

EDY CÉLIA COELHO

**TEORIA DA RESPOSTA AO ITEM - DESAFIOS E
PERSPECTIVAS EM EXAMES MULTIDISCIPLINARES**

CURITIBA

2014

EDY CÉLIA COELHO

**TEORIA DA RESPOSTA AO ITEM - DESAFIOS E
PERSPECTIVAS EM EXAMES MULTIDISCIPLINARES**

Tese apresentada como requisito parcial à obtenção do grau de Doutor em Ciências, pelo Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Setor de Tecnologia, Universidade Federal do Paraná.

Orientador: Prof. PhD. Paulo Justiniano Ribeiro Jr.

CURITIBA

2014

Termo de Aprovação

EDY CÉLIA COELHO

TEORIA DA RESPOSTA AO ITEM - DESAFIOS E PERSPECTIVAS EM EXAMES MULTIDISCIPLINARES

Tese aprovada como requisito parcial para obtenção do grau de Doutor em Ciências, pelo Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Setor de Tecnologia, Universidade Federal do Paraná, pela seguinte banca examinadora:

Prof. PhD. Paulo Justiniano Ribeiro Jr.
Universidade Federal do Paraná

Prof. Dr. Pedro Alberto Barbetta
Universidade Federal de Santa Catarina

Prof. Dr. Paulo Sérgio Macuchen Noga
Pontifícia Universidade Católica do Paraná

Prof. Dr. Julio Sílvio de Souza Bueno Filho
Universidade Federal de Lavras

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

Dedico esta extensão da vida, Tese, à:

Deus, pela existência dos espíritos

Jesus, pelo exemplo de amor

Maria, pela proteção de mãe. “Passa na frente...”

Minha Mãe, Cléia ao som infinito

“Tem dias que eu acordo pensando em você
Em fração de segundos vejo o mundo desabar
E aí que cai a ficha que eu não vou te ver
Será que esse vazio um dia vai me abandonar?
Tem gente que tem cheiro de rosa de avelã
Tem um perfume doce de toda manhã
Você tem tudo, você tem muito
Muito mais que um dia eu sonhei pra mim
Tem a pureza de um anjo querubim
Eu trocaria tudo pra te ter aqui
Eu troco minha paz por um beijo seu
Eu troco meu destino pra viver o seu
Eu troco minha cama pra dormir na sua
Eu troco mil estrelas pra te dar a lua
E tudo que você quiser” (L.Santana)

Minha Paixão, Lipe e Mary Luly ao som infinito

“Vivemos esperando
O dia em que seremos melhores
Melhores no amor
Melhores na dor
Melhores em tudo” (Jota Quest)

A vida:

Metal Contra as Nuvens:

“Não sou escravo de ninguém
Ninguém, senhor do meu domínio
Sei o que devo defender
E, por valor eu tenho
E temo o que agora se desfaz
Viajamos sete léguas
Por entre abismos e florestas
Por Deus nunca me vi tão só
É a própria fé o que destrói
Estes são dias desleais
Eu sou metal, raio, relâmpago e trovão
Eu sou metal, eu sou o ouro em seu brasão
Eu sou metal, me sabe o sopro do dragão
Reconheço meu pesar
Quando tudo é traição
O que venho encontrar
É a virtude em outras mãos
Minha terra é a terra que é minha
E sempre será
Minha terra tem a lua, tem estrelas
E sempre terá”

“Quase acreditei na sua promessa
E o que vejo é fome e destruição
Perdi a minha sela e a minha espada
Perdi o meu castelo e minha princesa
Quase acreditei, quase acreditei
E, por honra, se existir verdade
Existem os tolos e existe o ladrão
E há quem se alimente do que é roubado
Mas vou guardar o meu tesouro”

“É a verdade o que assombra
O descaso que condena
A estupidez, o que destrói
Eu vejo tudo que se foi
E o que não existe mais
Tenho os sentidos já dormentes
O corpo quer, a alma entende
Esta é a terra-de-ninguém
Sei que devo resistir
Eu quero a espada em minhas mãos
Eu sou metal, raio, relâmpago e trovão
Eu sou metal, eu sou o ouro em seu brasão
Eu sou metal, me sabe o sopro do dragão
Não me entrego sem lutar
Tenho, ainda, coração
Não aprendi a me render
Que caia o inimigo então”

“Tudo passa, tudo passará
E nossa estória não estará pelo avesso
Assim, sem final feliz
Teremos coisas bonitas pra contar
E até lá, vamos viver
Temos muito ainda por fazer
Não olhe pra trás
Apenas começamos
O mundo começa agora
Apenas começamos” (Renato Russo)

Agradecimentos

Agradeço a Deus pelo seu infinito amor e por me abençoar proporcionando pais, irmã, família, mestres e amizade de tantas pessoas que até mesmo através de atos simples da vida contribuíram para que eu chegasse ao fim deste trabalho. Com refrões musicais que marcaram cada etapa da jornada, gostaria de agradecer em especial:

Aos meus pais Eduardo e Cléia. “Eu tenho tanto prá lhe falar, mas com palavras não sei dizer como é grande o meu amor por vocês...”

A minha irmã. “Nada do que foi será. De novo do jeito que já foi um dia. Tudo passa tudo sempre passará. A vida vem em ondas como um mar. Num indo e vindo infinito”. “Eu não sei o que seria de mim sem você minha irmã”. “ Estendo a compreensão, as escritas, as discussões nos fundos da casa...por tudo Mario, de coração não sei como retribuir e sim dizer... ao segundo melhor,...agora...deixa para o Lue o primeiro.”

A minha Paixão. “So, so you think you can tell. Heaven from Hell. Blue skies from pain.” “O amor é paciente, é prestativo, não é invejoso, não se ostenta, não se incha de orgulho...Tudo desculpa, tudo crê, tudo espera, tudo suporta?” Eu não sei ficar sem seu amor...sem seu abraço..sem você...então acelera...porque? “ Além do horizonte existe um lugar, bonito e tranqüilo, pra gente se amar”.

Ao meu Lipe. “E eu sou feliz e canto e o universo é uma canção, e eu vou que vou, história, nossas histórias, dias de luta, dias de glória”. Exemplo de responsabilidade e filho pai do amor.

A minha Maria Luiza. “Aos olhos do Pai você é obra-prima que Ele planejou, com suas próprias mãos pintou. A cor de sua pele, os seus cabelos desenhou, cada detalhe num toque de amor”. Disputa desejada de amor, para alegrar e tagarelar...

