCHNER, Paul. Cognitive load theory: implications of cognitive load theory on the design of learning. **Learning and Instruction**, v. 12, n. 1, p. 1-10, 2002.

KLINE, Theresa J. B. **Psychological testing: A practical approach to design and evaluation**. pp. 91-105, Thousand Oaks, CA. Sage. 2005.

KRUSKAL, William H.; WALLIS, W. Allen. Use of ranks in one-criterion variance analysis. **Journal of the American statistical Association**, v. 47, n. 260, p. 583-621, 1952.

LOTERO, Luis Alejandro Andrade. Teoría de la carga cognitiva, diseño multimedia y aprendizaje: un estado del arte. **Magis, Revista Internacional de Investigación en Educación**, v. 5, n. 10, 2012.

LOW, Renae; OVER, Ray. Gender differences in solution of algebraic word problems containing irrelevant information. **Journal of Educational Psychology**, v. 85, n. 2, p. 331, 1993.

MACHADO, Milene. C. **Gênero e Desempenho em Itens da Prova de Matemática do Exame Nacional do Ensino Médio (Enem):** Relações com as Atitudes e Crenças de Autoeficácia Matemática. Tese de Doutorado. Universidade Estadual de Campinas, Campinas. p.224. 2014.

MACK, Nancy K. Learning fractions with understanding: Building on informal knowledge. **Journal for research in mathematics education**, p. 16-32, 1990.

MACK, Nancy K. Learning rational numbers with understanding: The case of informal knowledge. In T. P. Carpenter, E. Fennema & T. A. Romberg (Eds.), **Rational numbers: An integration of Research**. (pp. 85-106). Hillsdale, NJ.: Lawrence Erlbaum Association, 1993.

MACK, Nancy K. Building on informal knowledge through instruction in a complex content domain: Partitioning, units, and understanding multiplication of fractions. **Journal for Research in Mathematics Education**. v. 32, n. 3, p. 267, 2001.

MAROCO, João; GARCIA-MARQUES, Teresa. Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas? **Laboratório de psicologia**, p. 65-90, 2006.

MILLER, George A. The magical number seven plus or minus two : Some limits on our capacity for processing information. **Psychological Review**, v. 63, p. 2, p. 81, 1956.

MILLER, Heather, BICHSEL, Jacqueline. Anxiety, Working Memory, Gender, and Math Performance. **Personality and Individual Differences**. v. 37, n. 3, p. 591-606, 2004.

NASCIMENTO, Adriano Rodrigues et al. Contextualização e interdisciplinaridade no "Novo ENEM": uma análise de conteúdo das questões da prova de matemática no período 2009-2016. In: Encontro Goiano de Educação Matemática, 6, 193-205, 2017, Goiás. **Anais...** Goiânia: SBEM, 2017. ISSN 2176-3305.

NIEDERLE, Muriel; VESTERLUND, Lise. Explaining the Gender Gap in Math Test Scores: The Role of Competition, **Journal of Economic Perspective**, v. 24, n. 2, p.129-144, 2010.

OECD. **PISA 2003 Technical Report**. OCDE Publishing, 2003. Disponível em: <http://www.pisa.oecd.org/>. Acesso em: maio de 2015

OECD. **PISA 2009 at a Glance**, OECD Publishing, 2011. Disponível em: <www.oecd.org/pisa/46660259.pdf>. Acesso em: 15 ago. 2014.

OECD. **PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy**, OECD Publishing, 2013.

OECD. **PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy**. OECD Publishing, 2016.

ORTIGÃO, Maria Isabel R.; AGUIAR Glauco da Silva. Letramento em Matemática no PISA. In: Seminário Internacional de Pesquisa em Educação Matemática, 2012, Petrópolis. **Anais...** Petrópolis: V SIPEM, 2012. ISBN: 978-85-98092-15-7.

OWEN, Elizabeth; SWELLER, John. What do students learn while solving mathematics problems?. **Journal of Educational Psychology**, v. 77, n. 3, p. 272, 1985.

PAAS, Fred; RENKL, Alexander; SWELLER, John. Cognitive Load Theory and Instructional Design: Recent Developments. **Educational Psychologist**, v. 38, n. 1, p. 1-4, 2003.

PAAS, Fred. et al. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. **Educational Psychologist**, v. 38, n. 1, p. 63-71, 2003.

PAAS, Fred; van GOG, Tamara; SWELLER, John. Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. **Educational Psychology Review**, v. 22, n. 2, p. 115-121, 2010.

PAAS, Fred; van MERRIENBOER, Jeroen J. G. Variability of worked examples and transfer of geometric problem-solving skills: A cognitive load approach. **Journal of Educational Psychology**, v. 86, n. 1, p. 122-133, 1994.

PALMA FILHO, João Cardoso. **Política educacional brasileira: educação brasileira numa década de incertezas (1990 -2000): avanços e retrocessos**. São Paulo: Cte Editora, 2005.

PASQUALI, Luiz. Validade dos Testes Psicológicos: Será Possível Reencontrar o Caminho? **Psicologia: Teoria e Pesquisa**, v. 23, n. especial, p. 099-107, 2007.

PASQUALI, Luiz. Psicometria. **Revista da Escola de Enfermagem da USP**, São Paulo, v. 43, n. especial, p. 992-999, 2009.

PENIN, Sônia; MARTÍNEZ, Miguel. **Profissão docente: pontos e contrapontos**. São Paulo: Summus, 2009.

PETERSON, Robert A. A meta-analysis of Cronbach's coefficient alpha. **Journal of Consumer Research**, v. 21, n. 2, p. 381-391, 1994.

RABELO, Mauro. **Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro**. Rio de Janeiro. Sociedade Brasileira de Matemática, 258p. Coleção Profmat, 2013.

RAUBER, Marcelo Fernando. **Sistema Tutor Inteligente Aplicado ao Ensino de Ciências**: uma Proposta de Arquitetura. Dissertação de Mestrado. Universidade Federal de Santa Catarina. p.158. 2016.

RODRIGUES, Marcio Urel. **Análise das questões de matemática do novo ENEM (2009 á 2012)**: reflexões para professores de matemática. Curitiba: SBEM, 2013.

de SOUZA, Nelson Pinheiro Coelho. **Teoria da Carga Cognitiva**: Origem, Desenvolvimento e Diretrizes Aplicáveis ao Processo Ensino-Aprendizagem. Dissertação de Mestrado. Universidade Federal do Pará. p.176. 2010.

SANTOS, Leila Maria Araújo; TAROUCO, Liane Margarida Rockenbach. A importância do estudo da teoria da carga cognitiva em uma educação tecnológica. **RENOTE**, v. 5, n. 1, 2007.

SPEARMAN, Charles. Correlation calculated from faulty data. **British Journal of Psychology**, v. 3, n.3, p. 271-195, 1910.

SWELLER, John. Cognitive load during problem solving: Effects on learning. **Cognitive Science**, v. 12, n. 32, p. 257-285, 1988.

SWELLER, John. Cognitive load theory, learning difficulty, and instructional design. **Learning and Instruction**, v. 4, n. 4, p. 295-312, 1994.

SWELLER, John. Evolution of human cognition architecture. **The psychology of learning and motivation**, v. 43, p. 215-266, 2003.

SWELLER, John. Element interactivity and intrinsic, extraneous, and germane cognitive load. **Educational Psychology Review**, v. 22, n. 2, p. 123-138, 2010.

SWELLER, John; CHANDLER, Paul. Evidence for cognitive load theory. **Cognition and Instruction**, v. 8, n. 4, p. 351-362, 1991.

SWELLER, John; van MERRIENBOER, Jeroen J. G.; PAAS, Fred. Cognitive architecture and instructional design. **Educational Psychology Review**, v. 10, p. 251-296, 1998.

SWELLER, John; Ayres, Paul, & Kalyuga, Slava. Measuring Cognitive Load. In: **Cognitive load theory**. p. 71-85, Springer, New York, 2011.

TAKIR, Aygil. **The Effect of an Instruction Designed by Cognitive Load Theory Principles on 7th Grade Students' Achievement in Algebra Topics and Cognitive Load**. Tese de Doutorado. The Graduate School of Social Sciences of Middle East Technical University. p.237. 2011

VIANNA, Heraldo Marelím. Validade de Construtos em Testes Educacionais. **Educação e Seleção**, n. 08, p. 35-44, 2013.

Van MERRIENBOER, Jeroen J. G.; KIRSCHNER, Paul A., & KESTER, Liesbeth. Taking the load off a learner's mind: Instructional design for complex learning. **Educational Psychologist**, v. 38, n. 1, p. 5-13, 2003.

WALKINGTON, Candace, CLINTON, Virginia, SHIVRAJ, Pooja. How Readability Factors Are Differentially Associated With Performance for Students of Different Backgrounds When Solving Mathematics Word Problems. **American Educational Research**. v. 55, n. 2, p. 362-414, 2018.

APÊNDICE I - TALE e TCLE



**Ministério da Educação
Secretaria de Educação Profissional e Tecnológica
Instituto Federal Educação, Ciência e Tecnologia de São Paulo
Comitê de Ética em Pesquisa**

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Seu (sua) filho(a) está sendo convidado para participar da pesquisa "Desempenho entre homens e mulheres em questões de matemática no ENEM que apresentam, ou não, leitura e interpretação". Os objetivos deste estudo são verificar os efeitos no desempenho de homens e mulheres, em uma avaliação de longa duração, de questões de matemática que possuem redução de carga cognitiva, que são os processos que as pessoas utilizam ao resolverem um problema. Quando um problema possui muita informação desnecessária, isso pode causar um ruído nos processos que são relevantes na resolução. A proposta de reduzir a carga cognitiva é tentar medir mais o conhecimento de matemática do estudante, e menos outras capacidades. O estudante foi selecionado pelo fato de estar no segundo ou terceiro ano do ensino médio, e sua participação não é obrigatória, nem remunerada. A qualquer momento o estudante pode desistir de participar e retirar seu consentimento. Sua recusa não trará nenhum prejuízo em sua relação com o pesquisador ou com a instituição. A participação nesta pesquisa consistirá em resolver um prova de múltipla escolha com questões de matemática. Como a pesquisa se trata de uma prova de múltipla escolha nos moldes do Enem, o risco para o aluno é baixíssimo, equivalente a fazer uma prova de 1h30 de duração. Os benefícios relacionados com a participação do aluno são: de fazer um simulado longo de matemática, portanto praticar competências e habilidades nessas condições; e contribuir para uma pesquisa sobre questões de matemática do tipo Enem. As informações obtidas através dessa pesquisa serão confidenciais e asseguramos o sigilo sobre a participação de cada aluno. Os dados não serão divulgados de forma a possibilitar a identificação uma vez que os únicos dados entregues pelos alunos são: o sexo e as respostas das questões. Sobre os dados: 1) As provas físicas ficarão com os alunos; 2) Os gabaritos serão mantidos em arquivo por 20 anos, sendo picotados e reciclados após esse tempo; 3) Os resultados digitais das escolhas nos gabaritos serão mantidos por 30 anos. Você receberá uma via deste termo onde constam o telefone e o endereço institucional do pesquisador principal e do CEP, podendo tirar suas dúvidas sobre o projeto e sua participação, agora ou a qualquer momento.

Dr. Mauricio Urban Kleinke
Orientador
E-mail: kleinke@ifp.unicamp.br
Universidade Estadual de Campinas - Campinas - SP
Rua Sergio Buarque de Holanda, 777
Telefone: (19) 3521-5350

Emiliano Augusto Chagas
Estudante de Doutorado
E-mail: emiliano@ifsp.edu.br
Instituto Federal de São Paulo - São Paulo - SP
Rua Pedro Vicente, 625 Canindé
Telefone: (11) 5579-8258

COMITÊ DE ÉTICA EM PESQUISA
Rua Pedro Vicente, 625 Canindé – São Paulo/SP
Telefone: (11) 3775-4569
E-mail: cep_ifsp@ifsp.edu.br

Declaro que entendi os objetivos, riscos e benefícios de minha participação na pesquisa e concordo em participar.

Responsável: assinatura e nome



**Ministério da Educação
Secretaria de Educação Profissional e Tecnológica
Instituto Federal Educação, Ciência e Tecnologia de São Paulo
Comitê de Ética em Pesquisa**

TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidado para participar da pesquisa "Desempenho entre homens e mulheres em questões de matemática no ENEM que apresentam, ou não, leitura e interpretação". Os objetivos deste estudo são verificar os efeitos no desempenho de homens e mulheres, em uma avaliação de longa duração, de questões de matemática que possuem redução de carga cognitiva, que são os processos que as pessoas utilizam ao resolverem um problema. Quando um problema possui muita informação desnecessária, isso pode causar um ruído nos processos que são relevantes na resolução. A proposta de reduzir a carga cognitiva é tentar medir mais o conhecimento de matemática do estudante, e menos outras capacidades. Você foi selecionado pelo fato de estar no segundo ou terceiro ano do ensino médio, e sua participação não é obrigatória, nem remunerada. A qualquer momento você pode desistir de participar e retirar seu consentimento. Sua recusa não trará nenhum prejuízo em sua relação com o pesquisador ou com a instituição. Sua participação nesta pesquisa consistirá em resolver um prova de múltipla escolha com questões de matemática. Como a pesquisa se trata de uma prova de múltipla escolha nos moldes do Enem, o risco para os participantes é baixíssimo, equivalente a fazer uma prova de 1h30 de duração. Os benefícios relacionados com a participação dos alunos são: de fazer um simulado longo de matemática, portanto praticar competências e habilidades nessas condições; e contribuir para uma pesquisa sobre questões de matemática do tipo Enem. As informações obtidas através dessa pesquisa serão confidenciais e asseguramos o sigilo sobre a participação de cada aluno. Os dados não serão divulgados de forma a possibilitar a identificação uma vez que os únicos dados entregues pelos alunos são: o sexo e as respostas das questões. Sobre os dados: 1) As provas físicas ficarão com os alunos; 2) Os gabaritos serão mantidos em arquivo por 20 anos, sendo picotados e reciclados após esse tempo; 3) Os resultados digitais das escolhas nos gabaritos serão mantidos por 30 anos. Você receberá uma via deste termo onde constam o telefone e o endereço institucional do pesquisador principal e do CEP, podendo tirar suas dúvidas sobre o projeto e sua participação, agora ou a qualquer momento.

Dr. Mauricio Urban Kleinke
Orientador
E-mail: kleinke@ifunicamp.br
Universidade Estadual de Campinas - Campinas - SP
Rua Sergio Buarque de Holanda, 777
Telefone: (19) 3521-5350

Emiliano Augusto Chagas
Estudante de Doutorado
E-mail: emiliano@ifsp.edu.br
Instituto Federal de São Paulo - São Paulo - SP
Rua Pedro Vicente, 625 Canindé
Telefone: (11) 5579-8258

COMITÊ DE ÉTICA EM PESQUISA
Rua Pedro Vicente, 625 Canindé – São Paulo/SP
Telefone: (11) 3775-4569
E-mail: cep_ifsp@ifsp.edu.br

Declaro que entendi os objetivos, riscos e benefícios de minha participação na pesquisa e concordo em participar.

Participante da pesquisa: assinatura e nome

APÊNDICE II - Folha de Respostas e Questionário Socioeconômico

LOCALXYZWA/B

AO FINAL CONFIRA SE O GABARITO ESTÁ COMPLETAMENTE PREENCHIDO
 DESTAQUE ESSA FOLHA DE GABARITO E ENTREGUE NA SAÍDA
 MARQUE TODAS AS SUAS RESPOSTAS NA PROVA PARA CONFERIR DEPOIS
 A PARTIR DE 26/10 VOCÊ PODE TER ACESSO À RESOLUÇÃO DA PROVA
 VOCÊ PODE USAR LÁPIS OU CANETA PARA PREENCHER O GABARITO

Sexo (marque com um X): Masculino Feminino

Gabarito: preencher totalmente as alternativas corretas desse modo ●

01 - (A) (B) (C) (D) (E)
 02 - (A) (B) (C) (D) (E)
 03 - (A) (B) (C) (D) (E)
 04 - (A) (B) (C) (D) (E)
 05 - (A) (B) (C) (D) (E)
 06 - (A) (B) (C) (D) (E)
 07 - (A) (B) (C) (D) (E)
 08 - (A) (B) (C) (D) (E)

09 - (A) (B) (C) (D) (E)
 10 - (A) (B) (C) (D) (E)
 11 - (A) (B) (C) (D) (E)
 12 - (A) (B) (C) (D) (E)
 13 - (A) (B) (C) (D) (E)
 14 - (A) (B) (C) (D) (E)
 15 - (A) (B) (C) (D) (E)
 16 - (A) (B) (C) (D) (E)

17 - (A) (B) (C) (D) (E)
 18 - (A) (B) (C) (D) (E)
 19 - (A) (B) (C) (D) (E)
 20 - (A) (B) (C) (D) (E)
 21 - (A) (B) (C) (D) (E)
 22 - (A) (B) (C) (D) (E)
 23 - (A) (B) (C) (D) (E)
 24 - (A) (B) (C) (D) (E)

Questionário Socioeconômico (marque com um X):

Escolaridade da mãe:

- () Não sei
- () Ensino Fundamental incompleto
- () Ensino Fundamental completo
- () Ensino Médio incompleto
- () Ensino Médio completo
- () Ensino Superior completo
- () Pós graduação

Escolaridade do pai:

- () Não sei
- () Ensino Fundamental incompleto
- () Ensino Fundamental completo
- () Ensino Médio incompleto
- () Ensino Médio completo
- () Ensino Superior completo
- () Pós graduação

Renda mensal da sua família (em salários mínimos):

- () Não sei
- () Entre 0 e 2 (até R\$ 1874,00)
- () Entre 2 e 4 (entre R\$ 1874,00 e R\$ 3748,00)
- () Entre 4 e 6 (entre R\$ 3748,00 e R\$ 5622,00)
- () Entre 6 e 8 (entre R\$ 5622,00 e R\$ 7496,00)
- () Entre 8 e 10 (entre R\$ 7496,00 e R\$ 9370,00)
- () Maior que 10 (maior que R\$ 9370,00)

APÊNDICE III - Teoria Clássica dos Testes e Alfa de Crombach

A teoria clássica dos testes se fundamenta essencialmente "nos aspectos do escore de um teste constituído de múltiplos itens" (KLINE, 2005). Variáveis de interesse, como por exemplo o sexo e escolaridade dos pais no nosso estudo, possuem interesses com relação a tomadas de decisão ou devolutivas, e essas variáveis se relacionam tanto com o escore total de um teste quanto à uma camada mais fundamental como o item, como aponta Kline (2005). Existe uma grande disponibilidade de análises disponíveis para avaliar as características individuais e globais dos itens.

Estamos interessados, para os propósitos desse estudo, particularmente em quatro indicadores de qualidade dos itens: dificuldade do item, discriminação do item, Alfa de Cronbach e distância de Cohen. Não foram realizadas análises de frequências de respostas nas alternativas pois, em um primeiro momento, não nos ajudariam a responder a pergunta de pesquisa.

A dificuldade de um item é dada pela proporção total de acertos que esse item possui, em outras palavras, é a fração de candidatos que acertaram esse item. Portanto, é um valor entre zero e um, de modo que quanto mais perto de um, mais fácil é o item para o conjunto de respondentes, e quanto mais perto de zero, mais difícil. É importante mencionar que se a dificuldade de um item de múltipla escolha com cinco alternativas, como no Enem, está próxima de 0,2 o resultado sugere que essa alternativa foi escolhida ao acaso.