A cumadi Ana Luisa. “Esse ano quero paz no meu coração...O tempo passa e com ele caminhamos todos juntos sem parar. Nossos passos pelo chão vão ficar...”. Entra ano, passa o ano e a esperança sempre esteve conosco, né cumadi... Obrigada por estar sempre comigo e fazer parte da minha família, te amo.

Aos amigos de festa, “Uh...uh...ah...ah...Oh! Mundo tão desigual. Tudo é tão

desigual”. Pelo companheirismo de tudo, angustia, alegrias, loucuras enfim pelos copos e festas tirando-me do mundo real para o imaginário. Isso tudo graças a prima/amiga Bay e ao Marcelo com seu “ bom dia ...bom dia..hoje estou tão feliz...” acordando-nos quase todo domingo!

Ao primo Marcos, Marlize e cia. Ltda, “tira o sol daqui...tire o sol de lá...”, porque nosso lema é ser feliz.

As minhas amigas e amigos de Cascavel por estarem sempre perto do meu coração. “Longe de casa ...a mais de uma semana por milhas e milhas...”. Ah!!! minha vizinhas/amigas que saudades de ir “tchau...tchau...”.

As amigas em especial Vanessa, Naimara e Karine, enfim aos amigos do LEG. “Deixa a vida me levar...vida leva eu...”, nem vou citar todos os nomes, porque cada um em especial nem sei como agradecer pela ajuda, pelos seminários, discussões e com o toque todo especial das risadas, lágrimas, consolos e confidências compartilhados.

Ao meu orientador Prof. Paulo Justiniano por estar disposto, durante todo este tempo a me orientar, pela amizade, conselhos e pelo lanche das 4h “Café...Café...” sem contar da disposição contagiante do “Pedala Curitiba”. “...Fila do próximo...”. O verdadeiro professor com a surpreendente facilidade de criar e contextualizar o conhecimento. Admiração incontestável.

Ao Prof. Wagner H. Bonat, simplesmente Waagnnerrr...fala lá... dos porque...para que...e daí, pelas orientações e pelas discussões, enfim, pela paciência e amizade... conhecido como anjo do LEG. Fica aqui o reconhecimento e a total admiração pela pessoa humana, humilde e por compartilhar. Que Deus o abençoe em todo seu caminhar!

Ao Prof. Elias Teixeira Krainski que apesar da distância, ainda assim respondia com rapidez minhas dúvidas, ou seja, não existe fuso horário para o grande “ELIAS”.

Ao Prof. Walmes Zeviani pela ajuda no aprimoramento de comandos e escrita.

Ao Prof. Pedro Barbeta por se dispor à discutir sobre pontos críticos da tese, encaminhando documentos e estudos; e por me receber sempre com simpatia e hospitalidade na UFSC.

À Emilio Kavamura o professor “anjo”, pela amizade, por candidatar-se a leitura dessa tese em tex, com palavras de energia ...“calma” e com preciosas contribuições. Obrigada!

Aos grandes amigos Prof. Mozart, Prof. Willian e Prof. Paulo Nogas pela

recepção, amizade e companheirismo na PUC, pelas discussões, pelos dados e ajuda no caminhar. Amigos do presente e para sempre.

A todas as amigas e amigo da sala 308 pela força nos ajustes finais da tese, meu carinho e obrigada. Especialmente a Gianna por dedicar seu tempo em corrigir “sem pregar me na parede” e a Katya pela sintonia da amizade.

À Secretaria do Estado da Educação do Paraná, que promoveu o afastamento por dois anos para finalizar o doutorado, o meu reconhecimento como profissional do quadro próprio do magistério; e principalmente pela oportunidade de avançar na carreira como educadora.

Sumário

Lista de Figuras	xii
Lista de Tabelas	xv
Lista de Siglas	xix
Resumo	xxi
Abstract	xxii
1 INTRODUÇÃO	1
1.1 CONTEXTO E JUSTIFICATIVA	1
1.2 OBJETIVO.....	6
1.3 ESTRUTURA DA TESE	7
2 AVALIAÇÃO EDUCACIONAL	8
2.1 DADOS DE AVALIAÇÕES INTERNAS.....	9
2.1.1 Conjunto de dados I	10
2.1.2 Conjunto de dados II.....	20
2.2 DADOS DE AVALIAÇÃO EXTERNA	25
2.2.1 Conjunto de dados III	27
3 METODOLOGIA	35
3.1 TEORIA CLÁSSICA DOS TESTES	36
3.1.1 Modelo da Teoria Clássica dos Testes.....	37
3.2 TEORIA DA RESPOSTA AO ITEM	38

3.2.1	Considerações	39
3.2.2	Modelos da TRI	41
3.2.3	Modelos de k dimensões	43
3.2.4	Modelos Exploratório e Confirmatório	55
3.2.5	Estimação dos Parâmetros - Inferência	55
3.2.6	Métodos Computacionais	60
3.2.6.1	Pacotes <i>ltm</i> e <i>mirt</i>	64
3.2.7	Dimensionalidade dos dados	65
3.2.7.1	Procedimentos para determinar o número necessário de dimensões na TRI	67
3.2.8	Escores da TRI	70
4	RESULTADOS E DISCUSSÕES	72
4.1	MODELO LOGÍSTICO DA TRI NO CONJUNTO DE DADOS I	73
4.1.1	Dimensionamento do conjunto de dados I	74
4.1.2	TRIUI no conjunto de dados I	76
4.1.2.1	Estimação dos parâmetros nos modelos da TRIUI no conjunto de dados I .	80
4.1.2.2	Curva característica dos itens do conjunto de dados I - TRIUI	82
4.1.2.3	Escores das habilidades do conjunto de dados I - TRIUI	86
4.1.3	TRIM no conjunto de dados I	92
4.1.3.1	Estimação dos parâmetros nos modelos da TRIM no conjunto de dados I .	95
4.1.3.2	Cargas fatoriais e CCI dos itens na TRIM no conjunto de dados I	96
4.1.3.3	Escores das habilidades da TRIM no conjunto de dados I	99
4.2	MODELO LOGÍSTICO DA TRI NO CONJUNTO DE DADOS II	102
4.2.1	Dimensionamento do conjunto de dados II	103
4.2.2	TRIM no conjunto de dados II	104
4.2.2.1	Estimação dos parâmetros nos modelos da TRIM no conjunto de dados II	106
4.2.2.2	CCI e cargas fatoriais dos itens na TRIM no conjunto de dados II	107

4.2.2.3	Escores das habilidades da TRIM no conjunto de dados II	113
4.3	MODELO LOGÍSTICO DA TRI NO CONJUNTO DE DADOS III	117
4.3.1	Dimensionamento do conjunto de dados III.....	118
4.3.2	TRIM no conjunto de dados III.....	119
4.3.2.1	Estimação dos parâmetros nos modelos da TRIM no conjunto de dados III	121
4.3.2.2	CCI e cargas fatoriais dos itens na TRIM no conjunto de dados III	122
4.3.2.3	Escores das habilidades da TRIM no conjunto de dados III	127
4.3.3	Estimação Bayesiana para o parâmetro do acerto ao acaso.....	130
5	CONCLUSÕES E SUGESTÕES DE TRABALHOS FUTUROS	140
	Referências	143
	APÊNDICE A – Índices de facilidade conforme a proporção de acertos, índices de discriminação de acordo com o ponto bisserial e o <i>Cronbach's alpha</i> na prova geral dos 443 respondentes.....	150
	APÊNDICE B – Índices de facilidade conforme a proporção de acertos, índices de discriminação de acordo com o ponto bisserial e o <i>Cronbach's alpha</i> na prova geral dos 436 respondentes com os 32 itens, excluindo as questões (G1, Q23) e as (G1, G2 e Q23).....	153
	APÊNDICE C – Código fonte R - estimação dos parâmetros dos itens com métodos para acerto ao acaso.....	156
	APÊNDICE D – Código fonte R - transformação dos escores dos modelos da TRI para escala padronizada entre 0 a 1	159
	APÊNDICE E – Escore dos 94 respondentes das habilidades conforme os modelos propostos pela TRI.....	162
	APÊNDICE F – Escores e escala padronizada dos 94 e 75 respondentes do MLD1 indicados pela TRIU e TRIM	165

APÊNDICE G – Escores e escala padronizada dos 189 e 154 respondentes do conjunto de dados II.....	168
APÊNDICE H – Estimativas dos parâmetros dos modelos do conjunto de dados III - excluindo G1;Q23 e G1;G2;Q23.....	173
APÊNDICE I – Cargas fatoriais dos modelos para as 32 questões e retirando a G1 e Q23	176
APÊNDICE J – Escores e escala padronizada do Enade no MLD1-3PL para 436 graduados de estatística - 2009.....	178
ANEXO A – Dados do exame multidisciplinar do conjunto de dados I.	184
ANEXO B – Dados do exame multidisciplinar do conjunto de dados II.	186

Lista de Figuras

Figura 1	Porcentagem dos acertos das questões com as seis habilidades	13
Figura 2	Frequência de acertos do valor observado da habilidade total e o número esperado de acerto, casual	14
Figura 3	Somatório dos acertos das questões das seis habilidades com a habilidade geral	15
Figura 4	Frequência de acertos do valor observado da habilidade total e o número esperado de acerto casual	24
Figura 5	Frequência de acertos do valor observado e o número esperado de acerto casual dos 436 alunos que responderam o Enade/2009	33
Figura 6	Diferentes modelos da TRIU de acordo com a hierarquia	50
Figura 7	CCI do modelo unidimensional (a) e multidimensional (b) com o ML2 do mesmo item	51
Figura 8	Diagnóstico da dimensão com: (a) 94 respondentes e (b) 75 respondentes.	74
Figura 9	CCI para a habilidade geral dos 94 respondentes do $M_2(\alpha, \beta_i, 0)$	83
Figura 10	CCI para H.QGI - $M_2(\alpha, \beta_i, 0)$ H.IEQ - $M_1(1, \beta_i, 0)$, H.GAA - $M_4(\alpha, \beta_i, c)$, H.CDI, H.FGE e H.EST - $M_1(1, \beta_i, 0)$ dos 94 respondentes	84

Figura 11	CCI para a habilidade geral dos 75 respondentes do $M_4 (\alpha, \beta_i, c)$	85
Figura 12	CCI para H.QGI e H.CDI com o $M_1 (1, \beta_i, 0)$, H.FGE $M_2 (\alpha, \beta_i, 0)$ e H.IEQ, H.GAA E H.EST com o $M_4 (\alpha, \beta_i, c)$ dos 75 respondentes	86
Figura 13	CCI da Q20 (a), Q21 (b) e Q4 (c) do MLD2- 2PL com 94 respondentes	98
Figura 14	Escores dos 94 e 75 respondentes na escala padronizada	100
Figura 15	Correlação entre: os escores do pacote <i>mirt</i> / MLD1-2PL & <i>ltm</i> / M_5 e a escala padronizada com a MLD1-2PL e TCT	101
Figura 16	Diagnóstico da dimensão dos 189 (a) e 154 (b) respondentes	103
Figura 17	CCI dos 154 respondentes da Q29 e Q2 do MLD1/3PL	108
Figura 18	CCI dos 189 e 154 respondentes do MLD1- 3PL	109
Figura 19	Influências dos zeros do MLD1- 3PL na CCI	110
Figura 20	CCI da Q30 dos 189 (a) e 154 (b) respondentes MLD2-2PLf	113
Figura 21	Escalas padronizadas dos 189 e 154 respondentes	116
Figura 22	Diagnóstico da dimensão dos 436 respondentes com: 32 questões; 30 questões excluindo a G1 e Q23	118
Figura 23	CCI do MLD1-3PL para 436 respondentes com 32, 30 e 29 itens	123

Figura 24	Curva característica dos itens G1, G2 e Q23 no MLD1-3PL	125
Figura 25	CCI do MLD1-3PL dos itens corretos para os alunos 5 e 12 com 32 itens e excluindo G1, G2 e Q23	129
Figura 26	Escore dos respondentes pela TRIU: (a) dispersão entre as Notas/TRIU dos respondentes e (b) histograma das Notas/TRIU	130
Figura 27	CCI do MLD1-3PLp : (a) com 32 itens e (b) itens Q1, Q2 e Q23 do Enade-2009 de estatística	134
Figura 28	Escala padronizada e a correlação das notas da TCT com a escala padronizada do MLD1-3PLp	135
Figura 29	Diagnóstico da dimensão dos 436 respondentes com 24 questões específicas	136
Figura 30	Correlação das notas com a escalas padronizadas das 32 e 24 questões	138