Segundo Haladyna (2004), é desejável que os itens apresentem índice de dificuldade entre 0,45 e 0,75, uma vez que itens com índice de dificuldade nessa faixa impactam de maneira positiva para abrir o espectro de classificação dos alunos. Itens com dificuldade menor que 0,25 ou maior que 0,91 devem ser evitados, a menos que o construto desses itens seja essencial na presença do instrumento de avaliação.

A denominação de dificuldade de um item parece contraintuitiva, uma vez que se o valor da dificuldade do item for perto de um isso significa que ele teve uma grande taxa de acerto. Por esse motivo será mais utilizado, daqui para frente, a expressão taxa de acerto ao fazer referência à dificuldade do item.

A discriminação de um item está relacionada a capacidade de "diferenciar indivíduos com habilidades ou proficiências distintas" (RABELO, 2013). Para tal

propósito devemos, em um teste, separar os respondentes em três grupos de acordo com seu escore total, ou seja, a soma dos pontos no teste. Os grupos são: Superior, os 27% que obtiveram o maior desempenho; Inferior, os 27% de menor desempenho; Intermediário, os demais candidatos.

Em cada item analisamos o percentual de acerto dos grupos superior e inferior. Então temos para o item i o percentual de acerto do grupo superior como P_{sup_i} e para o grupo inferior como P_{inf_i} . A discriminação do item i (D_i) é dada por:

$$D_i = P_{sup_i} - P_{inf_i}$$

Veja que quanto maior for a diferença entre os dois grupos em um item, mais esse item separa indivíduos de alta e baixa proficiência e quanto mais perto de zero, menor é o potencial de discriminação do item. A preocupação para que instrumentos de avaliação não tenham itens com taxa de acerto perto dos extremos (0,2 e 0,9) se justifica nesse momento. É muito mais razoável conseguir um valor numérico de D_i consideravelmente maior que zero quando a taxa de acerto estiver entre 0,3 e 0,7. Como é de se esperar, pela própria fórmula de cálculo, o valor da discriminação de um item é um valor que na maioria das vezes está entre 0,0 e 0,5.

Segundo Ebel e Frisbie (1986), valores numéricos de discriminação de itens podem ser classificados da seguinte maneira.

Tabela 25 - Categorização da qualidade de um item pela sua discriminação

D	Qualidade	Recomendação
>0,39	Excelente	Manter
0,30 a 0,39	Bom	Sujeito a aprimoramento
0,20 a 0,29	Marginal	Sujeito a reelaboração
<0,20	Deficiente	Descartado ou profundamente reelaborado

A questão de verificar a taxa de acerto dos itens e conseqüentemente obter a discriminação de cada item está diretamente relacionada com a proposta do instrumento de avaliação. Ao construir um teste, o avaliador se propõe a verificar um conjunto de competências e habilidades (construtos) e, como já mencionado no começo do trabalho, existem tipos de avaliações, cada um com sua finalidade. Para testes classificatórios, a escolha adequada dos itens que compõem o instrumento de avaliação é fundamental, uma vez que existe uma pressuposição de que os sujeitos de pesquisa, com toda sua variabilidade de proficiência, sejam classificados da maior proficiência para a menor proficiência independente do teste aplicado.

O que está por trás dessa ideia é a chamada validade do teste, que essencialmente versa sobre que informação desejamos e como buscamos essa informação. Pasquali (2007) alerta para a confusão e variedade do conceito de validade de um teste, mas nesse trabalho vamos nos ater apenas a pergunta: o instrumento está medindo o que pretendia medir? Em outras palavras, interpretaremos validade aqui como adequação do instrumento, "ao aspecto da medida ser congruente com a propriedade medida dos objetos e não com a exatidão com que a mensuração, que descreve esta propriedade do objeto, é feita" (PASQUALI, 2009, p. 993).

A discussão de validade está intimamente ligada à questão de confiabilidade do instrumento de avaliação, ou seja, do quanto e como o instrumento foi interpretado e compreendido pelos respondentes, para finalmente respondermos ao questionamento de ser possível obter uma mesma classificação desses respondentes. Espera-se que uma avaliação seja válida e confiável se, obtivermos uma determinada classificação dos respondentes após uma aplicação do instrumento e, ao reaplicarmos esse mesmo instrumento, nas mesmas condições e para o mesmo conjunto de respondentes, com a condição de ser possível apagar da memória deles a primeira avaliação, tivéssemos a mesma classificação, ou classificação muito parecida.

Dadas todas as condicionais utilizadas no parágrafo anterior, especialmente na questão da amnésia dos respondentes durante a segunda aplicação, não temos certeza sobre a confiabilidade e consistência dos instrumentos de avaliação. O que é possível definir é uma consistência interna do instrumento, dividindo ele em duas metades, por exemplo o conjunto dos itens pares e o conjunto dos itens ímpares, e calculando a correlação entre os escores desses dois conjuntos de todos os candidatos. Espera-se que candidatos de alta proficiência tenham notas boas nos dois conjuntos, e candidatos de baixa proficiência notas menores nos dois conjuntos.

Essa é a ideia da medida de meia prova de Spearman-Brown (SPEARMAN, 1910; BROWN 1910) para confiabilidade de um instrumento de avaliação. Lembrando que aqui essa confiabilidade se traduz em consistência interna da prova. Uma possibilidade mais robusta seria tomar uma média entre todos os conjuntos de meia prova (de mesma cardinalidade) possíveis para essa mesma medida de confiabilidade. Essa medida é a que utilizaremos nesse trabalho e que por anos tem sido a mais utilizada por psicometristas, o Alfa de Cronbach (CRONBACH, 1951).

$$\alpha = \frac{k}{k-1} \times \left[1 - \frac{\sum_{j=1}^k S_j^2}{S_T^2} \right]$$

Aqui, k é o número de itens do instrumento de avaliação, S_T é o desvio padrão dos escores observados do teste e S_j o desvio padrão do item j .

Esse valor varia entre 0 e 1 e "estima quão uniformemente os itens contribuem para a soma não ponderada do instrumento" (MAROCO; GARCIA-MARQUES, 2006), quanto mais perto de 1, mais consistência interna o instrumento possui. Os autores ainda chamam a atenção que não existe uma concordância sobre valores aceitáveis de Alfa de Cronbach, uma vez que os cenários são bem distintos entre as pesquisas que se utilizam desse número.

Em uma meta análise, Peterson (1994) compila uma tabela de valores de Alfa aceitáveis sob as condições específicas para diversos autores, na tabela 26 abaixo.

Tabela 26 - Categorização de valores de Alfa de Cronbach por autores

Autor	Condição	Alfa aceitável
Davis, 1964, p. 24	Previsão individual	Acima de 0,75
	Previsão para grupos de 25-50 indivíduos	Acima de 0,5
Kaplan & Sacuzzo, 1982, p. 106	Investigação fundamental	0,7-0,8
	Investigação aplicada	0,95
Murphy & Davidsholder, 1988, p. 89	Fiabilidade inaceitável	<0,6
	Fiabilidade baixa	0,7
	Fiabilidade moderada a elevada	0,8-0,9
	Fiabilidade Elevada	>0,9
Nunnally, 1978, p. 245-246	Investigação preliminar	0,7
	Investigação fundamental	0,8
	Investigação aplicada	0,9-0,95

Fonte - Peterson, 1994

APÊNDICE IV - Estatística clássica dos testes para cada prova

3366A

<u>Number of subject</u>	<u>Number of item</u>	<u>Number of missing response</u>	<u>Score's mean</u>	<u>Score's standard deviation</u>	<u>Cronbach's Alpha</u>
154	24	0	16,461	4,666	0,818

<u>Item</u>	<u>Mean (difficulty)</u>	<u>Standard deviation</u>	<u>Correlation (discrimination)</u>
ITEM1	0,838	0,369	0,329
ITEM2	0,883	0,321	0,353
ITEM3	0,604	0,489	0,426
ITEM4	0,500	0,500	0,550
ITEM5	0,896	0,305	0,332
ITEM6	0,714	0,452	0,392
ITEM7	0,630	0,483	0,444
ITEM8	0,779	0,415	0,298
ITEM9	0,909	0,287	0,305
ITEM10	0,721	0,449	0,524
ITEM11	0,883	0,321	0,385
ITEM12	0,727	0,445	0,483
ITEM13	0,429	0,495	-0,126
ITEM14	0,506	0,500	0,454
ITEM15	0,766	0,423	0,451
ITEM16	0,656	0,475	0,231
ITEM17	0,584	0,493	0,290
ITEM18	0,630	0,483	0,245
ITEM19	0,565	0,496	0,398
ITEM20	0,422	0,494	0,324
ITEM21	0,714	0,452	0,510
ITEM22	0,701	0,458	0,518
ITEM23	0,604	0,489	0,351
ITEM24	0,799	0,401	0,277

5050A

<u>Number of subject</u>	<u>Number of item</u>	<u>Number of missing response</u>	<u>Score's mean</u>	<u>Score's standard deviation</u>	<u>Cronbach's Alpha</u>
147	24	0	16,551	4,638	0,816
<u>Item</u>	<u>Mean (difficulty)</u>	<u>Standard deviation</u>	<u>Correlation (discrimination)</u>		
ITEM1	0,810	0,393	0,479		
ITEM2	0,932	0,252	0,334		
ITEM3	0,653	0,476	0,450		
ITEM4	0,592	0,491	0,399		
ITEM5	0,789	0,408	0,090		
ITEM6	0,667	0,471	0,318		
ITEM7	0,707	0,455	0,491		
ITEM8	0,810	0,393	0,319		
ITEM9	0,864	0,343	0,081		
ITEM10	0,857	0,350	0,374		
ITEM11	0,789	0,408	0,616		
ITEM12	0,762	0,426	0,617		
ITEM13	0,401	0,490	-0,194		
ITEM14	0,456	0,498	0,305		
ITEM15	0,741	0,438	0,571		
ITEM16	0,646	0,478	0,299		
ITEM17	0,619	0,486	0,420		
ITEM18	0,463	0,499	0,290		
ITEM19	0,646	0,478	0,412		
ITEM20	0,565	0,496	0,327		
ITEM21	0,741	0,438	0,416		
ITEM22	0,667	0,471	0,460		
ITEM23	0,524	0,499	0,469		
ITEM24	0,850	0,357	0,353		

5050B

<u>Number of subject</u>	<u>Number of item</u>	<u>Number of missing response</u>	<u>Score's mean</u>	<u>Score's standard deviation</u>	<u>Cronbach's Alpha</u>
154	24	0	16,643	4,289	0,785
<u>Item</u>	<u>Mean (difficulty)</u>	<u>Standard deviation</u>	<u>Correlation (discrimination)</u>		
<i>ITEM1</i>	0,792	0,406	0,443		
<i>ITEM2</i>	0,935	0,246	0,352		
<i>ITEM3</i>	0,610	0,488	0,419		
<i>ITEM4</i>	0,558	0,497	0,408		
<i>ITEM5</i>	0,948	0,222	0,239		
<i>ITEM6</i>	0,779	0,415	0,261		
<i>ITEM7</i>	0,649	0,477	0,473		
<i>ITEM8</i>	0,786	0,410	0,221		
<i>ITEM9</i>	0,909	0,287	0,199		
<i>ITEM10</i>	0,701	0,458	0,293		
<i>ITEM11</i>	0,896	0,305	0,258		
<i>ITEM12</i>	0,805	0,396	0,428		
<i>ITEM13</i>	0,383	0,486	-0,045		
<i>ITEM14</i>	0,506	0,500	0,282		
<i>ITEM15</i>	0,753	0,431	0,469		
<i>ITEM16</i>	0,649	0,477	0,195		
<i>ITEM17</i>	0,513	0,500	0,306		
<i>ITEM18</i>	0,565	0,496	0,119		
<i>ITEM19</i>	0,643	0,479	0,487		
<i>ITEM20</i>	0,455	0,498	0,284		
<i>ITEM21</i>	0,695	0,460	0,501		
<i>ITEM22</i>	0,727	0,445	0,457		
<i>ITEM23</i>	0,571	0,495	0,402		
<i>ITEM24</i>	0,812	0,391	0,396		

6633B

<u>Number of subject</u>	<u>Number of item</u>	<u>Number of missing response</u>	<u>Score's mean</u>	<u>Score's standard deviation</u>	<u>Cronbach's Alpha</u>
154	24	0	17,286	3,861	0,740
<u>Item</u>	<u>Mean (difficulty)</u>	<u>Standard deviation</u>	<u>Correlation (discrimination)</u>		
<i>ITEM1</i>	0,838	0,369	0,347		
<i>ITEM2</i>	0,948	0,222	0,213		
<i>ITEM3</i>	0,630	0,483	0,270		
<i>ITEM4</i>	0,630	0,483	0,255		
<i>ITEM5</i>	0,922	0,268	0,142		
<i>ITEM6</i>	0,708	0,455	0,398		
<i>ITEM7</i>	0,675	0,468	0,287		
<i>ITEM8</i>	0,792	0,406	0,100		
<i>ITEM9</i>	0,942	0,235	0,190		
<i>ITEM10</i>	0,831	0,375	0,216		
<i>ITEM11</i>	0,896	0,305	0,232		
<i>ITEM12</i>	0,825	0,380	0,323		
<i>ITEM13</i>	0,422	0,494	0,090		
<i>ITEM14</i>	0,578	0,494	0,356		
<i>ITEM15</i>	0,818	0,386	0,483		
<i>ITEM16</i>	0,636	0,481	0,189		
<i>ITEM17</i>	0,675	0,468	0,264		
<i>ITEM18</i>	0,468	0,499	0,171		
<i>ITEM19</i>	0,610	0,488	0,362		
<i>ITEM20</i>	0,545	0,498	0,206		
<i>ITEM21</i>	0,747	0,435	0,357		
<i>ITEM22</i>	0,701	0,458	0,453		
<i>ITEM23</i>	0,604	0,489	0,541		
<i>ITEM24</i>	0,844	0,363	0,300		

APÊNDICE V - Desempenho de Homens e Mulheres por Item em Cada Prova

	F5050A		M5050A		F5050B		M5050B	
	Média	Dp	Média	Dp	Média	Dp	Média	Dp
Q1	0,784	0,412	0,826	0,379	0,707	0,455	0,844	0,362
Q2	0,892	0,311	0,967	0,178	0,890	0,313	0,978	0,147
Q3	0,541	0,498	0,707	0,455	0,512	0,500	0,700	0,458
Q4	0,527	0,499	0,641	0,480	0,488	0,500	0,633	0,482
Q5	0,784	0,412	0,793	0,405	0,951	0,215	0,956	0,206
Q6	0,581	0,493	0,728	0,445	0,780	0,414	0,767	0,423
Q7	0,649	0,477	0,728	0,445	0,537	0,499	0,744	0,436
Q8	0,838	0,369	0,772	0,420	0,829	0,376	0,756	0,430
Q9	0,824	0,381	0,902	0,297	0,866	0,341	0,956	0,206
Q10	0,851	0,356	0,870	0,337	0,585	0,493	0,789	0,408
Q11	0,743	0,437	0,826	0,379	0,890	0,313	0,911	0,285
Q12	0,703	0,457	0,815	0,388	0,817	0,387	0,833	0,373
Q13	0,459	0,498	0,337	0,473	0,390	0,488	0,400	0,490
Q14	0,351	0,477	0,554	0,497	0,524	0,499	0,489	0,500
Q15	0,662	0,473	0,783	0,412	0,646	0,478	0,867	0,340
Q16	0,622	0,485	0,652	0,476	0,610	0,488	0,700	0,458
Q17	0,500	0,500	0,707	0,455	0,463	0,499	0,567	0,496
Q18	0,432	0,495	0,500	0,500	0,488	0,500	0,644	0,479
Q19	0,581	0,493	0,652	0,476	0,598	0,490	0,689	0,463
Q20	0,527	0,499	0,576	0,494	0,415	0,493	0,500	0,500
Q21	0,689	0,463	0,750	0,433	0,634	0,482	0,778	0,416
Q22	0,554	0,497	0,750	0,433	0,695	0,460	0,767	0,423
Q23	0,392	0,488	0,565	0,496	0,476	0,499	0,656	0,475
Q24	0,811	0,392	0,902	0,297	0,805	0,396	0,833	0,373
Média	15,297		17,304		15,598		17,756	

	F3366A		M3366A		F6633B		M6633B	
	Média	Dp	Média	Dp	Média	Dp	Média	Dp
Q1	0,730	0,444	0,918	0,274	0,800	0,400	0,881	0,324
Q2	0,851	0,356	0,918	0,274	0,971	0,167	0,941	0,236
Q3	0,568	0,495	0,643	0,479	0,543	0,498	0,663	0,473
Q4	0,392	0,488	0,592	0,491	0,557	0,497	0,693	0,461
Q5	0,919	0,273	0,898	0,303	0,871	0,335	0,941	0,236
Q6	0,568	0,495	0,806	0,395	0,643	0,479	0,693	0,461
Q7	0,554	0,497	0,714	0,452	0,571	0,495	0,752	0,432
Q8	0,757	0,429	0,816	0,387	0,757	0,429	0,802	0,399
Q9	0,878	0,327	0,929	0,258	0,914	0,280	0,950	0,217
Q10	0,662	0,473	0,765	0,424	0,843	0,364	0,802	0,399
Q11	0,865	0,342	0,898	0,303	0,900	0,300	0,881	0,324
Q12	0,689	0,463	0,806	0,395	0,729	0,445	0,871	0,335
Q13	0,473	0,499	0,367	0,482	0,357	0,479	0,436	0,496
Q14	0,419	0,493	0,592	0,491	0,557	0,497	0,574	0,494
Q15	0,676	0,468	0,837	0,370	0,671	0,470	0,891	0,312
Q16	0,595	0,491	0,724	0,447	0,614	0,487	0,644	0,479
Q17	0,486	0,500	0,633	0,482	0,543	0,498	0,752	0,432
Q18	0,595	0,491	0,704	0,456	0,500	0,500	0,475	0,499
Q19	0,500	0,500	0,633	0,482	0,557	0,497	0,663	0,473
Q20	0,419	0,493	0,429	0,495	0,500	0,500	0,594	0,491
Q21	0,649	0,477	0,765	0,424	0,714	0,452	0,792	0,406
Q22	0,554	0,497	0,796	0,403	0,600	0,490	0,752	0,432
Q23	0,554	0,497	0,622	0,485	0,457	0,498	0,713	0,452
Q24	0,716	0,451	0,857	0,350	0,757	0,429	0,871	0,335
Média	15,068		17,663		15,929		18,030	

APÊNDICE VI - Desempenho em cada item estratificado por sexo e por tipo

F é o sexo feminino e M o sexo masculino

Desempenho no Item 1 estratificado por sexo e por tipo

Tipo	Sexo	Não Acertou	Acertou	Total
Original	F	58	168	226
	M	33	254	287
Reduzida	F	16	58	74
	M	16	75	91
Total		123	555	678

Desempenho no Item 2 estratificado por sexo e por tipo

Original	F	19	129	148
Reduzida	M	11	178	189
	F	11	141	152
	M	8	181	189
	Total	49	629	678

Desempenho no Item 3 estratificado por sexo e por tipo

Original	F	72	84	156
Reduzida	M	61	126	187
	F	66	78	144
	M	61	130	191
	Total	260	418	678

Desempenho no Item 4 estratificado por sexo e por tipo

Original	F	66	78	144
Reduzida	M	64	127	191
	F	87	69	156
	M	72	115	187
	Total	289	389	678

Desempenho no Item 5 estratificado por sexo e por tipo

Original	F	16	58	74
Reduzida	M	18	73	91
	F	19	207	226
	M	20	267	287
	Total	73	605	678