Lista de Tabelas

Tabela 1	Disciplinas avaliadas e suas siglas	11
Tabela 2	Proporção de acertos por questões	12
Tabela 3	Frequência de acertos do valor observado e o número esperado de acertos sob a hipótese de respostas aleatórias	13
Tabela 4	Correlação ponto bisserial e <i>Cronbach's alpha</i>	18
Tabela 5	Resultados descritivos do exame multidisciplinar do conjunto de dados II	21
Tabela 6	Frequência de acertos do valor observado e o número esperado de acertos casuais do conjunto de dados II	23
Tabela 7	Classificações das questões objetivas segundo índice facilidade e o índice de discriminação - Enade/2009	29
Tabela 8	Índices de facilidade conforme a proporção de acertos, índices de discriminação de acordo com o ponto bisserial e o <i>Cronbach's alpha</i> para as questões da formação geral e específica	30
Tabela 9	Frequência de acertos do valor observado e o número esperado de acerto casual do CD3	32
Tabela 10	Resultados dos ajustes de diferentes modelos da TRIU considerado para	

	a prova	77
Tabela 11	Resultados dos ajustes de diferentes modelos da TRIU por áreas	78
Tabela 12	Modelos da TRI escolhidos conforme as habilidades	80
Tabela 13	Parâmetros estimados da TRI para a prova geral e por áreas	81
Tabela 14	Comparação dos escores de alguns alunos posicionados - TCT e TRI	87
Tabela 15	Comparação dos escores da TCT e da TRI de alguns alunos posicionados com 94 e 75 respondentes	89
Tabela 16	Comparação dos escores e escala padronizada da TCT e da TRIU de alguns alunos com 94 e 75 respondentes	90
Tabela 17	Resultados das estatísticas de diagnósticos da qualidade dos 94 e 75 respondentes do ajuste dos modelos do conjunto de dados I	94
Tabela 18	Resultados das estimativas dos parâmetros dos MLD1 e MLD2 para os 94 e MLD1 para 75 respondentes	95
Tabela 19	Cargas fatoriais e comunalidades dos MLD1 e MLD2 para o CD1	97
Tabela 20	Comparação dos escores dos 20 primeiros alunos dos 94 e 75 respondentes dos modelos indicados para TRIM	99
Tabela 21	Valores observados no pacote <i>ltm</i> (M_5) e <i>mirt</i> (2PL) no MLD1	102
Tabela 22	Resultados das estatísticas de diagnósticos da qualidade dos 189 e 154 respondentes do ajuste dos modelos do conjunto de dados II	105

Tabela 23	Resultados das estimativas dos parâmetros dos MLD1 e MLD2 para os 189 e 154 respondentes do CD2	107
Tabela 24	Cargas fatoriais e comunalidades dos MLD1 e MLD2 do CD2	111
Tabela 25	Escore e escala padronizada dos 189 e 154 respondentes	114
Tabela 26	Resultados das estatísticas de diagnósticos da qualidade do ajuste dos modelos	120
Tabela 27	Estimativas dos parâmetros dos modelos do conjunto de dados III	121
Tabela 28	Cargas fatoriais dos modelos para as 32 questões e retirando a G1 e Q23	126
Tabela 29	Escore e escala padronizada do Enade no MLD1-3PL para 436 graduados de estatística - 2009	128
Tabela 30	Resultados das estatísticas de diagnóstico da qualidade do ajuste dos modelos	131
Tabela 31	Estimação dos parâmetros com o acerto ao acaso a <i>priori</i> no MLD1 e MLD2	132
Tabela 32	Cargas fatoriais dos MLD1 e MLD2 com o 3PLp no CD3	133
Tabela 33	Resultados das estatísticas de diagnóstico da qualidade do ajuste dos modelos para as 24 questões específicas	136
Tabela 34	Estimação dos parâmetros com o acerto ao acaso a <i>priori</i> no MLD1 e MLD2	137

Tabela 35	Cargas fatoriais dos MLD1 e MLD2 com priori	138
-----------	---	-----

Lista de Siglas

TRI	Teoria de Resposta ao Item
Enade	Exame Nacional de Desempenho dos Estudantes
MEC	Ministério da Educação e Cultura
TCT	Teoria Clássica dos Testes
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais
SAEB	Sistema de Avaliação da Educação Básica
ENEM	Exame Nacional do Ensino Médio
SINAES	Sistema de Avaliação do Ensino Superior
PUCPR	Pontifícia Universidade Católica do Paraná
CD1	Conjunto de Dados I
GAA	Geometria Analítica e Álgebra Linear
CDI	Cálculo Diferencial e Integral
FGE	Física Geral e Experimental
QGI	Química Geral e Inorgânica
EST	Estatística
IEQ	Introdução à Engenharia Química
<i>ltm</i>	<i>Latent Trait Model</i>
AC	Cronbach's alpha
CPB	Correlação Ponto Bisserial
CD2	Conjunto de Dados II
OEI	Organização dos Estados Ibero-Americanos
BID	Banco Interamericano de Desenvolvimento
CD3	Conjunto de Dados III
CNE	Conselho Nacional de Educação

N.A	Nenhuma Alternativa
IF	Índice de Facilidade
ID	Índice de Discriminação
N.Esp	Número Esperado
ML1	Modelo Logístico de um parâmetro
ML2	Modelo Logístico com dois parâmetros
ML3	Modelo Logístico de três parâmetros
TRIM	Teoria de Resposta ao Item Multidimensional
TRIU	Teoria de Resposta ao Item Unidimensional
MLD1	Modelo logístico de uma dimensão
MLD2	Modelo logístico de duas dimensões
df	Número de parâmetros do modelo
MDISCi	Norma do Vetor Discriminações
MDIFF	Dificuldade Multidimensional
MVM	Máxima Verossimilhança Marginal
EM	Esperança e Maximização
MH-RM	Metropolis Hastings -Robbins Monro
MCMC	Monte Carlo via Cadeias de Markov
logLik	Verossimilhança maximizada
AIC	Critério de Informação de Akaike
BIC	Monte Carlo via Cadeias de Markov
EP	Escala Padronizada
FA	Análise Fatorial
PC	Componentes Principais
HA	Habilidades dos Alunos
CCI	Curvas Características do Itens

Resumo

A utilização da Teoria da Resposta ao Item (TRI) na avaliação educacional tem fornecido dados sobre às habilidades necessárias e competências dos educandos, visando mensurar seus conhecimentos não no teste como um todo, mas em cada item individualmente, apresentando-se como uma alternativa viável agregada à realidade do crescimento de métodos para avaliar alunos. Esta pesquisa fornece uma visão geral dos modelos da TRI e as implicações concretas em exames multidisciplinares, compostos por três conjunto de dados. O primeiro conjunto tem 94 alunos da graduação do curso de engenharia; o segundo com 189 alunos da graduação de administração da Pontifícia Universidade Católica do Paraná e o último conjunto com 436 alunos concluintes da graduação de Estatística do Exame Nacional de Avaliação de Desempenho de Estudantes (Enade) no ano de 2009. Nos dois primeiros conjuntos de dados, o conhecimento e a preparação para o Enade foram considerados como avaliação interna da instituição, pois avaliam o desempenho. Para o terceiro conjunto de dados, foram consideradas somente as provas objetivas. A flexibilidade dos benefícios da TRI e as implicações do uso de modelos em processos educacionais são discutidas e abordadas a partir dos modelos encaixados: unidimensional (TRIU) e multidimensional (TRIM). Assim, para avaliar tanto a estimação dos parâmetros dos itens e os traços latentes dos respondentes, foi necessário que esses representassem a realidade observada e como são avaliados por tais modelos. Logo, o tema central desta pesquisa é mostrar a importância de um modelo da TRI multidimensional, capaz de levar em consideração uma prova que seja composta e, conseqüentemente, respondida por múltiplas habilidades latentes, para os testes realizados na avaliação. Os estudos descritos neste trabalho mostraram que cada modelo proposto da TRI tem suas particularidades. Foram aplicados critérios estatísticos, os quais permitiram que se fizessem críticas, escolhas e avaliações dos modelos para justificar metodológica e computacionalmente um modelo mais complexo. A interpretação dos resultados da escolha do modelo pode levar a conclusões precipitadas da TRI, principalmente em exames multidisciplinares, cuja dimensionalidade depende dos itens, os quais expressam as habilidades avaliadas. Os escores obtidos possibilitaram comparações convenientes da teoria clássica com a TRI na proposta da escala padronizada entre 0 e 1. O trabalho proporciona também análises para as habilidades dos alunos e métodos computacionais ainda não utilizados nas avaliações multidisciplinares. Os conjuntos de dados, foram sugeridos e aceitáveis por análises de diagnóstico estatístico, conforme o modelo, por representarem práticas em estudo, com ajustes e estimativas de parâmetros precisos e adequados.

Palavras-chave: Avaliação. Habilidades. Modelos-TRI.

Abstract

The Item Response Theory (IRT) application on the educational evaluation has been providing data related to the student's required skills and their acquired competences, aiming to quantitate their knowledge on each item of the entire process, not only on the tests individually. Therefore, it is seen as a viable alternative to be associated with the current increase of student's evaluation methods. This research allows a wide view of the IRT models and their real implications on the educational evaluations, composed by three sets of data. The first one has 94 undergraduate Engineering student's from Pontificia Universidade Católica do (PUC) Paraná. The second group is composed by 189 undergraduate students from the Business Administration Course from PUC - Paraná and the last one had 436 undergraduate students of Statistics Course from the National Examination of Students Performance (ENADE) in 2009. In the first two data sets, both knowledge and preparation for Enade were considered an internal evaluation of the institution, since they assess the student's performance. For the third data set, only the objective evaluations were considered. The TRI flexibility benefits and the implications of applying the models in educational processes are discussed and addressed from the following models: unidimensional (UIRT) and multidimensional (MIRT). Hence, in order to assess both estimation of the parameters and the latent traits of the respondents, it was necessary that they did represent the observed reality and how they are evaluated by such models. Therefore, the main objective of this research is to show the importance of an IRT multidimensional model, which is able to take into account that a test is formed and consequently answered by multiple latent skills, for the tests performed in the evaluation. The performed studies during this study have shown that each model proposed by IRT has its peculiarities. The applied statistical criteria allowed the use of critics, choices and evaluations of the model to justify both computational and methodologically a more complex model. The results interpretation concerning the model of choice can lead to IRT early conclusions, mainly in multidisciplinary tests, which dimensionality depends on the items, which express the assessed skills. The obtained scores allowed convenient comparisons of classical theory with the TRI based on the propose of the standardized scale between 0 and 1. This work also provides analyses concerning the students' abilities and computational methods which have not yet been used in these assessments. However, the data sets were suggested and accepted by consistent statistical analyses according to the model since they represent the ongoing reality, with adjustments and estimates of accurate and appropriate parameters.

Key-word: Evaluation. Abilities. IRT Models.

1 INTRODUÇÃO

Esta tese insere-se em uma área interdisciplinar, composta tanto por tópicos de Estatística como de Educação para tratar do tema Teoria de Resposta ao Item (TRI), que aplica métodos estatísticos em avaliações multidisciplinares.

A avaliação do aprendizado está em desenvolvimento e aprimora-se, uma vez que, agora, ela está voltada para as competências adquiridas do educando. No entanto, os métodos e critérios de avaliação estão ainda baseados no sistema tradicional de aquisição de conceitos e informações. Desta maneira, a transição para a avaliação por competências deve ser feita de forma criteriosa e metódica, para que os resultados possam ter métricas comparáveis, além de terem características multi e interdisciplinares.

Com objetivo de contribuir na qualidade do sistema educacional e na construção de alternativas que contemplem o avaliar, buscou-se a compreensão do desempenho e da evolução do educando, isto é, da construção da habilidade desenvolvida do saber como fazer. Esse conjunto dos saberes é a competência de como saber agir e integrar saberes múltiplos, que é desejado e apreciado nas avaliações multidisciplinares e que também é o âmago desta pesquisa.

Neste capítulo a seção 1.1 refere-se às motivações, a seção 1.2 aos objetivos da pesquisa e a seção 1.3 à estrutura dos capítulos deste trabalho.

1.1 CONTEXTO E JUSTIFICATIVA

A avaliação é um importante instrumento de medição do aprendizado. Marcar uma opção em uma prova pode ser um ato que revele plena consciência do assunto indicado no enunciado, mas, também, pode ser apenas mais um item assinalado na expectativa de não se deixar escapar alguma possível chance de acerto.

As avaliações aplicadas tanto à educação como na psicologia, para aferir o aprendizado ou a satisfação das pessoas, vêm ocupando posição de destaque, e mostra-se cada

vez mais urgente, além da necessidade de se conhecer o *status quo* num determinado momento, é importante para que se possa comparar os resultados com os de avaliações anteriores (ter uma métrica e rastreabilidade), para assim planejar e adequar as futuras avaliações de acordo com o progresso atual (ser adaptável e coerente).

Para avaliar o aprendizado, a aquisição de habilidades e as competências, de acordo com o perfil dos estudantes, instituições de ensino superior estão adotando a avaliação multidisciplinar. Este tipo de avaliação parte da ideia de que as competências profissionais são construídas ao longo do curso. Portanto, as avaliações são específicas a cada série, embora cada série dependa uma da outra. Os resultados formam uma base de informações que oferecem subsídios para a análise do perfil dos cursos, contribuindo para o trabalho de melhoria constante da qualidade do ensino e aprendizagem.