Desempenho no Item 6 estratificado por sexo e por tipo

Original	F	88	130	218
Reduzida	M	75	214	289
	F	18	64	82
	M	20	69	89
	Total	201	477	678

Desempenho no Item 7 estratificado por sexo e por tipo

Original	F	101	125	226
	M	75	212	287
Reduzida	F	26	48	74
	M	25	66	91
	Total	227	451	678

Desempenho no Item 8 estratificado por sexo e por tipo

Original	F	29	115	144
	M	40	151	191
Reduzida	F	32	124	156
	M	39	148	187
	Total	140	538	678

Desempenho no Item 9 estratificado por sexo e por tipo

Original	F	22	126	148
	M	16	173	189
Reduzida	F	17	135	152
	M	8	181	189
	Total	63	615	678

Desempenho no Item 10 estratificado por sexo e por tipo

Original	F	59	97	156
	M	42	145	187
Reduzida	F	22	122	144
	M	32	159	191
	Total	155	523	678

Desempenho no Item 11 estratificado por sexo e por tipo

Original	F	19	55	74
	M	16	75	91
Reduzida	F	26	200	226
	M	30	257	287
	Total	91	587	678

Desempenho no Item 12 estratificado por sexo e por tipo

Original	F	15	67	82
	M	14	75	89
Reduzida	F	64	154	218
	M	49	240	289
	Total	142	536	678

Desempenho no Item 13 estratificado por sexo e por tipo

Original	F	79	69	148
	M	123	66	189
Reduzida	F	95	57	152
	M	109	80	189
	Total	406	272	678

Desempenho no Item 14 estratificado por sexo e por tipo

Original	F	48	26	74
Reduzida	M	41	50	91
	F	113	113	226
	M	127	160	287
	Total	329	349	678

Desempenho no Item 15 estratificado por sexo e por tipo

Original	F	72	146	218
Reduzida	M	47	242	289
	F	29	53	82
	M	11	78	89
	Total	159	519	678

Desempenho no Item 16 estratificado por sexo e por tipo

Original	F	89	137	226
Reduzida	M	89	198	287
	F	28	46	74
	M	32	59	91
	Total	238	440	678

Desempenho no Item 17 estratificado por sexo e por tipo

Original	F	82	74	156
Reduzida	M	74	113	187
	F	69	75	144
	M	52	139	191
	Total	277	401	678

Desempenho no Item 18 estratificado por sexo e por tipo

Original	F	72	84	156
Reduzida	M	60	127	187
	F	77	67	144
	M	98	93	191
	Total	307	371	678

Desempenho no Item 19 estratificado por sexo e por tipo

Original	F	31	43	74
Reduzida	M	32	59	91
	F	101	125	226
	M	97	190	287
	Total	261	417	678

Desempenho no Item 20 estratificado por sexo e por tipo

Original	F	91	65	156
Reduzida	M	100	87	187
	F	70	74	144
	M	80	111	191
	Total	341	337	678

Desempenho no Item 21 estratificado por sexo e por tipo

Original	F	56	100	156
Reduzida	M	42	145	187
	F	43	101	144
	M	44	147	191
	Total	185	493	678

Desempenho no Item 22 estratificado por sexo e por tipo

Original	F	66	82	148
Reduzida	M	43	146	189
	F	53	99	152
	M	46	143	189
	Total	208	470	678

Desempenho no Item 23 estratificado por sexo e por tipo

Original	F	114	112	226
Reduzida	M	97	190	287
	F	45	29	74
	M	40	51	91
	Total	296	382	678

Desempenho no Item 24 estratificado por sexo e por tipo

Original	F	37	119	156
Reduzida	M	28	159	187
	F	31	113	144
	M	22	169	191
	Total	118	560	678

APÊNDICE VII - Análise dos outros Itens da Prova

O item 2, em suas duas versões, está no quadro 19 abaixo. Assim como no item 1, a redução de carga cognitiva estranha nesse item também proporcionou o efeito de deixar o desempenho das mulheres estatisticamente igual ao dos homens. Como o quadro 19 a seguir nos mostra, o item original apresenta uma RC próximo de 2,4 para o acerto de homens em comparação ao acerto das mulheres de maneira significativa. Já o item reduzido possui uma RC próxima de 1,7 com um p-valor $> 0,05$, ou seja, sem diferença estatística. Se olharmos apenas o desempenho de todos os estudantes entre o item original e o reduzido temos uma RC perto de 1,7, portanto os estudantes tiveram uma tendência de score maior na versão reduzida, muito embora o p-valor seja ligeiramente maior do que 0,05.

Quadro 19 - Item 2 da Prova e sua respectiva versão Reduzida

<u>Item 2 - Original</u> (Enem 2011 - 138)	<u>Item 2 - Reduzido</u>
<p>O dono de uma oficina mecânica precisa de um pistão das partes de um motor, de 68 mm de diâmetro, para o conserto de um carro. Para conseguir um, esse dono vai até um ferro velho e lá encontra pistões com diâmetros iguais a 68,21 mm; 68,102 mm; 68,001 mm; 68,02 mm e 68,012 mm. Para colocar o pistão no motor que está sendo consertado, o dono da oficina terá de adquirir aquele que tenha o diâmetro mais próximo do que precisa.</p> <p>Nessa condição, o dono da oficina deverá comprar o pistão de diâmetro</p> <p>a) 68,21 mm. b) 68,102 mm. c) 68,02 mm. d) 68,012 mm. e) 68,001 mm.</p>	<p>Um motor utiliza pistões com 68 mm de diâmetro.</p> <p>Qual dos pistões da lista abaixo possui o diâmetro mais próximo de 68 mm?</p> <p>a) 68,21 mm. b) 68,102 mm. c) 68,02 mm. d) 68,012 mm. e) 68,001 mm.</p>

Quadro 20 - Valores de Razão de Chance e p-valor para o Item 2 por tipo de item e sexo, e taxa de acerto por sexo.

	RC	p-valor
Tipo de Item x Desempenho	1,656	0,127
Acerto x Sexo - Original	2,383	0,04
Acerto x Sexo - Reduzida	1,765	0,335

Gráfico de barras empilhadas mostrando a proporção de acerto por sexo (F e M) para o item Original e Reduzido. O eixo Y representa a proporção de 0,00 a 1,00. Para o item Original, a proporção de acerto é maior para os homens (M) do que para as mulheres (F). Para o item Reduzido, as proporções de acerto são semelhantes para ambos os sexos.

O contexto dessa questão original é também mais masculino e novamente aborda carros com palavras como, oficina mecânica, pistão e motor. Na versão original do problema, a taxa de acerto das mulheres (0,87) era consideravelmente menor que a dos homens (0,94) e na versão reduzida as mulheres (0,93) empataram estatisticamente com o desempenho dos homens (0,96). Avaliamos que o construto pretendido no item está associado a comparar números na forma decimal em uma situação contextualizada.

Efetuamos duas mudanças estruturais no item: a primeira, de maneira mais sutil foi, retirar algumas palavras mais masculinas do problema, como carro e oficina mecânica, deixamos apenas motor e pistão; a segunda mudança foi mais substancial, removemos uma parte do contexto, deixando apenas elementos básicos para a comparação dos diâmetros, desse modo acreditamos que foi retirada uma grande quantidade de carga cognitiva estranha, que prejudicava o desempenho dos estudantes. Para esse item, utilizamos as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que eliminamos algumas palavras que podiam acarretar carga estranha e também simplificamos a linguagem; e Extirpação, já que boa parte do texto associado ao contexto foi removido.

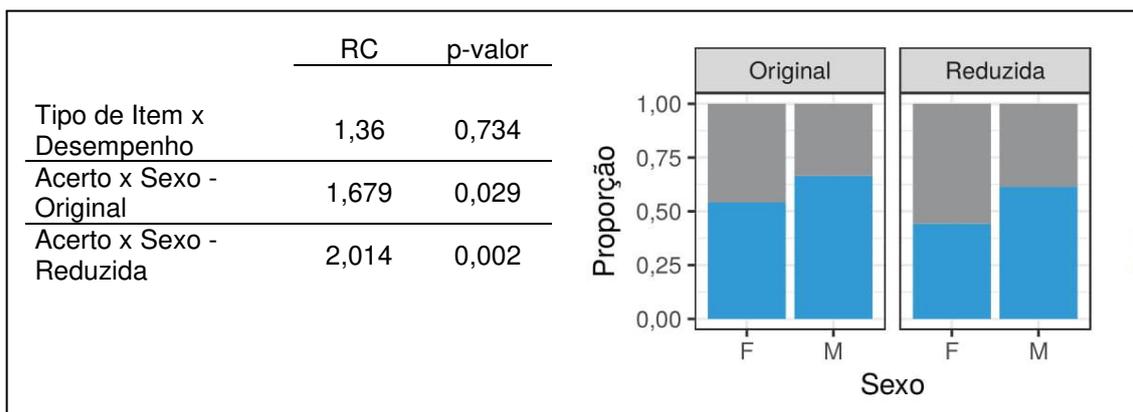
O item 4, em suas duas versões, está localizado no quadro 21 a seguir. Assim como o item 3, a redução de carga cognitiva estranha não proporcionou o efeito de melhoria do desempenho das mulheres em relação aos homens, bem como a redução de carga cognitiva estranha não impactou significativamente a taxa global de acerto. Como é mostrado no quadro 23 a seguir, o item original apresenta RC perto de 1,7 para o acerto dos homens em relação ao das mulheres, já o item reduzido possui RC próximo de 2,0, ambos com diferença estatística significativa. A RC entre a questão original e reduzida é de aproximadamente 0,7, em outras palavras, o efeito da redução de carga cognitiva teve uma tendência de deteriorar o desempenho global dos estudantes.

Acreditamos que o item original não possui viés em relação ao sexo, uma vez que o contexto lida com medidas de um terreno. O construto do item afere sobre área do retângulo em um problema contextualizado. Para a versão original do problema, a taxa de acerto das mulheres (0,54) foi consideravelmente menor que a dos homens (0,67), e para a versão reduzida, o desempenho das mulheres (0,44) foi significativamente menor que o dos homens (0,61).

Quadro 21 - Item 4 da Prova e sua respectiva versão Reduzida

Item 4 - Original (Enem 2011 - 142)	Item 4 - Reduzido
<p>Em uma certa cidade, os moradores de um bairro carente de espaços de lazer reivindicam à prefeitura municipal a construção de uma praça. A prefeitura concorda com a solicitação e afirma que irá construí-la em formato retangular devido às características técnicas do terreno. Restrições de natureza orçamentária impõem que sejam gastos, no máximo, 180m de tela para cercar a praça. A prefeitura apresenta aos moradores desse bairro as medidas dos terrenos disponíveis para a construção da praça:</p> <p>Terreno 1: 55 m por 45 m, Terreno 2: 55 m por 55 m, Terreno 3: 60 m por 30 m Terreno 4: 70 m por 20 m, Terreno 5: 95 m por 85 m</p> <p>Para optar pelo terreno de maior área, que atenda às restrições impostas pela prefeitura, os moradores deverão escolher o terreno.</p> <p>a) 1 b) 2 c) 3 d) 4 e) 5</p>	<p>Dentre os terrenos retangulares abaixo, qual apresenta <u>maior área</u> com um perímetro de 180m:</p> <p>a) Terreno 1: 55m por 45m b) Terreno 2: 55m por 55m c) Terreno 3: 60m por 30m d) Terreno 4: 70m por 20m e) Terreno 5: 95m por 85m</p>

Quadro 22 - Valores de Razão de Chance e p-valor para o Item 4 por tipo de item e sexo, e taxa de acerto por sexo.



No item original, o enunciado apresenta uma situação que motiva a resolução do problema. Nesse texto existem muitas informações que não nos pareciam ser relevantes para o construto pretendido, como mencionar a existência de um bairro carente em uma cidade. Algumas partes do texto nos levaram a crer que a carga cognitiva estranha poderia ser muito alta, como por exemplo: reivindicar, solicitar e características técnicas. Para o item reduzido removemos todo o contexto, e mudamos o formato dos distratores.

Para o item 4 utilizamos as estratégias de redução de carga cognitiva estranha de: Tradução, já que eliminamos muitas palavras e também simplificamos a linguagem; Extirpação, uma vez que grande parte do texto associado ao contexto foi removido; e

Estética, já que as alternativas foram modificadas de maneira mais adequada para a resolução.

O item 7 em suas duas versões está localizado no quadro 23 abaixo. A redução de carga cognitiva estranha no item 7 apresentou o efeito de igualar o desempenho das mulheres com o dos homens. Como mostra o quadro 25 a seguir, o item original apresenta RC maior que 2,2 para o acerto dos homens em relação ao das mulheres, de modo significativo. Já o item reduzido possui RC próximo de 1,4 sem diferença estatística significativa. A RC entre a questão original e reduzida é de aproximadamente 1,1 e apresenta um p-valor bem maior que a significância, em outras palavras, o efeito da redução de carga cognitiva não fez um efeito global nos estudantes, mas melhorou o desempenho das mulheres.

Quadro 23 - Item 7 da Prova e sua respectiva versão Reduzida

<u>Item 7 - Original</u> (Enem 2011 - 164)	<u>Item 7 - Reduzido</u>
<p>Cerca de 20 milhões de brasileiros vivem na região coberta pela caatinga, em quase 800 mil km² de área. Quando não chove, o homem do sertão e sua família precisam caminhar quilômetros em busca da água dos açudes. A irregularidade climática é um dos fatores que mais interferem na vida do sertanejo.</p> <p>Disponível em: http://www.wwf.org.br. Acesso em: 23 abr. 2010.</p> <p>Segundo este levantamento, a densidade demográfica da região coberta pela caatinga, em habitantes por km², é de</p> <p>a) 250 b) 25 c) 2,5 d) 0,25 e) 0,025</p>	<p>Considere uma população de 20 milhões de pessoas vivendo em uma área de 800 mil km². O número de habitantes por km² é:</p> <p>a) 250 b) 25 c) 2,5 d) 0,25 e) 0,025</p>

Quadro 24 - Valores de Razão de Chance e p-valor para o Item 7 por tipo de item e sexo, e taxa de acerto por sexo

	RC	p-valor
Tipo de Item x Desempenho	1,167	0,478
Acerto x Sexo - Original	2,284	<0,001
Acerto x Sexo - Reduzida	1,43	0,374

Gráfico de barras empilhadas mostrando a proporção de acerto por sexo (F e M) para as versões Original e Reduzida do item 7. O eixo Y representa a proporção de 0,00 a 1,00. O eixo X representa o sexo (F e M). Para a versão Original, a proporção de acerto é maior para os homens (M) do que para as mulheres (F). Para a versão Reduzida, a proporção de acerto é semelhante para ambos os sexos.

Um efeito que deve ser mencionado é que a na versão original do problema, a taxa de acerto das mulheres (0,55) era consideravelmente menor que a dos homens (0,74) e na versão reduzida as mulheres (0,65) tiveram um desempenho próximo dos homens (0,73), e que esses essencialmente tiveram o mesmo desempenho, tanto no item original quanto na sua versão reduzida.

Esse efeito foi parecido com o que aconteceu no Item 1, mas diferentemente do que ocorreu naquele caso, em que o contexto se mostrava mais favorável ao sexo masculino, nesse caso temos uma situação neutra em relação aos sexos, onde os estudantes são apresentados à densidade demográfica na caatinga. Nesse sentido acreditamos que o construto do item afere sobre o cálculo de densidade demográfica à partir de um contexto.

Para o item reduzido removemos o contexto, deixando apenas os elementos básicos, explicitando a área e a população, para a sua resolução. Utilizamos as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que eliminamos muitas palavras e simplificamos a linguagem também; e Extirpação, uma vez que parte do texto associado ao contexto foi removido.

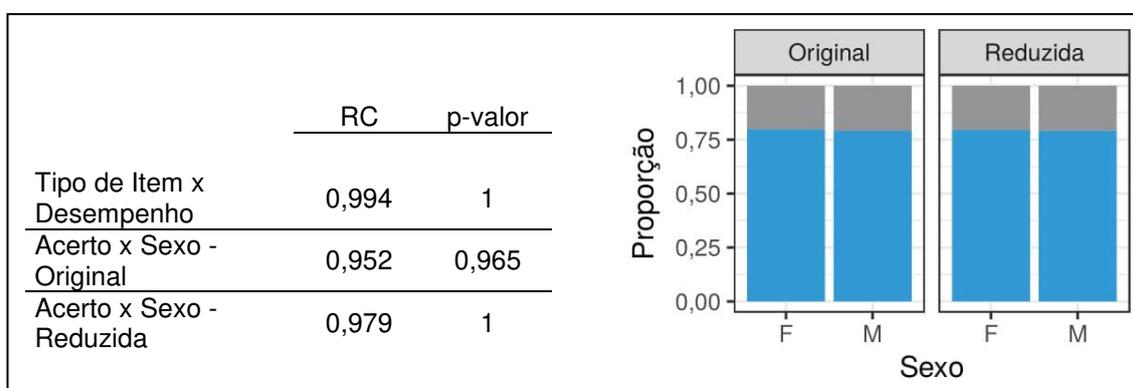
Quadro 25 - Item 8 da Prova e sua respectiva versão Reduzida

<u>Item 8 - Original</u> (Enem 2011 - 171)	<u>Item 8 - Reduzido</u>
<p>Nos últimos cinco anos, 32 mil mulheres de 20 a 24 anos foram internadas nos hospitais do SUS por causa de AVC. Entre os homens da mesma faixa etária, houve 28 mil internações pelo mesmo motivo.</p> <p>Época. 26 abr. 2010 (adaptado).</p> <p>Suponha que, nos próximos cinco anos, haja um acréscimo de 8 mil internações de mulheres e que o acréscimo de internações de homens por AVC ocorra na mesma proporção. De acordo com as informações dadas, o número de homens que seriam internados por AVC, nos próximos cinco anos, corresponderia a</p> <p>a) 4 mil b) 9 mil c) 21 mil d) 35 mil e) 39 mil</p>	<p>Em uma cidade, 32 mil mulheres e 28 mil homens foram hospitalizados no ano de 2017. Suponha que em 2018, ocorra um acréscimo de 8 mil mulheres hospitalizadas.</p> <p>Considerando que ocorra um acréscimo de homens hospitalizados na mesma proporção, qual o total de homens hospitalizados em 2018?</p> <p>a) 4 mil b) 9 mil c) 21 mil d) 35 mil e) 39 mil</p>

O item 8 em suas duas versões está no quadro 25 acima. A redução de carga cognitiva estranha nesse item não apresentou efeito de melhoria do desempenho das mulheres em relação aos homens, além disso não ocorreu impacto significativo na taxa global de acerto. Como mostra o quadro 27 a seguir, o item original apresenta RC perto

de 1,0 para o acerto dos homens em relação ao das mulheres, já o item reduzido também possui RC próximo de 1,0, ambos sem diferença estatística significativa. A RC entre a questão original e reduzida novamente é de próxima de 1,0, em outras palavras, o efeito da redução de carga cognitiva não alterou o desempenho global do estudantes. Para a versão original problema, a taxa de acerto das mulheres (0,80) foi igual ao dos homens (0,79), para a versão reduzida, o desempenho das mulheres (0,79) também foi igual ao dos homens (0,79).

Quadro 26 - Valores de Razão de Chance e p-valor para o Item 8 por tipo de item e sexo, e taxa de acerto por sexo



O item se apresenta neutro em relação ao viés entre os sexos, uma vez que ele aborda uma questão de saúde que envolve homens e mulheres. O construto do item afere sobre razão e proporção aplicada a um contexto. No item original o enunciado apresenta um texto com informações relevantes à resolução do problema, mas nesse texto existem muitas informações que não são necessárias para aferir o construto pretendido, como as siglas AVC e SUS, que devem exercer carga cognitiva estranha aos respondentes.

Para o item reduzido essas siglas foram removidas do corpo do item, mas o mesmo contexto foi mantido. Foram utilizadas as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que eliminamos muitas palavras e também simplificamos a linguagem; e Extirpação, uma vez que parte do texto associado ao contexto foi removido.

No item 11, apresentado em suas duas versões no quadro 27 a seguir, não houve discrepância entre taxa de acerto de homens e mulheres, tanto na versão original quanto na versão reduzida. Como mostra o quadro 29 a seguir, o item original apresenta RC

maior que 1,6 para o acerto dos homens em relação ao das mulheres, sem diferença estatística, já o item reduzido possui RC próximo de 1,1 entre os sexos, também sem diferença estatística. Em geral os estudantes acertaram mais o item reduzido em comparação ao item original, com uma RC próxima de 2,2 e com um p-valor bem abaixo do nível de significância.

Quadro 27 - Item 11 da Prova e sua respectiva versão Reduzida

<u>Item 11 - Original</u> (Enem 2012 - 156)	<u>Item 11 - Reduzido</u>
Nos <i>shopping centers</i> costumam existir parques com vários brinquedos e jogos. Os usuários colocam créditos em um cartão, que são descontados por cada período de tempo de uso dos jogos. Dependendo da pontuação da criança no jogo, ela recebe um certo número de tíquetes para trocar por produtos nas lojas dos parques. Suponha que o período de uso de um brinquedo em certo <i>shopping</i> custa R\$ 3,00 e que uma bicicleta custa 9 200 tíquetes. Para uma criança que recebe 20 tíquetes por período de tempo que joga, o valor, em reais, gasto com créditos para obter a quantidade de tíquetes para trocar pela bicicleta é	Uma pessoa compra 20 tíquetes com R\$ 3,00. Para adquirir 9200 tíquetes ela deve pagar, em reais:
a) 153	a) 153
b) 460	b) 460
c) 1218	c) 1218
d) 1380	d) 1380
e) 3066	e) 3066

Quadro 28 - Valores de Razão de Chance e p-valor para o Item 11 por tipo de item e sexo, e taxa de acerto por sexo

	RC	p-valor
Tipo de Item x Desempenho	2,197	0,001
Acerto x Sexo - Original	1,619	0,283
Acerto x Sexo - Reduzida	1,114	0,813

Item	Sexo	Proporção de Acerto
Original	F	0,74
	M	0,83
Reduzida	F	0,89
	M	0,90

Na versão original do item, a taxa de acerto das mulheres (0,74) foi menor do que a dos homens (0,83), embora estatisticamente igual, e na versão reduzida as mulheres (0,89) empataram o desempenho com os homens (0,9). A redução de carga cognitiva estranha mostrou uma leve tendência de melhorar o escore feminino.

Sobre o construto do item, entendemos que é obter, à partir de um contexto, o valor de uma operação aritmética. Não verificamos viés de sexo para esse item, uma vez

que o problema aborda um contexto neutro sobre brinquedo em shopping center. Para o item reduzido removemos quase todo o contexto, e junto com ele informações irrelevantes para a resolução do problema, como a explicação sobre carregar créditos no cartão.

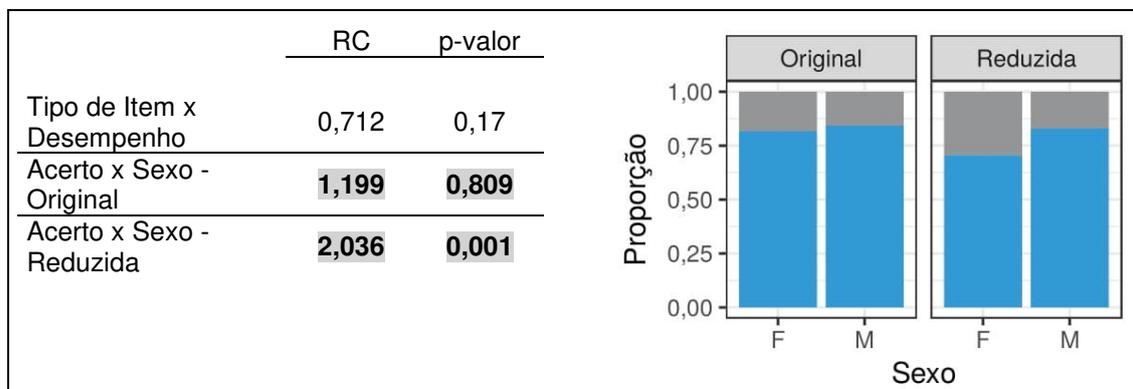
Foram feitas duas mudanças estruturais no item: no comando do item original a pergunta sobre o valor gasto em reais não está direta, o que pode acarretar confusão aos respondentes, por esse motivo o comando do item reduzido está bem direto; e a redução de carga cognitiva estranha relativa ao período de tempo mencionado no item original. Esse período de tempo se mostrou uma variável a mais como elemento cognitivo, que foi retirado na versão reduzida. Utilizamos as seguintes técnicas de redução de carga cognitiva estranha: Tradução, uma vez que a contagem de palavras foi drasticamente reduzida; Extirpação, já que recortamos partes inteiras do corpo do item; e Estética, pois entendemos que o comando não estava muito claro na versão original, e as informações foram colocadas de maneira mais natural para a resolução dos estudantes.

O item 12, em suas duas versões, está apresentado no quadro 29 abaixo. Nesse item, o desempenho das mulheres foi pior que o dos homens quando comparamos a questão original da sua versão reduzida. Como mostra o quadro 31 a seguir, o item original apresenta RC próximo de 1,2 para o acerto dos homens em relação ao das mulheres, sem diferença estatística, já o item reduzido possui RC próximo de 2,0 com um p-valor bem abaixo da significância. A RC entre a questão original e reduzida é de aproximadamente 0,7 e com um p-valor de 0,17, portanto apresenta uma tendência de que a versão reduzida prejudicou globalmente os estudantes.

Quadro 29 - Item 12 da Prova e sua respectiva versão Reduzida

<u>Item 12 - Original</u> (Enem 2012 - 157)	<u>Item 12 - Reduzido</u>
<p>João decidiu contratar os serviços de uma empresa por telefone através do SAC (Serviço de Atendimento ao Consumidor). O atendente ditou para João o número de protocolo de atendimento da ligação e pediu que ele anotasse. Entretanto, João não entendeu um dos algarismos ditados pelo atendente e anotou o número 13□98207, sendo que o espaço vazio é o do algarismo que João não entendeu.</p> <p>De acordo com essas informações, a posição ocupada pelo algarismo que falta no número de protocolo é a de</p> <ul style="list-style-type: none"> a) centena b) dezena de milhar c) centena de milhar d) milhão e) centena de milhão 	<p>Considere o número 13_98207, de oito algarismos, onde um deles está vazio.</p> <p>No lugar vazio se encontra um algarismo na posição de:</p> <ul style="list-style-type: none"> a) centena b) dezena de milhar c) centena de milhar d) milhão e) centena de milhão

Quadro 30 - Valores de Razão de Chance e p-valor para o Item 12 por tipo de item e sexo, e taxa de acerto por sexo



A redução de carga cognitiva estranha não reduziu o desempenho dos homens na versão reduzida em comparação com o item original, mas prejudicou mais as mulheres. Os índices de acerto nas duas versões do item foram altos, na versão original do problema, a taxa de acerto das mulheres (0,82) era equivalente a taxa de acerto dos homens (0,83), mas na versão reduzida as mulheres (0,71) tiveram um decréscimo considerável de desempenho, enquanto que os homens (0,83) tiveram praticamente o mesmo desempenho.

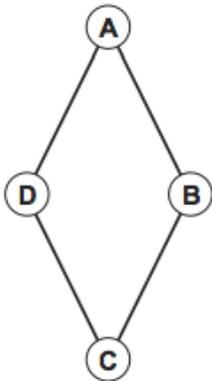
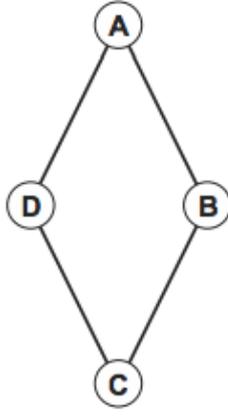
O construto pretendido pelo item, ao nosso ver, é encontrar através de um contexto, uma posição numérica no sistema decimal. O item original apresenta uma situação onde é plausível conectar o conteúdo matemático com uma situação cotidiana, mas a carga cognitiva estranha no corpo do item é grande, com a apresentação de dois personagens e uma sigla (SAC) a ser explicada. Utilizamos o mesmo posicionamento adotado no item 10, em que excluímos a contextualização devido ao grande número de acertos do item original.

A redução de carga cognitiva estranha para esse item está em suprimir os elementos que não são relevantes, para esse propósito foram utilizadas as estratégias de: Tradução, uma vez que a contagem de palavras foi bem reduzida, e Extirpação, ao remover elementos que não contribuíam ao construto pretendido.

O item 13, apresentado em suas duas versões no quadro 31 a seguir, apresentou uma inversão de desempenho entre os homens e as mulheres. No item original as mulheres tiveram um escore melhor que os homens, na versão reduzida a situação se inverteu. Como está apresentado no quadro 33 a seguir, o item original possui RC perto de 0,6 para o acerto dos homens em relação ao das mulheres, de maneira significativa,

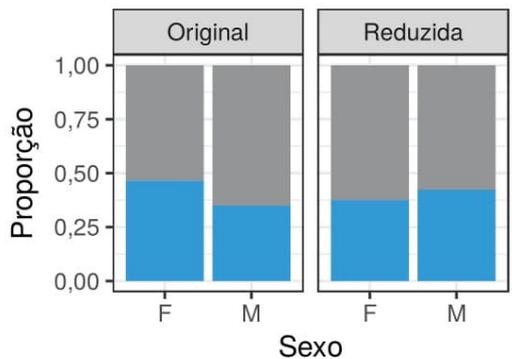
enquanto que na versão reduzida a RC ficou próxima de 1,2, sem diferença estatística significativa. A RC entre a questão original e reduzida é de aproximadamente 1,0, em outras palavras, o efeito da redução de carga cognitiva não alterou o desempenho global do estudantes.

Quadro 31 - Item 13 da Prova e sua respectiva versão Reduzida

Item 13 - Original (Enem 2013 - 161)	Item 13 - Reduzido
<p>Um artesão de joias tem à sua disposição pedras brasileiras de três cores: vermelhas, azuis e verdes. Ele pretende produzir joias constituídas por uma liga metálica, a partir de um molde no formato de um losango não quadrado com pedras nos seus vértices, de modo que dois vértices consecutivos tenham sempre pedras de cores diferentes. A figura ilustra uma joia, produzida por esse artesão, cujos vértices A, B, C e D correspondem às posições ocupadas pelas pedras.</p>	<p>A figura ilustra um losango de vértices A, B, C e D.</p>
	
<p>Joia com pedras (Foto: Enem)</p> <p>Com base nas informações fornecidas, quantas joias diferentes, nesse formato, o artesão poderá obter?</p> <p>a) 6 b) 12 c) 18 d) 24 e) 36</p>	<p>Se cada vértice desse losango deve ser pintado de vermelho, azul ou verde, e dois vértices consecutivos tenham que ter sempre cores distintas, quantas formas possíveis existem?</p> <p>a) 6 b) 12 c) 18 d) 24 e) 36</p>

Quadro 32 - Valores de Razão de Chance e p-valor para o Item 13 por tipo de item e sexo, e taxa de acerto por sexo

Item 13	RC	p-valor
Tipo de Item x Desempenho	1,005	1
Acerto x Sexo - Original	0,614	0,039
Acerto x Sexo - Reduzida	1,223	0,428



Para a versão original do item, a taxa de acerto das mulheres (0,47) foi ligeiramente maior que a dos homens (0,35), para a versão reduzida, o desempenho das mulheres (0,37) foi ligeiramente menor que a dos homens (0,42). Esses valores, juntamente com os testes de hipótese, sugerem que a redução de carga cognitiva estranha foi prejudicial para o escore feminino.

O item nos parece neutro em relação à possibilidade de viés entre os sexos, uma vez que ele aborda o número de joias construídas por um artesão. O construto do item afere sobre contagem aplicada a um contexto. No item original o enunciado apresenta um texto com informações relevantes à resolução do problema, mas nesse texto existem muitas informações que para nós poderiam elevar a carga cognitiva estranha, como por exemplo: liga metálica, molde e losango não quadrado.

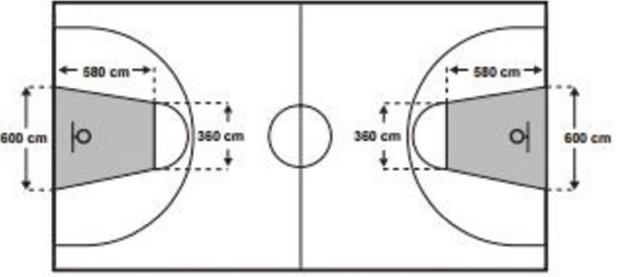
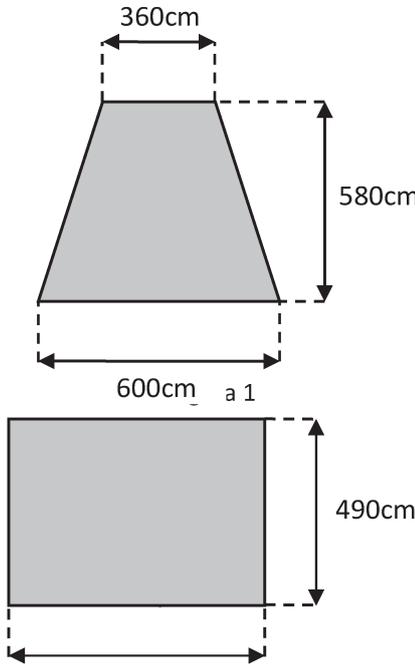
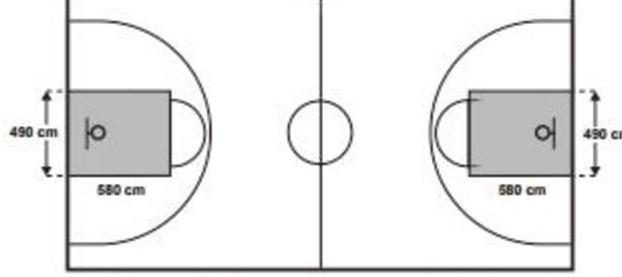
Para o item reduzido essas partes foram removidas do corpo do item, uma vez que decidimos retirar o contexto do número de joias, perguntando apenas o número de figuras que é possível obter. Utilizamos as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que eliminamos muitas palavras e também simplificamos a linguagem; e Extirpação, uma vez que parte do texto associado ao contexto foi removido.

O item 14 está apresentado, em suas duas versões, no quadro 33 a seguir. A redução de carga cognitiva estranha nesse item, proporcionou o efeito de igualar o desempenho das mulheres ao dos homens. Como mostra o quadro 34 a seguir, o item original apresenta RC maior que 2,3 para o acerto dos homens em relação ao das mulheres, com diferença estatística significativa. Já o item reduzido possui RC próximo de 1,3 sem diferença estatística significativa. A RC entre a questão original e reduzida é de aproximadamente 1,3, com um p-valor de 0,13, sugerindo que a redução de carga cognitiva possa ter uma tendência de melhorar o escore global dos estudantes.

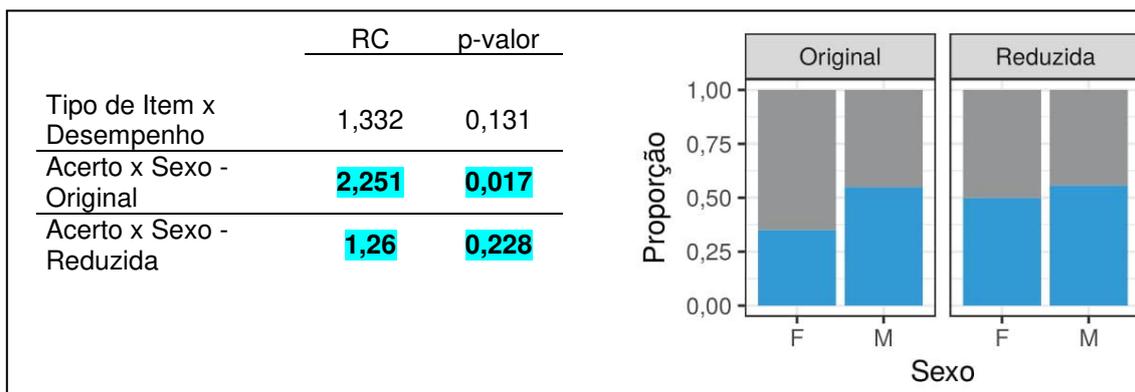
Um efeito que deve ser mencionado é que na versão original do problema, a taxa de acerto das mulheres (0,35) era consideravelmente menor que a dos homens (0,55) e na versão reduzida as mulheres (0,5) tiveram um desempenho próximo dos homens (0,55), que por sua vez mesmo desempenho, tanto no item original quanto na sua versão reduzida. A tendência de que a redução de carga cognitiva estranha tenha afetado globalmente os estudantes se deve exclusivamente às mulheres.

Acreditamos que o item original possui um pequeno viés masculino, uma vez que o contexto versa sobre basquete. O construto do item afere sobre a diferença de áreas entre duas figuras geométricas, um retângulo e um trapézio, em um contexto. No item original a explicação sobre o que são os garrafões, o que é a sigla FIBA e a motivação do cálculo, exercem uma carga cognitiva estranha desnecessária nos respondentes.

Quadro 33 - Item 14 da Prova e sua respectiva versão Reduzida

Item 14 - Original (Enem 2015 - 161)	Item 14 - Reduzido
<p>O Esquema I mostra a configuração de uma quadra de basquete. Os trapézios em cinza, chamados de garrafões, correspondem a áreas restritivas.</p>	<p>Considere o trapézio hachurado na figura 1 e o retângulo hachurado na figura 2 a seguir.</p>
 <p>Esquema I: área restritiva antes de 2010</p>	 <p>Figura 2</p>
<p>Visando atender as orientações do Comitê Central da Federação Internacional de Basquete (Fiba) em 2010, que unificou as marcações das diversas ligas, foi prevista uma modificação nos garrafões das quadras, que passariam a ser retângulos, como mostra o Esquema II.</p>	<p>Comparando as áreas das figuras, a área da figura 2 (em relação a figura 1), possui um(a):</p> <p>a) centena b) dezena de milhar c) centena de milhar d) milhão e) centena de milhão</p>
 <p>Esquema II: área restritiva a partir de 2010</p>	
<p>Após executadas as modificações previstas, houve uma alteração na área ocupada por cada garrafão, que corresponde a um(a)</p> <p>a) centena b) dezena de milhar c) centena de milhar d) milhão e) centena de milhão</p>	

Quadro 34 - Valores de Razão de Chance e p-valor para o Item 14 por tipo de item e sexo, e taxa de acerto por sexo



Para o item reduzido, tomamos a decisão de remover o contexto, os alunos tiveram apenas que calcular a diferença de área entre as duas figuras. Utilizamos as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que eliminamos muitas palavras e simplificamos a linguagem; Extirpação, já que parte do texto associado ao contexto foi removido e eliminamos efeitos visuais desnecessários, no caso o resto da quadra de basquete; e Estética, ao apresentar apenas as figuras desejadas, melhores dispostas para os respondentes.