Para direcionar as instituições de ensino quanto ao seu papel pedagógico, é elaborado um exame nacional multidisciplinar, o Exame Nacional de Desempenho dos Estudantes (Enade). O exame contribui com o processo de planejamento e com a autoavaliação dos envolvidos: alunos, professores e diretores. Neste contexto, nota-se a busca por avanços nos processos de avaliação educacional.

As avaliações multidisciplinares têm o sentido figurado de uma “ponte”, tendo como ponto inicial as avaliações internas feitas pelos professores em sala, ou seja, os exames das habilidades específicas. Do outro lado, o ponto final, o conjunto dos conhecimentos (saberes), representado pelo Enade e supervisionado pelo Ministério da Educação (MEC). Tal procedimento prepara e fornece como alternativas resultados das competências que estão sendo desenvolvidas no âmbito educacional (aluno, professor e instituição), em busca de qualidade. Essa ponte a ser percorrida é um preparo para uma avaliação externa, o Enade, que credencia, aprova e classifica os saberes.

O foco inicial deste trabalho são as avaliações multidisciplinares internas aplicadas para pequenos grupos de respondentes. A elaboração das questões, o enfoque nas competências específicas de cada um dos cursos avaliados formam a trajetória, “ponte”, em que cada curso está direcionado e focado. Mesmo aplicado a uma amostra da população discente, é de suma importância utilizar os resultados desta avaliação para se entender se suas perspectivas atendem os resultados almejados nas competências e saberes do curso e da instituição.

O Enade hoje, não aplica a TRI e é visto pelas instituições como um sistema de ranqueamento de seus desempenhos tanto interno como externo. A grande escala de respondentes torna mais significativos os resultados obtidos, o que o faz um retrato

real, ainda que parcial, das instituições de ensino. Por isso é tratado como um dos pontos principais das avaliações multidisciplinares, gerando competições internas e externas entre as instituições.

O ranqueamento leva muitas instituições privadas à competição pela nota máxima no Enade. Isto é visto como um bônus para toda a instituição, pois mantém o curso em destaque e principalmente angaria mais alunos. A competição por notas máximas no Enade, de certa forma, gera automaticamente a busca da qualidade na educação, levando em consideração aspectos intrínsecos, como o corpo docente e a infraestrutura. O questionamento que se deve ter em mente é se estas avaliações multidisciplinares, que avaliam múltiplas habilidades, apresentam questões que atendem às competências desejadas dos alunos no desenvolvimento do processo ensino e aprendizagem.

Para tentar chegar mais próximo de uma conclusão real acerca do verdadeiro conhecimento de um indivíduo avaliado, foi desenvolvida, a partir das limitações da Teoria Clássica dos Testes (TCT), a Teoria da Resposta ao Item (TRI).

A TRI usa modelos estatísticos construídos para representar a probabilidade de um indivíduo responder corretamente a um item em um determinado teste. Com a intenção de discutir e apresentar o modelo estatístico que represente a avaliação, é necessário fornecer a avaliação com suas questões e escores, pois o modelo deve incorporar além da visão de que o escore seja baseado somente nas questões assinaladas, mas definido pelas questões que compõem a prova.

A TRI vem sendo utilizada nacionalmente, principalmente por órgãos de avaliação institucional como o MEC, no âmbito do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Quanto mais rápido for o contato do meio educacional com esta teoria, mais rapidamente ocorrerá a disseminação de seu uso, o que poderá trazer uma benéfica contribuição à educação, por se tratar de um elemento diferencial para melhor conhecer e retratar a avaliação multidisciplinar. Reitera-se que a avaliação tem o propósito de coletar, interpretar e aferir resultados do conhecimento, mas que erroneamente, muitas vezes, visa somente à certificação educacional ou ao ingresso nas instituições, sejam elas públicas ou privadas, nas provas de desempenho ou em processos seletivos.

A busca pela qualidade do ensino e da aprendizagem em seus diversos níveis de interesse (instituições, professores e alunos) torna-se um grande desafio para a TRI como um sistema de diagnóstico e medidor pedagógico por meio do exame multidisciplinar aplicado em pequenas amostras.

Há necessidade de incorporar os procedimentos da TRI no sistema de avaliação educacional como método e indicador de decisões para contribuir com os profissionais da educação, a fim de permitir uma compreensão melhor dos mecanismos geradores de facilidade no aprendizado, bem como um acompanhamento do desempenho dos cursos ao longo do tempo. Portanto, foram consultados instrumentos de avaliação educacional que utilizam a TRI como o Sistema de Avaliação da Educação Básica (SAEB), o Exame Nacional do Ensino Médio (ENEM) e a Prova Brasil.

Além disso, um resultado que foi obtido é a comparação das abordagens da TRI e da TCT. A comparação das abordagens da TRI e da TCT para se verificar os valores das notas/escores, destaca os pontos em que a TRI tem mais a contribuir do que a TCT para o processo avaliativo.

Neste contexto, as aplicações da TRI em avaliação multidisciplinar nas IESs que utilizam o exame mesmo quando não aplicam a TRI, representa um cenário de crescimento para as instituições pela alternativa de adaptar um outro modelo na avaliação. A Lei nº10.861, 14 de abril de 2004, possibilitou a implantação de um Sistema de Avaliação do Ensino Superior, o SINAES. Por conseguinte, foi necessário a verificação de profissionais competentes e capacitados para o desenvolvimento de novas tecnologias e para a atuação no mercado global - já que sem um modelo e uma metodologia adequados e eficaz isso seria mais difícil. Assim, com a TRI, espera-se que esta pesquisa possibilite uma reflexão e leve os docentes de disciplinas correlatas a terem subsídios para direcionar o esforço empreendido no processo de ensino e aprendizagem, de forma a contemplar a melhor abordagem pedagógica e o mais pertinente método didático a cada disciplina.

A todo momento, em situações e de formas variadas, os conhecimentos dos indivíduos acerca de determinados assuntos são postos à prova. Alunos são classificados em disputas por uma vaga em instituições de ensino, candidatos concorrem a oportunidades de emprego na área pública, entre outros. Mas é crucial e deve ser levado em consideração, na escolha ou aprovação de um indivíduo nestes exemplos de seleções os seguintes questionamentos: será que a competência que ele realmente adquiriu representa o conhecimento que as questões da prova aparentemente avaliam? Será que a habilidade em resolver a prova realmente avalia o conhecimento que o candidato deve ter? A TRI não é somente para avaliar estudantes, mas para aperfeiçoar os métodos de ensino e da avaliação.