Para o item 15, apresentado em suas duas versões no quadro 35 a seguir, a redução de carga cognitiva estranha não provocou efeito de melhoria do desempenho das mulheres em relação aos homens, bem como essa redução de carga cognitiva não teve impacto significativo na taxa global de acerto. Como mostra o quadro 35 a seguir, o item original apresenta RC perto de 2,5 para o acerto dos homens em relação ao das mulheres, já o item reduzido possui RC próximo de 3,9, com um desempenho ainda melhor dos homens, ambos com p-valor menor que 0,001. A RC entre a questão original e reduzida é de aproximadamente 1,0, em outras palavras, o efeito da redução de carga cognitiva não alterou o desempenho global do estudantes.

Para a versão original problema, a taxa de acerto das mulheres (0,67) foi consideravelmente inferior a dos homens (0,84), e para a versão reduzida, a diferença de desempenho entre as mulheres (0,65) e os homens (0,87), ficou ainda maior. Os escores desses dois grupos foram muito parecidos comparando as duas versões.

Quadro 35 - Item 15 da Prova e sua respectiva versão Reduzida

<u>Item 15 - Original</u> (Enem 2013 - 148)			<u>Item 15 - Reduzido</u>		
Cinco empresas de gêneros alimentícios encontram-se à venda. Um empresário, almejando ampliar os seus investimentos, deseja comprar uma dessas empresas. Para escolher qual delas irá comprar, analisa o lucro (em milhões de reais) de cada uma delas, em função de seus tempos (em anos) de existência, decidindo comprar a empresa que apresente o maior lucro médio anual.			O quadro apresenta o lucro (em milhões de reais) acumulado ao longo do tempo (em anos) de existência de cada empresa.		
O quadro apresenta o lucro (em milhões de reais) acumulado ao longo do tempo (em anos) de existência de cada empresa.					
Empresa	Lucro (em milhões de reais)	Tempo (em anos)	Empresa	Lucro (em milhões de reais)	Tempo (em anos)
F	24	3,0	A	24	3,0
G	24	2,0	B	24	2,0
H	25	2,5	C	25	2,5
M	15	1,5	D	15	1,5
P	9	1,5	E	9	1,5
O empresário decidiu comprar a empresa			A empresa que possui o maior lucro médio anual é:		
a) F			a) A		
b) G			b) B		
c) H			c) C		
d) M			d) D		
e) P			e) E		

Quadro 36 - Valores de Razão de Chance e p-valor para o Item 15 por tipo de item e sexo, e taxa de acerto por sexo

Item 15	RC	p-valor
Tipo de Item x Desempenho	1,004	1
Acerto x Sexo - Original	2,539	<0,001
Acerto x Sexo - Reduzida	3,88	0,001

Sexo	Original (Proporção)	Reduzida (Proporção)
F	~0,65	~0,65
M	~0,85	~0,85

O item nos parece neutro em relação à possibilidade de viés entre os sexos, uma vez que ele aborda lucro de empresa. O construto do item afere sobre média aplicada a

um contexto. No item original o enunciado apresenta um texto motivador do problema, explicando o motivo do empresário sondar a compra de uma das empresas e o critério que ele vai usar.

Para o item reduzido retiramos do enunciado esse contexto do empresário, deixando apenas a parte da empresa, desse modo o corpo ficou menor. Foram modificados os nomes das empresas, para valores mais intuitivos: de A a E, por esse motivo as alternativas foram modificadas também. Para esse item utilizamos as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que eliminamos muitas palavras e também simplificamos a linguagem; Extirpação, já que parte do texto associado ao contexto foi removido; e Estética, ao modificar de maneira mais lógica o nome das empresas, no corpo e nas alternativas.

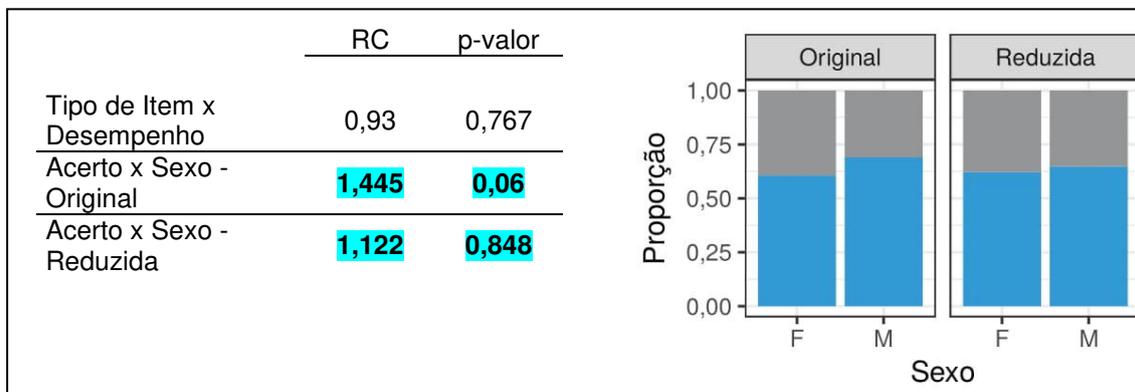
Quadro 37 - Item 16 da Prova e sua respectiva versão Reduzida

<u>Item 16 - Original</u> (Enem 2013 - 154)	<u>Item 16 - Reduzido</u>										
<p>As projeções para a produção de arroz no período de 2012 – 2021, em uma determinada região produtora, apontam para uma perspectiva de crescimento constante da produção anual. O quadro apresenta a quantidade de arroz, em toneladas, que será produzida nos primeiros anos desse período, de acordo com essa projeção.</p> <table border="1"> <thead> <tr> <th>Ano</th> <th>Projeção da produção (t)</th> </tr> </thead> <tbody> <tr> <td>2012</td> <td>50,25</td> </tr> <tr> <td>2013</td> <td>51,50</td> </tr> <tr> <td>2014</td> <td>52,75</td> </tr> <tr> <td>2015</td> <td>54,00</td> </tr> </tbody> </table> <p>A quantidade total de arroz, em toneladas, que deverá ser produzida no período de 2012 a 2021 será de</p> <p>a) 497,25 b) 500,85 c) 502,87 d) 558,75 e) 563,25</p>	Ano	Projeção da produção (t)	2012	50,25	2013	51,50	2014	52,75	2015	54,00	<p>Considere uma sequência de crescimento constante em que o primeiro termo é 50,25, o segundo termo é 51,50, o terceiro termo é 52,75 e assim sucessivamente.</p> <p>Ao somarmos do primeiro ao décimo termo dessa sequência, obtemos:</p> <p>a) 497,25 b) 500,85 c) 502,87 d) 558,75 e) 563,25</p>
Ano	Projeção da produção (t)										
2012	50,25										
2013	51,50										
2014	52,75										
2015	54,00										

O item 16 está apresentado, em suas duas versões, no quadro 37 acima. A redução de carga cognitiva estranha nesse item mostrou uma tendência de igualar o desempenho entre homens e mulheres no item reduzido, mas não aumentou o escore global dos respondentes. Como mostra no quadro 37 a seguir, a RC de acerto de homens e mulheres na versão original é próxima de 1,5 com uma tendência de desempenho melhor masculino.

Na versão reduzida do item, a RC ficou perto de 1,1, sem diferença estatística entre homens e mulheres. Em geral os estudantes acertaram o item reduzido e o item original na mesma proporção, com uma RC próxima de 0,9, sem diferença estatística.

Quadro 38 - Valores de Razão de Chance e p-valor para o Item 16 por tipo de item e sexo, e taxa de acerto por sexo



Na versão original do problema, a taxa de acerto das mulheres (0,61) foi ligeiramente menor que a dos homens (0,69) e na versão reduzida as mulheres (0,62) tiveram um desempenho estatisticamente parecido com o dos homens (0,65). O fato da RC global ficar próximo de 0,9 se deve aos homens que tiveram um desempenho pior na versão reduzida.

Não verificamos viés de sexo nesse item, uma vez que o problema aborda um contexto de produção de arroz. O construto do item foi avaliado em encontrar a soma dos termos de uma sequência numérica, no caso uma progressão aritmética, em uma situação problema. Esse item 16 difere do Item 5, que também aborda sequência. Enquanto nesse item é pedida a soma dos termos de uma sequência, no outro item é pedido um termo específico da sequência.

Decidimos remover o contexto desse item, deixá-lo apenas matemático, definindo a sequência e os termos, desse modo reduzimos a carga cognitiva estranha ao eliminar as explicações preliminares sobre a safra de arroz e particularmente a contagem do período de interesse, de 2012 em diante. Utilizamos as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que a contagem de palavras foi drasticamente reduzida, e além disso no contexto a pergunta foi feita relativa ao ano, e traduzimos apenas para um determinado termo; e Extirpação, já que reduzimos o problema na essência do que ele quer saber do estudante, juntamente com a eliminação da tabela dos termos, que não era necessária.

Quadro 39 - Item 17 da Prova e sua respectiva versão Reduzida

<u>Item 17 - Original</u> (Enem 2012 - 169)	<u>Item 17 - Reduzido</u>
Há, em virtude da demanda crescente de economia de água, equipamentos e utensílios como, por exemplo, as bacias sanitárias ecológicas, que utilizam 6 litros de água por descarga em vez dos 15 litros utilizados por bacias sanitárias não ecológicas, conforme dados da Associação Brasileira de Normas Técnicas (ABNT). Qual será a economia diária de água obtida por meio da substituição de uma bacia sanitária não ecológica, que gasta cerca de 60 litros por dia com a descarga, por uma bacia sanitária ecológica?	O vaso sanitário ecológico utiliza 6 litros de água por descarga, já o vaso sanitário convencional utiliza 15 litros de água por descarga. Considere que uma família com o vaso sanitário convencional utiliza 60 litros por dia com a descarga. Se essa família tivesse um vaso sanitário ecológico ela economizaria quanta água por dia?
a) 24 litros b) 36 litros c) 40 litros d) 42 litros e) 50 litros	a) 24 litros b) 36 litros c) 40 litros d) 42 litros e) 50 litros

O item 17 está apresentado em suas duas versões no quadro 39 acima. A redução de carga cognitiva estranha nesse item não produziu o efeito de igualar o desempenho entre mulheres e homens, mas aumentou o escore global dos respondentes, em comparação com o item original. Como mostra no quadro 39 a seguir, a RC de acerto de homens e mulheres na versão original é próxima de 1,7, onde os homens se sobressaem de maneira significativa. Na versão reduzida do item, a RC ficou perto de 2,5, também com vantagem masculina significativa. A RC entre a questão original e reduzida é de aproximadamente 1,5, o efeito da redução de carga cognitiva melhorou o desempenho global do estudantes.

Quadro 40 - Valores de Razão de Chance e p-valor para o Item 17 por tipo de item e sexo, e taxa de acerto por sexo

	RC	p-valor
Tipo de Item x Desempenho	1,475	0,016
Acerto x Sexo - Original	1,692	0,022
Acerto x Sexo - Reduzida	2,459	<0,001

Sexo	Original (Proporção)	Reduzida (Proporção)
F	~0,48	~0,52
M	~0,60	~0,72

Na versão original do problema, a taxa de acerto das mulheres (0,47) foi consideravelmente menor do que a dos homens (0,6) e na versão reduzida, as mulheres (0,52) continuaram com desempenho estatisticamente inferior aos homens (0,73). Nas duas versões as mulheres tiveram um desempenho inferior ao dos homens, mas a redução de carga cognitiva proporcionou um aumento global significativo no desempenho dos estudantes, tanto homens quanto mulheres tiveram aumento no escore.

Não verificamos viés de sexo nesse item, uma vez que o problema aborda um contexto neutro sobre vaso sanitário ecológico. Sobre o construto do item, entendemos que é obter, à partir de um contexto, o valor de uma operação aritmética. Para o item reduzido ainda deixamos um contexto para os respondentes, mas removemos a parte do contexto do corpo do item original que aciona a motivação para se utilizar um vaso sanitário ecológico, mencionando inclusive a ABNT, que aumenta a carga cognitiva estranha.

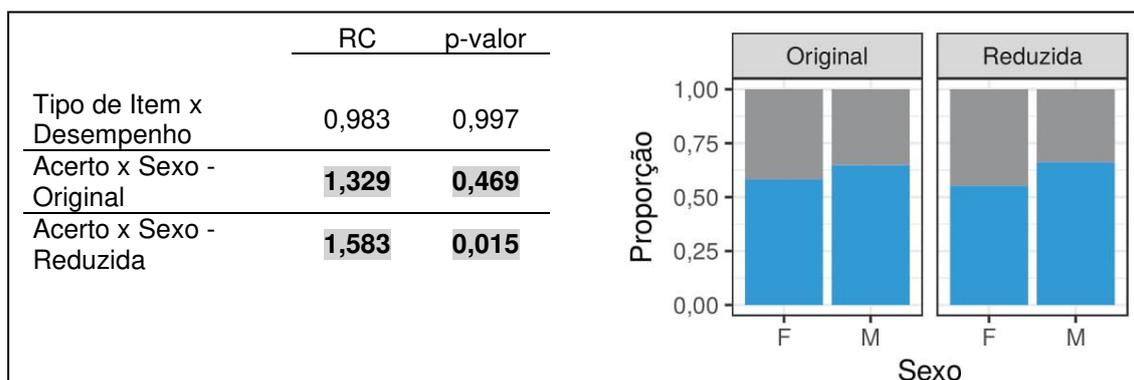
Outra mudança substancial para o item reduzido está no comando, uma vez que o item original ainda possui dados no comando e decidimos na versão reduzida dispor esse dado no próprio corpo. As estratégias de redução de carga cognitiva utilizadas foram: Tradução, uma vez que reduzimos um pouco a quantidade de palavras; Extirpação, já que retiramos parte do texto, reduzindo o problema à essência do que ele quer saber do estudante; e Estética, pois entendemos que o comando não estava muito claro na versão original, e as informações foram colocadas de maneira mais natural para a resolução dos estudantes.

Quadro 41 - Item 19 da Prova e sua respectiva versão Reduzida

<u>Item 19 - Original</u> (Enem 2013 - 147)	<u>Item 19 - Reduzido</u>
Para se construir um contrapiso, é comum, na constituição do concreto, se utilizar cimento, areia e brita, na seguinte proporção: 1 parte de cimento, 4 partes de areia e 2 partes de brita. Para construir o contrapiso de uma garagem, uma construtora encomendou um caminhão betoneira com 14 m ³ de concreto. Qual é o volume de cimento, em m ³ , na carga de concreto trazido pela betoneira?	Para produzir concreto é comum utilizar cimento, areia e brita na seguinte proporção: 1 parte de cimento, 4 partes de areia e 2 partes de brita. Em 14 m ³ de concreto, o volume de cimento, em m ³ , é de:
a) 1,75	a) 1,75
b) 2,00	b) 2,00
c) 2,33	c) 2,33
d) 4,00	d) 4,00
e) 8,00	e) 8,00

O item 19, em suas duas versões, está localizado no quadro 41 acima. A redução de carga cognitiva estranha aumentou a diferença de desempenho entre homens e mulheres, comparando a versão original com a versão reduzida. O escore global ficou praticamente inalterado com a redução de carga cognitiva.

Quadro 42 - Valores de Razão de Chance e p-valor para o Item 19 por tipo de item e sexo, e taxa de acerto por sexo



Como mostrado no quadro 42, a RC para de acerto entre homens e mulheres na versão original do item é próxima de 1,3, sem diferença significativa. Na versão reduzida do item, a RC ficou próxima de 1,6, de maneira significativa. A redução de carga cognitiva estranha provocou o efeito contrário do que gostaríamos: no item original os sexos eram indistinguíveis no escore, com a redução de carga cognitiva estranha, os homens se sobressaíram frente às mulheres. Por fim, o desempenho geral dos estudantes no item reduzido, em comparação ao item original, foi praticamente igual, com uma RC próxima de 1,0.

Na versão original do problema, a taxa de acerto das mulheres (0,58) foi equivalente ao dos homens (0,65) e na versão reduzida as mulheres (0,55) tiveram um desempenho estatisticamente inferior aos homens (0,66). O desempenho das mulheres, no item original e no reduzido, foi muito parecido, assim com o dos homens, mas na comparação entre os sexos em cada um dos tipos de item, esse leve aumento no desempenho dos homens, entre o item original e o reduzido, e a leve queda de desempenho das mulheres, acarretou uma diferença significativa no item reduzido. Pensando globalmente, os estudantes tiveram um desempenho ligeiramente menor no item reduzido, efeito causado pelo escore das mulheres.

Sobre o construto do item, entendemos que é obter, à partir de um contexto, o valor de uma razão, ou proporção. Acreditamos que o item apresenta um leve viés em

relação ao sexo uma vez que ele aborda um contexto de construção. Para o item reduzido foi retirada a parte em que é mencionado a construção de um contrapiso por uma construtora, e também removemos a menção à betoneira, pois todos esses elementos elevam a carga cognitiva estranha do problema, uma vez que não possuem relevância para o construto do item. Outra mudança no item reduzido está no comando, foi colocado um dado que o estudante utiliza imediatamente na resolução do item, e logo na sequência, a pergunta.

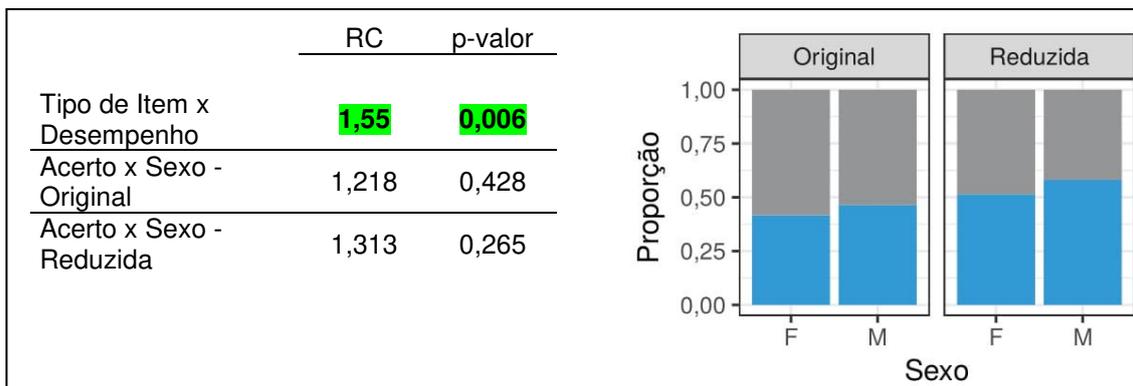
As estratégias de redução de carga cognitiva estranha utilizadas foram: Tradução, uma vez que reduzimos um pouco a quantidade de palavras; Extirpação, já que reduzimos o problema na essência do que ele quer saber do estudante, retirando menções externas que não estavam associadas ao construto; e Estética, pois entendemos que o comando do item original pudesse ser mais imediato, e no item reduzido as informações foram colocadas de maneira mais natural para a resolução.

Quadro 43 - Item 20 da Prova e sua respectiva versão Reduzida

Item 20 - Original (Enem 2014 - 137)	Item 20 - Reduzido
<p>Uma empresa que organiza eventos de formatura confecciona canudos de diploma a partir de folhas de papel quadradas. Para que todos os canudos fiquem idênticos, cada folha é enrolada em torno de um cilindro de madeira de diâmetro d em centímetros, sem folga, dando-se cinco voltas completas em torno de tal cilindro. Ao final, amarra-se um cordão no meio do diploma, bem ajustado, para que não ocorra o desenrolamento, como ilustrado na figura.</p>	<p>Uma folha de papel <u>quadrada</u> é enrolada em torno de um cilindro de madeira de diâmetro d, em centímetros, dando-se 5 voltas completas em tal cilindro, em seguida, retira-se o cilindro de madeira do meio do papel enrolado. Considere que a espessura do papel original seja desprezível. Qual é a medida, em centímetros, do lado dessa folha de papel.</p>
	
<p>Diploma (Foto: Reprodução/ENEM)</p>	
<p>Em seguida, retira-se o cilindro de madeira do meio do papel enrolado, finalizando a confecção do diploma. Considere que a espessura da folha de papel original seja desprezível.</p>	
<p>Qual é a medida, em centímetros, do lado da folha de papel usado na confecção do diploma?</p>	
<p>a) $\pi \cdot d$ b) $2\pi \cdot d$ c) $4\pi \cdot d$ d) $5\pi \cdot d$ e) $10\pi \cdot d$</p>	<p>a) $\pi \cdot d$ b) $2\pi \cdot d$ c) $4\pi \cdot d$ d) $5\pi \cdot d$ e) $10\pi \cdot d$</p>

O item 20, apresentado em suas duas versões no quadro 43 acima, não apresentou discrepância entre taxa de acerto de homens e mulheres, entre a versão original e reduzida. Em contrapartida, a redução de carga cognitiva melhorou o escore global dos estudantes.

Quadro 44 - Valores de Razão de Chance e p-valor para o Item 20 por tipo de item e sexo, e taxa de acerto por sexo



Como mostra o quadro 44 acima, a RC para de acerto entre homens e mulheres na versão original é próxima de 1,2, e na versão reduzida do item, a RC ficou próxima de 1,3, nos dois casos, sem diferença estatística significativa. Na versão reduzida do item, a RC ficou próxima de 1,6, de maneira significativa.

Na versão original do problema, a taxa de acerto das mulheres (0,42) foi ligeiramente menor do que a dos homens (0,46) e na versão reduzida as mulheres (0,51) continuaram com desempenho um pouco inferior aos homens (0,59). Nas duas versões as mulheres tiveram um desempenho inferior ao dos homens, mas não de maneira significativa, a redução de carga cognitiva proporcionou um aumento global significativo no desempenho dos estudantes.

Não verificamos viés de sexo nesse item, uma vez que o problema aborda um contexto neutro sobre canudo de formatura. Sobre o construto do item, entendemos que é obter, à partir de um contexto, o valor de uma medida em geometria. Para o item reduzido removemos a figura do canudo de formatura e a parte do contexto em que o corpo do item original menciona a empresa de formatura. Outra mudança substancial do item reduzido está na síntese do que o problema pede, uma vez que o problema original ainda possui muitas palavras para explicar o que devia, produzindo assim uma carga cognitiva estranha desnecessária ao aluno. Por último, e talvez o grande fator que pode ter atrapalhado os estudantes nesse item, é a menção à folha quadrada, logo no começo do item original. Um estudante que não levou essa informação à diante teve que parar o problema no meio para voltar ao enunciado, uma vez que não é possível resolver o problema sem considerar o formato da folha.

As estratégias de redução de carga cognitiva estranha utilizadas foram: Tradução, uma vez que reduzimos um pouco a quantidade de palavras; Extirpação, já que removemos parte do corpo, reduzindo o problema na essência do que ele quer saber do estudante; e Sinalização, ao sublinhar o formato da folha, pois entendemos que o fato do papel ser quadrado é muito relevante, e que ao perder essa informação no enunciado o aluno pode desperdiçar muito tempo na resolução do problema.

A redução de carga cognitiva estranha do item 21, apresentado em suas duas versões no quadro 45 a seguir, provocou o efeito de igualar o desempenho das mulheres ao dos homens. De um modo global, o desempenho dos estudantes não mudou significativamente entre essas duas versões.

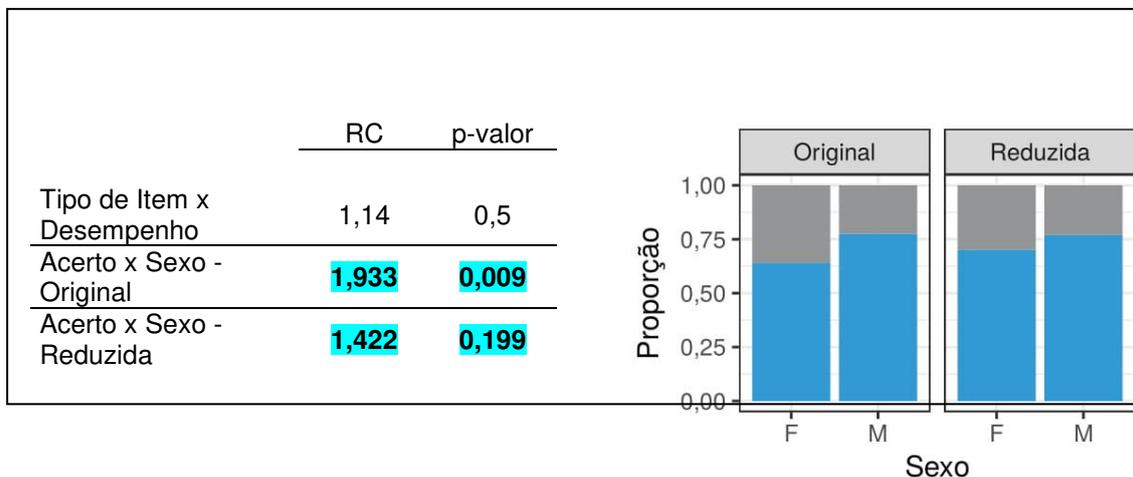
Quadro 45 - Item 21 da Prova e sua respectiva versão Reduzida

<u>Item 21 - Original</u> (Enem 2014 - 143)				<u>Item 21 - Reduzido</u>			
<p>O Ministério da Saúde e as unidades federadas promovem frequentemente campanhas nacionais e locais de incentivo à doação voluntária de sangue, em regiões com menos número de doadores por habitante, com o intuito de manter a regularidade de estoques nos serviços hemoterápicos. Em 2010, foram recolhidos dados sobre o número de doadores e o número de habitantes de cada região conforme o quadro seguinte. Os resultados obtidos permitiram que estados, municípios e o governo federal estabelecessem as regiões prioritárias do país para a intensificação das campanhas de doação de sangue.</p>				<p>Em 2010, foram recolhidos dados sobre o número de doadores e o número de habitantes de cada região conforme o quadro seguinte.</p>			
Taxa de doação de sangue, por região, em 2010				Taxa de doação de sangue, por região, em 2010			
Região	Doadores	Número de habitantes	Doadores/habitantes	Região	Doadores (x1000)	Habitantes (x1000)	Doadores/habitantes
Nordeste	820 959	53 081 950	1,5%	Nordeste	820	53000	1,5%
Norte	232 079	15 864 454	1,5%	Norte	232	16000	1,5%
Sudeste	1 521 766	80 364 410	1,9%	Sudeste	1521	80000	1,9%
Centro-Oeste	362 334	14 058 094	2,6%	Centro-Oeste	362	14000	2,6%
Sul	690 391	27 386 891	2,5%	Sul	690	27000	2,5%
Total	3 627 529	190 755 799	1,9%	Total	3628	190000	1,9%
<p>Disponível em: http://bvsmms.saude.gov.br. Acesso em: 2 ago, 2013 (adaptado).</p> <p>A campanha deveria ser intensificada nas regiões em que o percentual de doadores por habitantes fosse menor ou igual ao do país.</p> <p>a) Norte, Centro-Oeste e Sul. b) Norte, Nordeste e Sudeste. c) Nordeste, Norte e Sul. d) Nordeste, Sudeste e Sul. e) Centro-Oeste, Sul e Sudeste.</p>				<p>Disponível em: http://bvsmms.saude.gov.br. Acesso em: 2 ago, 2013 (adaptado).</p> <p>As regiões brasileiras que possuem uma quantidade de doadores menores ou iguais à do Brasil são:</p> <p>a) Norte, Centro-Oeste e Sul. b) Norte, Nordeste e Sudeste. c) Nordeste, Norte e Sul. d) Nordeste, Sudeste e Sul. e) Centro-Oeste, Sul e Sudeste.</p>			

Como mostra o quadro 46 a seguir, o item original apresenta RC perto de 1,9 para o acerto dos homens em relação ao das mulheres, de maneira significativa. Já o item

reduzido possui RC próximo de 1,4, sem diferença estatística significativa. A RC entre a questão original e reduzida é de aproximadamente 1,1.

Quadro 46 - Valores de Razão de Chance e p-valor para o Item 21 por tipo de item e sexo, e taxa de acerto por sexo



O efeito da redução de carga cognitiva estranha não fez um efeito global nos estudantes, mas melhorou o desempenho feminino. A taxa de acerto das mulheres (0,64) era consideravelmente menor que a dos homens (0,77) na versão original, e na versão reduzida as mulheres (0,7) tiveram um desempenho estatisticamente igual aos homens (0,77), portanto a redução da carga cognitiva estranha não alterou o desempenho dos homens, entre o item original e sua versão reduzida.

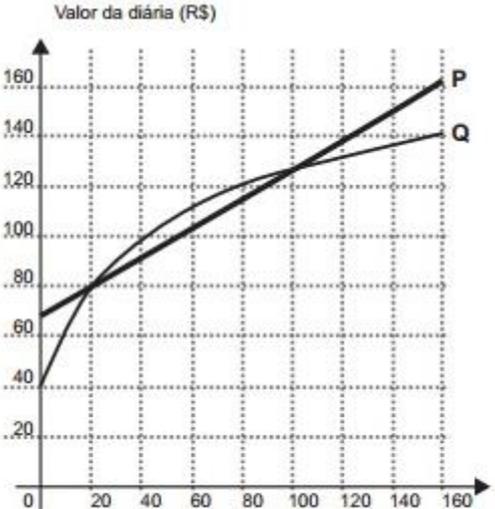
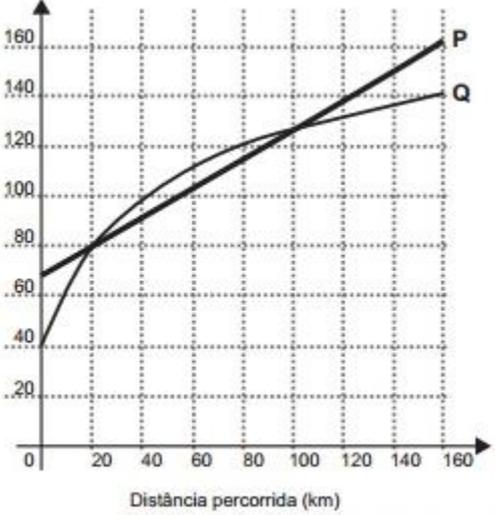
Acreditamos que o item original não possui viés em relação ao sexo, uma vez que o contexto versa sobre número de doadores de sangue no Brasil. O construto do item afere sobre ordem numérica (maior ou menor) em um contexto. No item original existe uma explicação muito prolongada sobre campanha de doação de sangue além de muitos termos, como serviços hemoterápicos, que não são relevantes para o construto e aumentam a carga cognitiva estranha. Para o item reduzido removemos substancialmente o contexto, mas ainda deixamos a mesma situação e uma tabela análoga sobre a taxa de doadores de sangue no Brasil. Nessa tabela substituímos valores que não eram relevantes à resolução do problema por números mais amigáveis, os alunos tiveram apenas que comparar o percentual do Brasil com os estados, como no item original.

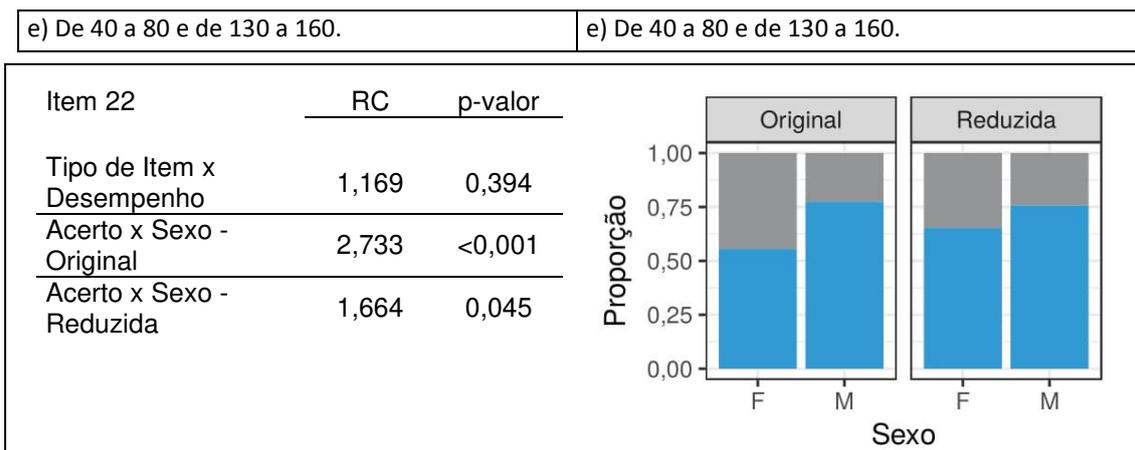
Para esse item utilizamos as estratégias de redução de carga cognitiva estranha de Tradução, uma vez que eliminamos muitas palavras e também simplificamos a

linguagem; Extirpação, uma vez que parte do texto associado ao contexto foi removido e a tabela ficou reduzida ao construto pretendido; e Simplificação Numérica, uma vez que acreditamos que os números não tinham influência no construto pedido.

Apresentamos o item 22, em suas duas versões, no quadro 47 abaixo. A redução de carga cognitiva estranha nesse item teve uma tendência de efeito de melhoria do desempenho das mulheres em relação aos homens, mas essa redução de carga cognitiva não teve impacto significativo na taxa global de acerto. Como mostra o quadro 47 a seguir, o item original apresenta RC perto de 2,7 para o acerto dos homens em relação ao das mulheres, de maneira significativa. Já o item reduzido possui RC próximo de 1,7, com um p-valor de 0,045, ainda estatisticamente significativo. A RC entre a questão original e reduzida é de aproximadamente 1,2, em outras palavras, o efeito da redução de carga cognitiva não alterou o desempenho global do estudantes.

Quadro 47 - Item 22 da Prova e sua respectiva versão Reduzida

Item 22 - Original (Enem 2015 - 141)	Item 22 - Reduzido
<p>Atualmente existem diversas locadoras de veículos, permitindo uma concorrência saudável para o mercado, fazendo com que os preços se tornem acessíveis. Nas locadoras P e Q, o valor da diária de seus carros depende da distância percorrida, conforme o gráfico.</p>	<p>Nas locadoras de veículo P e Q, o valor da diária de seus carros depende da distância percorrida, conforme o gráfico.</p>
 <p>Disponível em: www.sempretops.com. Acesso em: 7 ago. 2015.</p>	 <p>Disponível em: www.sempretops.com. Acesso em: 7 ago. 2015.</p>
<p>O valor pago na locadora Q é menor ou igual àquele pago na locadora P para distâncias, em quilômetros, presentes em qual(is) intervalo(s)?</p> <p>a) De 20 a 100. b) De 80 a 130. c) De 100 a 160. d) De 0 a 20 e de 100 a 160.</p>	<p>Para que intervalos, em quilômetros, o valor pago na locadora Q é <u>menor ou igual</u> àquele pago na locadora P?</p> <p>a) De 20 a 100. b) De 80 a 130. c) De 100 a 160. d) De 0 a 20 e de 100 a 160.</p>



Quadro 48 - Valores de Razão de Chance e p-valor para o Item 22 por tipo de item e sexo, e taxa de acerto por sexo

Para a versão do original problema, a taxa de acerto das mulheres (0,55) foi inferior dos homens (0,77), e para a versão reduzida, o desempenho das mulheres (0,65) melhorou mas ainda foi inferior aos homens (0,76), que por sua vez oscilaram no desempenho.

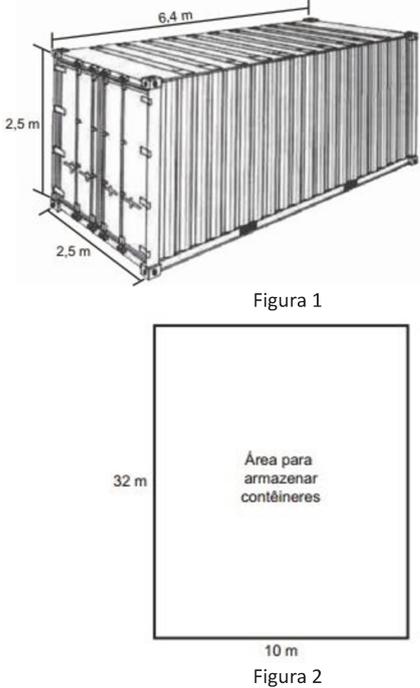
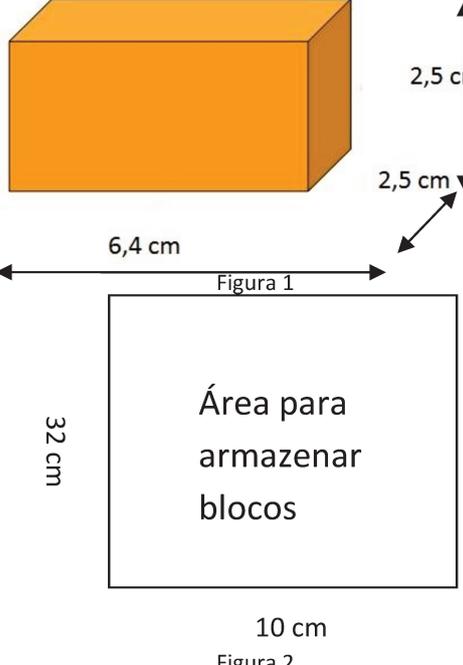
O item nos parece neutro em relação à possibilidade de viés entre os sexos, uma vez que ele aborda aluguel de veículos. O construto do item afere sobre análise de dados em um gráfico aplicado a um contexto. O item original oferece informações bem diretas para a resolução, com pouco espaço para redução de carga cognitiva estranha proporcionada por informações desnecessárias. O comando nos pareceu confuso, ao fazer referência ao valor pago em cada locadora e depois perguntar sobre a distância.

Para o item reduzido modificamos o comando, deixando explícito de maneira sublinhada, a ordem entre os valores das locadoras P e Q. Algumas palavras foram retiradas, mas o contexto permaneceu o mesmo. Utilizamos nesse item as estratégias de redução de carga cognitiva estranha de: Tradução, uma vez que eliminamos algumas palavras; Extirpação, uma vez que parte do texto associado ao contexto foi removido; Estética, já que o comando foi modificado de maneira mais lógica; e Sinalização, ao deixar evidente qual é a ordem procurada no comando.

O item 23, apresentado em suas duas versões no quadro 49 a seguir. A redução de carga cognitiva estranha nesse item teve uma tendência de efeito de melhoria do desempenho das mulheres em relação aos homens, assim como no item 22. Diferentemente do item 22, essa redução de carga cognitiva estranha teve um impacto negativo significativo na taxa global de acerto. Como mostra o quadro 50 a seguir, o item

original apresenta RC perto de 2,0 para o acerto dos homens em relação ao das mulheres, de maneira significativa. Já o item reduzido possui RC próximo de 2,0, mas com um p-valor de 0,046, ainda estatisticamente significativo. A RC entre a questão original e reduzida é de aproximadamente 0,7, com um p-valor de 0,027. Em outras palavras, o efeito da redução de carga cognitiva piorou o desempenho global do estudantes.