Estas e outras questões são de relevância para o processo de avaliação educacional, as quais, porém, não podem ser resumidas somente em conceitos formais e que envolvam atribuições de notas obrigatórias à decisão de avanço ou retenção em determina-

das disciplinas. Trata-se de verificar um melhor entendimento dos mecanismos geradores de dificuldades na resolução das questões das provas, além de gerar a avaliação por competências, buscando o desenvolvimento educacional. Os métodos estatísticos aplicando os modelos da TRI podem contribuir neste processo de avaliação, permitindo uma análise mais consistente com a realidade.

Por meio desse processo de avaliação, que não se resume à mecânica do conceito formal, ou seja, não simplesmente atribuir notas obrigatórias à decisão de avanço ou retenção em determinadas disciplinas, mas obter um melhor entendimento dos mecanismos geradores de dificuldades na resolução das questões das provas e, a partir disso, gerar a avaliação que contempla e proporciona o desenvolvimento educacional.

Assim, a tese discorre analisando inicialmente três conjuntos de dados de exames multidisciplinares, os quais foram analisados pela TCT. Ressalta-se que o objetivo das instituições que aplicam o exame multidisciplinar é avaliar a aquisição de habilidades e competências por parte dos alunos e também implementar políticas públicas para a área educacional. Com a aplicação da TRI, os resultados deverão oferecer subsídios e contribuirão para a melhoria da qualidade do ensino. Além disso, o exame coopera com o processo de planejamento e com a auto avaliação dos envolvidos na academia.

Com os estudos, da mensuração por escores viabiliza a observação de componentes do processo de avaliação, tais como: uma análise mais aprofundada nas respostas dos avaliados; as situações-problema referentes à prova; o grau de discrepância das questões; a verificação do nível de dificuldade encontrado pelos respondentes a partir das respostas feitas nos gabaritos; e outras análises que podem ser obtidas detalhando-se as informações do exame.

Ao buscar fundamentos na literatura para aplicar a TRI, observa-se que o material disponível ainda é muito recente em se tratando de avaliação educacional. De acordo com Andrade, Tavares e Valle (2000), a teoria vem sendo utilizada desde 1995, ganhando popularidade devido às provas do ENEM para a seleção de candidatos em algumas universidades do país em 2010. Alguns autores abordaram a TRI na área de avaliação, relatando as potencialidades da teoria na validade de testes, são eles: Nojosa (2001); Francisco (2005); Vendramini e Dias (2005); Andrade e Bortololli (2007) ; Andriola (2008); Travitzki (2011); Quaresma, Dias e Sartorio (2012) e Piton-Gonçalves (2012).

A aplicação de modelos de TRI em provas multidisciplinares é ainda pouco explorada na literatura. O modelo clássico de TRI impõe a restrição de que existe apenas uma habilidade latente atuante, e que essa é a principal responsável pela resposta do candidato

a todas as questões.

O problema que motiva este trabalho é decorrente da aplicação de uma prova multidimensional, onde se supõe que existem diversas habilidades atuantes.

O modelo clássico supõe que todas as questões estão avaliando a dimensão latente “conhecimento em engenharia”. Porém, como ficam as habilidades específicas? Não são avaliadas? Será que esta suposição é razoável? Como testá-la?

São perguntas coerentes com a realidade e que o modelo clássico de TRI não é capaz de responder. Além disso, quando se admite que existem várias habilidades sob avaliação, como identificar qual delas o candidato está usando para responder a cada questão? O conhecimento em cálculo diferencial e integral deve, sem sombra de dúvida ajudar o aluno a responder questões de estatística, ou mesmo de álgebra linear, e o contrário também pode ocorrer. Como saber se as questões de cálculo realmente avaliam o conhecimento em cálculo e como esta habilidade se relaciona com as demais, seja estatística, álgebra ou qualquer outra?

Estas questões são relevantes para o processo de avaliação educacional e devem ser levadas em consideração para que os modelos de TRI possam ser mais realísticos. Neste sentido, os objetivos desta tese são discutido na seção 1.2 .

1.2 OBJETIVO

O objetivo da tese foi motivado por estudos preliminares inicialmente gerados com base na aplicações de três conjuntos de dados e os resultados de provas multidisciplinares, que estão detalhados no Capítulo 2.

Ao longo do trabalho procurou-se investigar aspectos preliminares da pesquisa. O objetivo geral é identificar o modelo de Teoria de Resposta ao Item que melhor se ajuste a uma avaliação multidisciplinar, interna e externa, ou seja, conforme a avaliação que foi proposta e respondida, como aborda-lá na dimensionalidade. No entanto, faz-se necessário verificar os procedimentos pela TRI, que devem ser adotados, para que, avaliação dependa dos itens elaborados e de como eles foram interpretados pelo aluno.

Assim, para determinar a dimensionalidade, considera-se que uma prova seja composta e conseqüentemente respondida por múltiplas habilidades latentes. O modelo deve ser realista à prova multidisciplinar e oferecer benefícios/vantagens para que a TRI utilize os modelos multidimensional ou unidimensional.

Como contribuições para as áreas de educação e estatística, tem-se como objetivos específicos:

- verificar a dimensionalidade pela análise fatorial e componentes principais;
- avaliar os resultados da estimação dos parâmetros com modelos aninhados, com base nos critérios de diagnóstico estatístico;
- avaliar os modelos unidimensional/multidimensional considerando “prioris” estatística para estimar o parâmetro acerto ao acaso;
- analisar as cargas fatoriais nos modelo unidimensional/multidimensional na prova multidisciplinar, com abordagem das perdas e ganhos em relação ao modelo;
- propor um escore padronizado final e resumidor da avaliação, que seja comparável com a nota entre 0 a 1.

O procedimento adotado é aplicado a três exemplos de exames multidisciplinares de grande e pequena escala. O estudo oferece subsídios e perspectivas para provas que não utilizam a TRI, mas que buscam o desenvolvimento educacional para qualidade de ensino.

1.3 ESTRUTURA DA TESE

A tese esta estruturada em cinco capítulos. O presente capítulo, é introdutório e aborda o tema de estudo, a justificativa, os objetivos e a estrutura.

O segundo capítulo traz como tema a avaliação educacional e discute os aspectos da avaliação. Descreve-se o cenário de três conjuntos de dados diferentes de avaliações em pequena e larga escala, com suas aplicações e especificidades.

O terceiro capítulo apresenta a descrição da metodologia e busca mapear e abordar os modelos empregados nas avaliações utilizando a TCT e a TRI, com base em uma ampla revisão da literatura.