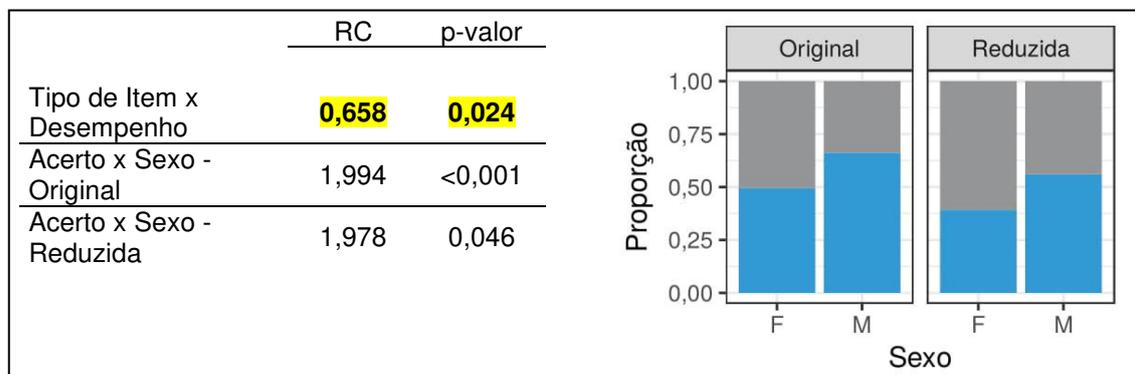
Quadro 49 - Item 23 da Prova e sua respectiva versão Reduzida

<p>Item 23 - Original (Enem 2015 - 143)</p> <p>Uma carga de 100 contêineres, idênticos ao modelo apresentado na Figura 1, deverá ser descarregada no porto de uma cidade. Para isso, uma área retangular de 10 m por 32 m foi cedida para o empilhamento desses contêineres (Figura 2).</p>  <p>De acordo com as normas desse porto, os contêineres deverão ser empilhados de forma a não sobrem espaços nem ultrapassem a área delimitada. Após o empilhamento total da carga e atendendo à norma do porto, a altura mínima a ser atingida por essa pilha de contêineres é</p> <ol style="list-style-type: none"> 12,5 m. 17,5 m. 25,0 m. 22,5 m. 32,5 m. 	<p>Item 23 - Reduzido</p> <p>Uma criança brinca com 100 peças idênticas de na forma de paralelepípedo, como mostra a Figura 1. Ela deverá armazenar os paralelepípedos uma área retangular de 10 cm por 32 cm (Figura 2).</p>  <p>A criança empilha essas peças de forma que não sobrem espaços e nem ultrapasse a área delimitada. Após o empilhamento total, a altura <u>mínima</u> atingida por essas peças é:</p> <ol style="list-style-type: none"> 12,5 cm. 17,5 cm. 25,0 cm. 22,5 cm. 32,5 cm.
--	--

Na versão original do problema, a taxa de acerto das mulheres (0,5) foi consideravelmente menor do que a dos homens (0,66) e na versão reduzida as mulheres (0,39) tiveram também um desempenho estatisticamente inferior aos homens (0,57). A

redução de carga cognitiva acarretou uma diminuição global significativa no desempenho dos estudantes. Sobre o construto do item, entendemos que é utilizar geometria plana e espacial em um contexto.

Quadro 50 - Valores de Razão de Chance e p-valor para o Item 23 por tipo de item e sexo, e taxa de acerto por sexo



É pouco provável que exista viés de sexo no item original, uma vez que o problema aborda um contexto sobre a disposição de contêineres. Para o item reduzido mudamos o contexto do problema, de contêineres para blocos de brinquedo, na esperança de que a carga cognitiva estranha fosse atenuada para um contexto mais próximo aos alunos. Além disso retiramos algumas palavras do texto e enfatizamos uma parte do comando, sublinhando o que era pedido.

As estratégias de redução de carga cognitiva utilizadas foram: Tradução, uma vez que reduzimos substancialmente a quantidade de palavras e deixamos o contexto mais simplificado; Extirpação, já que reduzimos o problema na essência do que ele quer saber do estudante; e Sinalização, uma vez que o problema pedia a altura mínima e ao sublinhar essa parte foi frisado o que se pedia no problema.

O item 24, está apresentado em suas duas versões no quadro 51 a seguir. A redução de carga cognitiva estranha nesse item teve uma tendência de efeito de piorar do desempenho das mulheres em relação aos homens. Comparando as duas versões do item, essa redução de carga cognitiva estranha não teve um impacto significativo na taxa global de acerto.

Quadro 51 - Item 24 da Prova e sua respectiva versão Reduzida

Item 24 - Original (Enem 2013 - 140)

Em um certo teatro, as poltronas são divididas em setores. A figura apresenta a vista do setor 3 desse teatro, no qual as cadeiras escuras estão reservadas e as claras não foram vendidas.

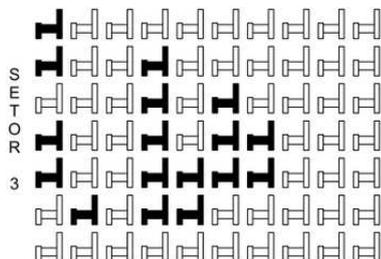


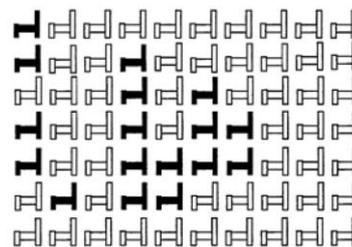
Figura (Foto: Reprodução)

A razão que representa a quantidade de cadeiras reservadas do setor 3 em relação ao total de cadeiras desse mesmo setor é

- a) $17/70$
- b) $17/53$
- c) $53/70$
- d) $53/17$
- e) $70/17$

Item 24 - Reduzido

A figura apresenta um setor de um certo teatro no qual as cadeiras escuras estão reservadas e as claras não foram vendidas.



A razão que representa a quantidade de cadeiras reservadas nesse setor em relação ao total de cadeiras nesse mesmo setor é:

- a) $17/70$
- b) $17/53$
- c) $53/70$
- d) $53/17$
- e) $70/17$

Quadro 52 - Valores de Razão de Chance e p-valor para o Item 24 por tipo de item e sexo, e taxa de acerto por sexo

	RC	p-valor
Tipo de Item x Desempenho	1,244	0,33
Acerto x Sexo - Original	1,766	0,055
Acerto x Sexo - Reduzida	2,107	0,02

Como mostra o quadro 52 acima, o item original apresenta RC perto de 1,8 para o acerto dos homens em relação ao das mulheres com um p-valor de 0,055, ou seja, existe uma maior tendência de acerto masculina. Já o item reduzido possui RC próximo de 2,1, de maneira significativa. A RC entre a questão original e reduzida é de aproximadamente 1,2, o efeito da redução de carga cognitiva não alterou o desempenho global do estudantes.

Os índices de acerto nos dois tipos de item foram altos, na versão original do problema, a taxa de acerto das mulheres (0,76) era consideravelmente menor que a dos

homens (0,85) e na versão reduzida as mulheres (0,78) não tiveram um aumento tão grande enquanto que os homens tiveram um desempenho significativamente melhor (0,89) que as mulheres. A redução de carga cognitiva em si não prejudicou as mulheres na versão reduzida, na verdade aumentou um pouco o escore feminino, mas beneficiou muito mais os homens.

Não verificamos viés em relação ao sexo nesse item, já que ele trata de escolha de poltronas em um teatro. O construto pretendido pelo item, é encontrar uma razão numérica a partir de um contexto. O item original não tinha muita mobilidade para mudanças substanciais na carga cognitiva estranha, retiramos a referência ao setor do teatro já que essa informação acarretava carga cognitiva estranha e não era útil para o construto do problema. Utilizamos para o item reduzido as estratégias de: Tradução, uma vez que a contagem de palavras foi reduzida; e Extirpação, ao remover o setor do teatro, tanto no texto quanto na figura.

ANEXO I - Histórico do SAEB

Histórico

1990

Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
1ª, 3ª, 5ª e 7ª séries do EF	Escolas públicas Amostral	Curriculos de sistemas estaduais	Língua Portuguesa, Matemática, Ciências Naturais, Redação

1993

Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
1ª, 3ª, 5ª e 7ª séries do EF	Escolas públicas Amostral	Curriculos de sistemas estaduais	Língua Portuguesa, Matemática, Ciências Naturais, Redação

1995

Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
1ª, 3ª, 5ª e 7ª séries do EF	Escolas públicas Amostral	Curriculos de sistemas estaduais	Língua Portuguesa, Matemática, Ciências Naturais, Redação

Inovações

Em 1995 foi adotada uma nova metodologia de construção do teste e análise de resultados: a Teoria de Resposta ao Item (TRI). Dessa forma, a comparabilidade entre os resultados das avaliações ao longo do tempo se tornou possível. Ocorreu também o levantamento de dados contextuais (Questionários).

1997

Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
4ª, 8ª séries do EF 3ª série do EM	Escolas públicas + Escolas Particulares Amostral	Matrizes de Referência - Avalia competências / Define descritores (conteúdo curriculares + operações mentais)	Língua Portuguesa, Matemática, Ciências Naturais, Física, Química e Biologia

1999			
Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
4ª, 8ª séries do EF 3ª série do EM	Escolas públicas + Escolas Particulares Amostral	Matrizes de Referência - Avalia competências / Define descritores (conteúdo curriculares + operações mentais)	Língua Portuguesa, Matemática, Ciências Naturais, Física, Química e Biologia História e Geografia
2001			
Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
4ª, 8ª séries do EF 3ª série do EM	Escolas públicas + Escolas Particulares Amostral	Matrizes de Referência - Avalia competências / Define descritores (conteúdo curriculares + operações mentais)	Língua Portuguesa, Matemática
2003			
Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
4ª, 8ª séries do EF 3ª série do EM	Escolas públicas + Escolas Particulares Amostral	Matrizes de Referência - Avalia competências / Define descritores (conteúdo curriculares + operações mentais)	Língua Portuguesa, Matemática
2005			
Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
4ª, 8ª séries do EF 3ª série do EM	Escolas públicas + Escolas Particulares Amostral + Estratos Censitários Ideb	Matrizes de Referência - Avalia competências / Define descritores (conteúdo curriculares + operações mentais)	Língua Portuguesa, Matemática
Inovações			
<p>Em 2005 o Saeb foi reestruturado pela Portaria Ministerial nº 931, de 21 de março de 2005. O sistema passou a ser composto por duas avaliações: Avaliação Nacional da Educação Básica (Aneb) e Avaliação Nacional do Rendimento Escolar (Anresc), conhecida como Prova Brasil. A Aneb manteve os procedimentos da avaliação amostral (atendendo aos critérios estatísticos de no mínimo 10 estudantes por turma) das redes públicas e privadas, com foco na gestão da educação básica que até então vinha sendo realizada no Saeb. A Anresc (Prova Brasil) passou a avaliar de forma censitária as escolas que atendessem aos critérios de no mínimo 30 estudantes matriculados na última etapa dos anos iniciais (4ª série/5º ano) ou dos anos finais (8ª série/9º ano) do Ensino Fundamental escolas públicas, permitindo gerar resultados por escola.</p>			

2007

Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
4ª, 8ª séries do EF 3ª série do EM	Escolas públicas + Escolas Particulares Amostral + Estratos Censitários Ideb	Matrizes de Referência - Avalia competências / Define descritores (conteúdo curriculares + operações mentais)	Língua Portuguesa, Matemática

2009

Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
4ª, 8ª séries do EF 3ª série do EM	Escolas públicas + Escolas Particulares Amostral + Estratos Censitários Ideb	Matrizes de Referência - Avalia competências / Define descritores (conteúdo curriculares + operações mentais)	Língua Portuguesa, Matemática

2011

Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
4ª, 8ª séries do EF 3ª série do EM	Escolas públicas + Escolas Particulares Amostral + Estratos Censitários Ideb	Matrizes de Referência - Avalia competências / Define descritores (conteúdo curriculares + operações mentais)	Língua Portuguesa, Matemática

2013

Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
5º ano, 9º ano do EF 3ª série do EM	Escolas públicas + Escolas Particulares Amostral + Estratos Censitários Ideb	Matrizes de Referência - Avalia competências / Define descritores (conteúdo curriculares + operações mentais)	Língua Portuguesa, Matemática

Inovações

Na edição de 2013 a Avaliação Nacional da Alfabetização (ANA), previsto no Pacto Nacional pela Alfabetização na Idade Certa – PNAIC, passou a compor o Saeb a partir da divulgação da portaria nº 482, de 7 de junho de 2013. Outra inovação desta edição foi a inclusão em caráter experimental da avaliação de Ciências, a ser realizada com os estudantes do 9º ano do Ensino Fundamental e do 3º ano do Ensino Médio. Neste ano foi aplicado, em caráter de estudo experimental, um pré-teste de Ciências Naturais, História e Geografia que não geraram resultados para a edição.

2015			
Público-alvo	Abrangência	Formulação dos Itens	Áreas do Conhecimento / Disciplinas Avaliadas
5º ano, 9º ano do EF 3ª série do EM	Escolas públicas + Escolas Particulares Amostral + Estratos Censitários Ideb	Matrizes de Referência - Avalia competências / Define descritores (conteúdo curriculares + operações mentais)	Língua Portuguesa, Matemática

Dados retirados de <http://portal.inep.gov.br/educacao-basica/saeb> acesso em 08/01/18

ANEXO II - Matrizes: Competência, Habilidade e Análise de Desempenho

As **competências** são as seguintes:

- I. *Demonstrar domínio básico da norma culta da Língua Portuguesa e do uso das diferentes linguagens: matemática, artística, científica, entre outras.*
- II. *Construir e aplicar conceitos das várias áreas do conhecimento para a compreensão de fenômenos naturais, de processos histórico-geográficos, da produção tecnológica e das manifestações artísticas.*
- III. *Selecionar, organizar, relacionar, interpretar dados e informações representados de diferentes formas, para enfrentar situações-problema, segundo uma visão crítica com vista à tomada de decisões.*
- IV. *Organizar informações e conhecimentos disponíveis em situações concretas, para a construção de argumentações consistentes.*
- V. *Recorrer aos conhecimentos desenvolvidos na escola para elaboração de propostas de intervenção solidária na realidade, considerando a diversidade sociocultural como inerente à condição humana no tempo e no espaço.*

As competências são gerais e expressam-se em 21 (vinte e uma) **habilidades**:

- 1) Dada a descrição discursiva ou por ilustração de um experimento real simples, de natureza técnico-científica (física, biológica, sociológica etc.), identificar variáveis relevantes e selecionar os instrumentos necessários para a realização e/ou a interpretação dos resultados do mesmo.
- 2) Em um gráfico cartesiano de variável socioeconômica ou técnico-científica em função do tempo:
 - identificar o valor da variável em dado instante ou em que instante a variável assume um dado valor;
 - identificar trechos em que este valor é crescente, decrescente ou constante;
 - analisar qualitativamente, em cada trecho, a taxa de variação.
- 3) Dado um diagrama de distribuição estatística de variável social, econômica, física, química ou biológica:
 - traduzir as informações disponíveis na linguagem ordinária;
 - identificar a representação de informações gráficas de diferentes maneiras;
 - reorganizar as informações, possibilitando interpolações ou extrapolações tendo em vista finalidades específicas.
- 4) Dada uma situação-problema no âmbito de determinada área de conhecimento, apresentada em linguagem comum, relacioná-la com sua formulação em diferentes linguagens; reciprocamente, dada uma destas formulações, relacioná-la a uma situação-problema descrita por um texto.
- 5) A partir da leitura de textos literários consagrados e de dados específicos sobre movimentos estéticos:

- identificar as principais características dos movimentos literários em que se situam;
- inferir as escolhas dos temas, gêneros e recursos lingüísticos dos autores;
- identificar seu contexto social, político, histórico e cultural;
- estabelecer relações entre textos de movimentos literários diversos.

6) Tendo como base textos orais e/ou escritos:

- identificar a função e a natureza da linguagem;
- distinguir as marcas das variantes lingüísticas de ordem sociocultural, geográfica, de registro, de estilo;
- analisar os elementos constituintes da linguagem oral e escrita;
- transformar as marcas da linguagem oral em linguagem escrita formal.

7) Reconhecer a conservação da energia em processos de transformação próprios da utilização ou da produção de recursos energéticos de uso social, como hidro-eletricidade ou derivados do petróleo.

8) Identificar e dimensionar processos mecânicos, elétricos e térmicos presentes na operação de instalações (residenciais ou sociais), em equipamentos (como veículos e outras máquinas) e em configurações naturais (como fenômenos atmosféricos):

- analisar perturbações ambientais decorrentes;
- analisar as implicações sociais e econômicas dos processos.

9) Demonstrar compreensão do significado e a importância da água e de seu ciclo para a determinação do clima e para a preservação da vida, sabendo quantificar variações de temperatura ou mudanças de fase em circunstâncias específicas.

10) Utilizar diferentes escalas de tempo para situar e descrever transformações planetárias (litosfera e biosfera), origem e evolução da vida, crescimento de diferentes populações.

11) Identificar uma unidade fundamental no fenômeno vital: padrões comuns aos processos metabólicos, nas estruturas intracelulares e nos códigos químicos de informação para a reprodução, que garantem a continuidade da vida, diante da diversidade de manifestações de vida e dos distintos níveis de complexidade, apresentados na forma de texto, diagramas ou outras ilustrações.

12) Reconhecer fatores socioeconômicos e ambientais que interferem nos padrões de saúde e desenvolvimento de populações humanas, por meio da interpretação ou da análise de gráficos e tabelas de indicadores.

13) Relacionar a diversidade de formas de vida à variedade de condições do meio, demonstrando compreensão do caráter dinâmico e sistêmico da vida no planeta por meio da análise de textos, diagramas ou outras formas de organização de dados.

14) Diante da riqueza e da diversidade de formas geométricas planas ou espaciais presentes na natureza ou imaginadas a partir delas, como polígonos, círculos, circunferências, prismas, pirâmides, cilindros, cones, esferas etc.:

- identificá-las e caracterizá-las através de propriedades;
- interpretar sua representação gráfica;
- perceber relações entre seus elementos, tendo em vista a realização de medidas de comprimentos, áreas e volumes em unidades adequadas;
- utilizar o conhecimento geométrico construído para o aperfeiçoamento da leitura, da compreensão e da ação sobre a realidade concreta.

15) Utilizar instrumentos adequados para descrição de fenômenos naturais, demonstrando compreensão dos aspectos aleatórios dos mesmos:

- em medidas e representação de frequências relativas;
- na construção de espaços amostrais, com a atribuição de probabilidades aos eventos elementares;
- no cálculo de probabilidades de eventos relevantes em situações concretas.

16) A partir da análise de diferentes situações-problema referentes à perturbação ambiental na atmosfera, na hidrosfera ou na litosfera:

- identificar fonte, transporte e sorvedouro dos poluentes e contaminantes;
- reconhecer algumas transformações químicas e biológicas que possam ocorrer durante o transporte do poluente;
- prever possíveis efeitos nos ecossistemas e no sistema produtivo que decorram das alterações ambientais apresentadas;
- propor formas de intervenção para reduzir os efeitos agudos e crônicos da poluição ambiental.

17) Apresentados alguns processos que envolvem transformações de materiais, como, por exemplo, a metalurgia do ferro, a produção do álcool:

- reconhecer as etapas intermediárias relevantes;
- identificar e calcular a conservação da massa, o rendimento, a variação de energia e a rapidez do processo;
- analisar o equilíbrio químico e suas perturbações ;
- analisar as perturbações ambientais;
- analisar as implicações sociais e econômicas dos processos.

18) Identificar os elementos que compõem a diversidade artística e cultural, manifestos no tempo e no espaço e que caracterizam a condição humana como fenômeno diverso e complexo.

19) Confrontar interpretações diversas de uma dada realidade histórico-geográfica:

- coordenando os diferentes pontos de vista em jogo;

- identificando os pressupostos de cada interpretação.

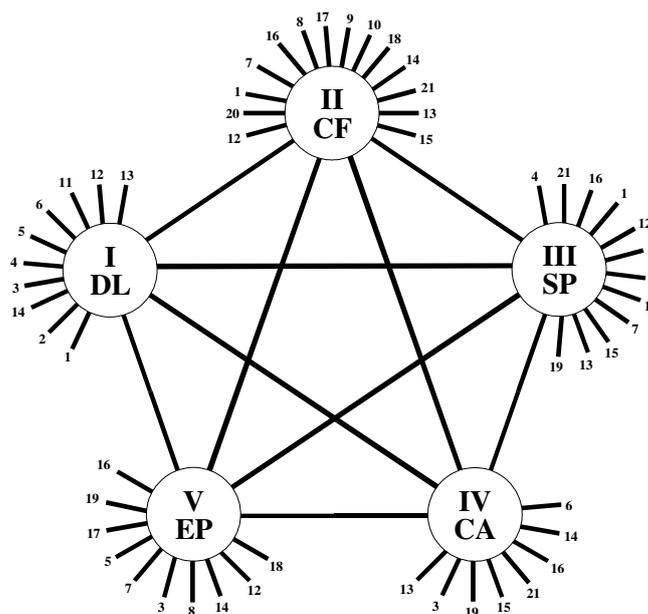
20) Comparar diferentes processos de formação socioeconômica:

- identificando-os em seu contexto histórico;
- estabelecendo entre eles uma seqüência temporal.

21) Dado um quadro informativo sobre uma realidade histórico-geográfica:

- contextualizar eventos históricos numa seqüência temporal;
- compreender a relação sociedade/natureza no arranjo espacial específico;
- destacar fatores sociais, econômicos, políticos e culturais constitutivos desses eventos em configurações sociais específicas;
- fundamentar o caráter constitutivo destes fatores, relacionando a vinculação de conceitos com unidades temporais e espaciais em que são significativos.
-

Matriz de Análise de Desempenho



- I. Dominar linguagens
- II. Compreender fenômenos
- III. Enfrentar situações-problema
- IV. Construir argumentações
- V. Elaborar propostas

Cada uma das habilidades foi medida três vezes – três questões objetivas para cada uma, devidamente calibradas pelo pré-teste como de dificuldade alta, média e baixa.

ANEXO III - Matriz de Referência de Matemática e Suas Tecnologias

Competência de área 1 - Construir significados para os números naturais, inteiros, racionais e reais.

H1 - Reconhecer, no contexto social, diferentes significados e representações dos números e operações - naturais, inteiros, racionais ou reais.

H2 - Identificar padrões numéricos ou princípios de contagem.

H3 - Resolver situação-problema envolvendo conhecimentos numéricos.

H4 - Avaliar a razoabilidade de um resultado numérico na construção de argumentos sobre afirmações quantitativas.

H5 - Avaliar propostas de intervenção na realidade utilizando conhecimentos numéricos.

Competência de área 2 - Utilizar o conhecimento geométrico para realizar a leitura e a representação da realidade e agir sobre ela.

H6 - Interpretar a localização e a movimentação de pessoas/objetos no espaço tridimensional e sua representação no espaço bidimensional.

H7 - Identificar características de figuras planas ou espaciais.

H8 - Resolver situação-problema que envolva conhecimentos geométricos de espaço e forma.

H9 - Utilizar conhecimentos geométricos de espaço e forma na seleção de argumentos propostos como solução de problemas do cotidiano.

Competência de área 3 - Construir noções de grandezas e medidas para a compreensão da realidade e a solução de problemas do cotidiano.

H10 - Identificar relações entre grandezas e unidades de medida.

H11 - Utilizar a noção de escalas na leitura de representação de situação do cotidiano.

H12 - Resolver situação-problema que envolva medidas de grandezas.

H13 - Avaliar o resultado de uma medição na construção de um argumento consistente.

H14 - Avaliar proposta de intervenção na realidade utilizando conhecimentos geométricos relacionados a grandezas e medidas.

Competência de área 4 - Construir noções de variação de grandezas para a compreensão da realidade e a solução de problemas do cotidiano.

H15 - Identificar a relação de dependência entre grandezas.

H16 - Resolver situação-problema envolvendo a variação de grandezas, direta ou inversamente proporcionais.

H17 - Analisar informações envolvendo a variação de grandezas como recurso para a construção de argumentação.

H18 - Avaliar propostas de intervenção na realidade envolvendo variação de grandezas.

Competência de área 5 - Modelar e resolver problemas que envolvem variáveis socioeconômicas ou técnico-científicas, usando representações algébricas.

H19 - Identificar representações algébricas que expressem a relação entre grandezas.

H20 - Interpretar gráfico cartesiano que represente relações entre grandezas.

H21 - Resolver situação-problema cuja modelagem envolva conhecimentos algébricos.

H22 - Utilizar conhecimentos algébricos/geométricos como recurso para a construção de argumentação.

H23 - Avaliar propostas de intervenção na realidade utilizando conhecimentos algébricos.

Competência de área 6 - Interpretar informações de natureza científica e social obtidas da leitura de gráficos e tabelas, realizando previsão de tendência, extrapolação, interpolação e interpretação.

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

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

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

Competência de área 7 - Compreender o caráter aleatório e não determinístico dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas, determinação de amostras e cálculos de probabilidade para interpretar informações de variáveis apresentadas em uma distribuição estatística.

H27 - Calcular medidas de tendência central ou de dispersão de um conjunto de dados

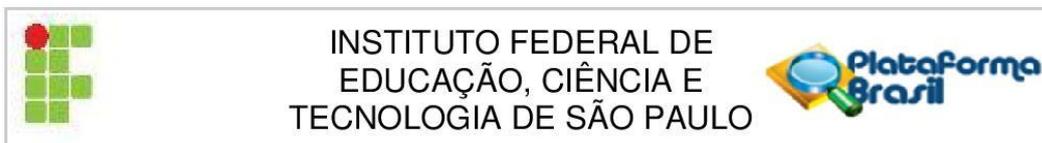
expressos em uma tabela de frequências de dados agrupados (não em classes) ou em gráficos.

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

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

H30 - Avaliar propostas de intervenção na realidade utilizando conhecimentos de estatística e probabilidade.

ANEXO IV - Parecer Consubstanciado do Comitê de Ética em Pesquisa



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Desempenho entre homens e mulheres em questões de matemática no ENEM que apresentam, ou não, leitura e interpretação

Pesquisador: EMILIANO AUGUSTO CHAGAS

Área Temática:

Versão: 3

CAAE: 70999417.0.0000.5473

Instituição Proponente: INSTITUTO FEDERAL DE EDUCACAO, CIENCIA E TECNOLOGIA DE SAO

Patrocinador Principal: Financiamento Próprio

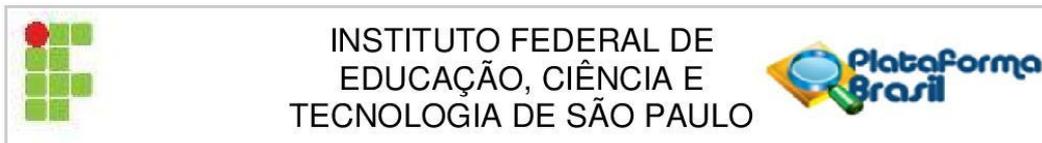
DADOS DO PARECER

Número do Parecer: 2.317.021

Apresentação do Projeto:

Segundo o embasamento teórico do projeto os homens levam vantagem em exames de matemática frente às mulheres. Explicações para esse fenômeno vão de contexto sócio cultural, do papel do homem e da mulher na sociedade, até problemas de ansiedade enfrentado por mulheres e o papel da competição que cativa mais os homens. Mas é fato que os homens têm uma aparente maior aptidão em matemática do que as mulheres. Em contrapartida, as mulheres levam vantagem em habilidades de leitura se comparadas com os homens. Surge uma questão bem relevante: ocorre diferença entre os gêneros em questões de matemática que envolvem texto, leitura e interpretação? Com base nos microdados do ENEM é possível explorar essa pergunta. O objetivo desse projeto é avaliar o desempenho em matemática obtido em cada questão da prova de múltipla escolha, particularmente nas questões que envolvem texto, leitura e interpretação. Para tal propósito, serão confrontados os resultados obtidos pelos microdados do ENEM e uma avaliação com redução de carga cognitiva, associada ao texto, aplicada a alunos na forma de um simulado. O principal objetivo buscado por esse projeto é apresentar para professores resultados referentes à discussão sobre matemática e sexo, para que eles possam se beneficiar dessas informações.

Endereço: Rua Pedro Vicente, 625
Bairro: Canindé **CEP:** 01.109-010
UF: SP **Município:** SAO PAULO
Telefone: (11)3775-4665 **Fax:** (11)3775-4570 **E-mail:** cep_ifsp@ifsp.edu.br



Continuação do Parecer: 2.317.021

Objetivo da Pesquisa:

De acordo com o autor:

Objetivos primários e secundários:

Considerando a nota da prova de matemática obtida pela teoria clássica de testes ou da teoria de resposta ao item, como é o desempenho feminino em relação ao masculino?

Dentre os eixos de conhecimento de matemática o desempenho feminino apresenta sempre o mesmo déficit em relação ao masculino?

Ocorre uma diferença de correlação entre a nota feminina e a nota masculina entre as provas de matemática e redação?

Qual é o impacto do efeito socioeconômico no desempenho entre homens e mulheres na prova de matemática do ENEM?

Ocorre uma correlação entre a nota feminina e a nota masculina entre as questões de matemática que apresentam texto e interpretação, e a redação?

Como utilizar as respostas das perguntas acima para diminuir ou anular a diferença de desempenho entre homens e mulheres em provas de matemática?

Contribuir para a informação dos professores sobre a diferença entre os gêneros em matemática e avaliações de matemática.

O principal objetivo buscado por esse projeto é apresentar para professores resultados referentes à discussão sobre matemática e gênero, para que eles possam se beneficiar dessas informações evitando estereótipos.

Avaliação dos Riscos e Benefícios:

Riscos:

A pesquisa envolve risco mínimo, cognitivos ou de viés de gênero, uma vez que as duas avaliações são equivalentes, e portanto, o construto matemático de cada item será apresentado para todos os respondentes

Benefícios:

A principal contribuição do trabalho é apresentar para os professores resultados sobre a diferença entre os gêneros em questões de matemática, particularmente nas que envolvem leitura e interpretação. Professores mais bem informados sobre a diferença entre os gêneros em matemática poderiam colocar questões em avaliações que envolvessem bastante texto e interpretação em diversos eixos do conhecimento matemático, dessa maneira o entendimento das mulheres, nesses eixos, melhoraria em média, e forçaria os homens a aprenderem matemática

Endereço: Rua Pedro Vicente, 625

Bairro: Canindé

CEP: 01.109-010

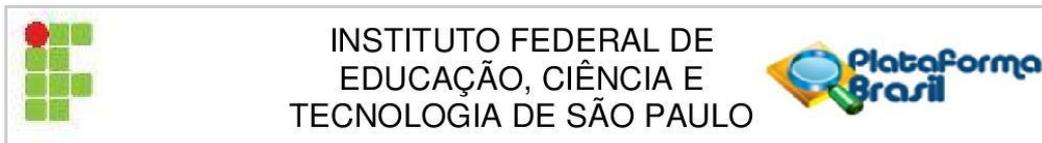
UF: SP

Município: SAO PAULO

Telefone: (11)3775-4665

Fax: (11)3775-4570

E-mail: cep_ifsp@ifsp.edu.br



Continuação do Parecer: 2.317.021

nesses eixos com base em texto, leitura e interpretação, que para eles é igualmente bom já que em média eles tem um desempenho pior nessa área que as mulheres.

Comentários e Considerações sobre a Pesquisa:

O projeto parte do princípio de que os homens levam vantagem em exames de matemática quando comparado com as mulheres. Segundo o pesquisador é fato que, pela literatura, os homens possuem um desempenho melhor em avaliações em matemática do que as mulheres. Em contrapartida, as mulheres levam vantagem em exames que exploram habilidades de leitura se comparadas com os homens.

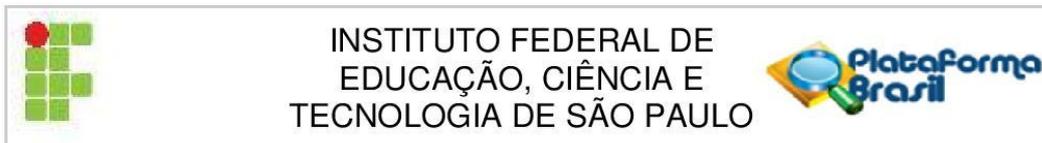
A proposta é partir de um banco de dados aberto de uma avaliação em larga escala nacional, como o ENEM, é possível comparar os desempenhos, e com uma prova de múltipla escolha aplicada a alunos do fim do ensino médio, é possível obter respostas com relação ao texto, contexto e interpretação, fazendo uma seleção adequada de itens. A hipótese é verificar se o desempenho em questões elementares do ENEM que necessitam (ou não) de leitura e interpretação apresentam alguma diferença que possa ser atribuída ao gênero dos candidatos.

Com os dados dos resultados do ENEM e com as informações dos candidatos (gênero, tipo de escola, se o aluno é ou não concludente do ensino médios) o pesquisador vai classificar as questões, conforme a matriz de referência do ENEM, nas competências disponíveis, e então identificar aquelas que apresentam texto (ou contexto) e cuja resolução passe pela extração de dados e informações desse texto. Serão selecionados apenas os candidatos concluintes do ensino médio, pressupondo que a maioria desses candidatos levam a avaliação a sério. Aplicando-se ferramentas estatísticas pretende-se observar as diferenças ocorridas entre os gêneros em questões de competências que apresentam e não apresentam muito texto (ou contexto), bem como as notas obtidas nas redações confrontadas com o desempenho nesse tipo de questão.

Após a análise desses dados serão selecionadas questões do ENEM que apresentam texto e contexto, e questões réplicas livres de contexto. Com diversos itens, com e sem contexto, é possível montar um simulado e aplicar para alunos, e então confrontar esses resultados com aqueles obtidos através dos microdados do ENEM.

Essas duas provas serão aplicadas como simulado ENEM de matemática para algumas turmas do 2o e 3o anos do ensino médio do IFSP campus São Paulo e, dependendo de parcerias, outros campus do IFSP. O projeto pretende analisar mais de 150 respondentes e em uma proporção não discrepante entre homens e mulheres (entre 40% e 60%). Ao final, um tratamento estatístico permitirá

Endereço: Rua Pedro Vicente, 625
Bairro: Canindé **CEP:** 01.109-010
UF: SP **Município:** SAO PAULO
Telefone: (11)3775-4665 **Fax:** (11)3775-4570 **E-mail:** cep_ifsp@ifsp.edu.br



Continuação do Parecer: 2.317.021

mensurar a diferença de desempenho, caso exista.

Segundo o autor o principal objetivo da pesquisa é apresentar para professores resultados referentes à discussão sobre matemática e gênero, para que eles possam se beneficiar dessas informações evitando estereótipos. O autor espera contribuir com o processo ensino aprendizagem dos docentes. Com base nas informações obtidas na pesquisa, os professores poderiam colocar questões em avaliações que envolvessem bastante texto e interpretação em diversos eixos do conhecimento matemático, melhorando dessa maneira o entendimento das mulheres, nesses eixos, melhoraria em média, e forçaria os homens a aprenderem matemática nesses eixos com base em texto, leitura e interpretação, que para eles é igualmente bom já que em média eles tem um desempenho pior nessa área que as mulheres.

Considerações sobre os Termos de apresentação obrigatória:

No Termo de Assentimento Livre e Esclarecido (TALE) e no Termo de Consentimento Livre e Esclarecido (TCLE) o pesquisador inseriu o título da pesquisa "Desempenho entre homens e mulheres em questões de matemática no ENEM que apresentam, ou não, leitura e interpretação", que deixa evidente o estudo de desempenho entre os sexos, além de explicitar mais uma vez nos objetivos .

Tanto no TALE quanto no TCLE foi inserido "Sobre os dados: 1) As provas físicas ficarão com os alunos; 2) Os gabaritos serão mantidos em arquivo por 20 anos, sendo picotados e reciclados após esse tempo; 3) Os resultados digitais das escolhas nos gabaritos serão mantidos por 30 anos". No projeto detalhado também inseri essa parte no final de metodologia.

Recomendações:

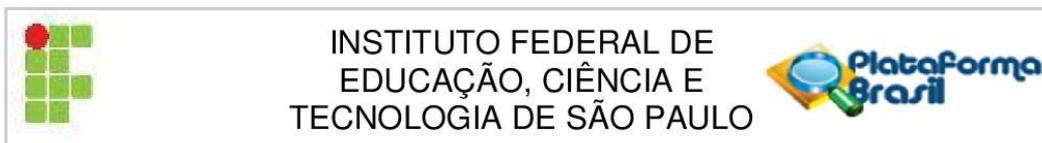
O projeto continua com alguns trechos que insinuam que há uma superioridade de um gênero sobre o outro, sugere-se uma revisão do projeto de forma que fique mais claro a colocação.

A metodologia também deveria ser revista ou melhor explicitada, pois, se uma pessoa consegue solucionar uma questão matemática após a leitura e interpretação de um texto, por que não o faria sem ter a necessidade destas etapas?

Conclusões ou Pendências e Lista de Inadequações:

Não há pendências ou lista de inadequações, as alterações solicitadas pelo CEP foram atendidas.

Endereço: Rua Pedro Vicente, 625
Bairro: Canindé **CEP:** 01.109-010
UF: SP **Município:** SAO PAULO
Telefone: (11)3775-4665 **Fax:** (11)3775-4570 **E-mail:** cep_ifsp@ifsp.edu.br



Continuação do Parecer: 2.317.021

Considerações Finais a critério do CEP:

Aprovado conforme parecer do relator.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_934909.pdf	21/09/2017 17:08:47		Aceito
Outros	Para_o_comite_ver2.pdf	21/09/2017 17:07:28	EMILIANO AUGUSTO CHAGAS	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_Detalhado_ver3.pdf	21/09/2017 17:06:19	EMILIANO AUGUSTO CHAGAS	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE_IFSP_ver2.pdf	21/09/2017 17:06:01	EMILIANO AUGUSTO CHAGAS	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_IFSP_ver2.pdf	21/09/2017 17:05:51	EMILIANO AUGUSTO CHAGAS	Aceito
Folha de Rosto	CapaEmiliano.pdf	21/09/2017 17:05:30	EMILIANO AUGUSTO CHAGAS	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

SAO PAULO, 05 de Outubro de 2017

Assinado por:
Thomas Edson Filgueiras Filho
(Coordenador)

Endereço: Rua Pedro Vicente, 625
Bairro: Canindé **CEP:** 01.109-010
UF: SP **Município:** SAO PAULO
Telefone: (11)3775-4665 **Fax:** (11)3775-4570 **E-mail:** cep_ifsp@ifsp.edu.br