No quarto capítulo, o tema principal desta tese está na diversidade de soluções e apresentam-se os resultados dos três conjuntos de dados. Discute-se acerca do emprego da TRI buscar, explicar e compreender a partir dos resultados obtidos; rever conceitos; redefinir estratégias; adequar necessidades ao meio computacional; impulsionar a melhoria da qualidade, constatando conhecimentos e analisar resultados esperados.

Por fim, o quinto capítulo tem como objetivo principal resumir as conclusões gerais da pesquisa e tecer comentários bem como sugestões para futuros estudos.

2 AVALIAÇÃO EDUCACIONAL

A avaliação é parte integrante do processo de ensino-aprendizagem e ganhou um amplo espaço com as mudanças ocorridas no mundo atual e as crescentes exigências na educação. A globalização tem levado as organizações a conceder um espaço maior às estruturas e aos processos de exames, assim, criam-se mecanismos de avaliação que requerem preparo educacional e grande capacidade de observação dos profissionais envolvidos.

Segundo Klein (2003), a avaliação educacional é um sistema de informações que tem como objetivo fornecer diagnósticos e subsídios para a implementação ou manutenção de políticas educacionais.

As aplicações da avaliação educacional, conforme Piton-Gonçalves (2012), Depresbiteris (2001, p. 144) e Penin e Martínez (2009, p. 23-24) passaram a ser identificadas a partir de duas dimensões de aplicação: interna e externa.

- A avaliação interna da aprendizagem, realizada pelo professor, é elaborada por especialistas *in loco*, durante um processo de ensino-aprendizagem para um determinado grupo de pessoas. Ela possui componentes objetivos (currículo) e subjetivos (conhecimento do professor), e é determinada de acordo com o planejamento escolar, planos de ensino, planos de aula, currículo do curso ou disciplina, etc. Geralmente, o propósito desta avaliação é ser Diagnóstica e/ou Formativa, com o objetivo de detectar as dificuldades de aprendizagem do conteúdo pelos alunos e, com isso, permitir a reflexão das estratégias de ensino do professor, as mudanças no currículo e na gestão escolar, a adoção ou o desenvolvimento de materiais didáticos, entre outras.
- A avaliação externa, de natureza sistêmica, realizada por agente externo à escola, é um importante instrumento significativo que fornece elementos para a formulação e o monitoramento de políticas públicas em escala nacional e internacional. Além disso, é relevante para a gestão da educação em nível de sistemas nas escolas estaduais e municipais. É elaborada por uma instituição pública ou privada e aplicada em um contexto de Avaliação Somativa. No Brasil, o INEP é uma autarquia federal

vinculada ao MEC que promove estudos, pesquisas e avaliações externas periódicas do sistema educacional brasileiro, com o objetivo subsidiar a formulação e implementação de políticas públicas para a área educacional. Geralmente, o propósito de uma avaliação externa é selecionar e certificar o examinado ou avaliar o sistema educacional, dentre os quais estão alunos, professores, diretores e coordenadores. São exemplos de avaliações externas os exames de certificação, os vestibulares, os concursos públicos, as provas de desempenho e os processos seletivos.

Tanto a avaliação interna como a externa precisam estar na pauta das discussões das escolas, pois objetivam diagnosticar as possíveis insuficiências na aprendizagem dos estudantes, assim como direcionar as ações pedagógicas, administrativas e financeiras com o intuito de melhorar a qualidade do ensino no País. Esse é o grande desafio que se tem, para o qual necessita-se encontrar soluções que aprimorem a qualidade da educação, para o qual necessita-se encontrar soluções que aprimorem a qualidade da educação com o objetivo de resultados efetivos de melhoria na aprendizagem dos alunos. Portanto, os indicadores sociais e educacionais mostram que se tem um longo caminho a trilhar, a construir, para garantir a aprendizagem.

Para o desenvolvimento da tese, escolheram-se três tipos de avaliação: duas internas de pequena escala e uma externa, de larga escala, que é elaborada por profissionais de fora do cotidiano escolar.

Na avaliação interna, como o elaborador é o professor, não foi encontrado e nem disponibilizado um exemplo de exame de larga escala, pois um professor não administra a mesma disciplina para todas as turmas da graduação do mesmo período, turno e curso. Devido a esses fatores, o número de alunos fica restrito pela instituição, pelo curso e período, logo, é inevitável que o conjunto de dados seja de pequena escala. Detalhes do conjunto de dados desta pesquisa são contempladas nas seções 2.1 pela avaliação interna e 2.2 quando trata da avaliação externa.

2.1 DADOS DE AVALIAÇÕES INTERNAS

Nas subseções 2.1.1 e 2.1.2, descrevem-se dois conjuntos de dados que são exemplos da avaliação multidisciplinar interna da mesma instituição. No contexto institucional, a utilização da avaliação interna ocorre porque o objetivo da instituição Pontifícia Universidade Católica do Paraná (PUCPR) é avaliar o desenvolvimento de competências com base na aplicação de um exame multidisciplinar.

As questões que compõem o exame são elaboradas pelos docentes que atuam em cada curso e validadas pela sua coordenação e por seu Núcleo Docente Estruturante. As questões são então encaminhadas à Diretoria de Graduação e Avaliação Institucional, na Pró-Reitoria Acadêmica, onde é feita uma conferência na editoração das questões e, em seguida, a composição das provas. Esta mesma diretoria se ocupa da reprodução das provas e da distribuição delas para aplicação em cada Escola no *Campus* Curitiba e em cada um dos *Campi* fora de sede: São José dos Pinhais, Toledo, Londrina e Maringá. Após a aplicação, os gabaritos retornam à Diretoria de Graduação e Avaliação Institucional para correção e posterior divulgação dos resultados e emissão de relatórios. Cada coordenação de curso utiliza-se dos relatórios para organizar a atividade devolutiva aos alunos, abordando especialmente aquelas questões cujo índice de acerto tenha sido muito baixo, e reflete sobre a elaboração daquelas que tiveram índice de acerto muito alto. Estas provas tem o objetivo de preparar os alunos que realizam o Enade.

A correção é feita pela contagem de acertos de questões objetivas, ou seja, não se utilizam outros modelos como o da TRI. Com a aplicação da nova proposta, os resultados devem oferecer subsídios contribuindo na melhoria constante do processo de verificação da aprendizagem e do desenvolvimento de competências.

2.1.1 Conjunto de dados I

No primeiro estudo de caso foi selecionada uma amostra da Escola Politécnica da PUCPR, de um único curso de graduação o da Engenharia Química, com trezentos e quarenta e três (343) alunos que realizaram a avaliação multidisciplinar. Entre eles, foram escolhidos somente os alunos do segundo período que realizaram a mesma prova referente aos conteúdos desenvolvidos de acordo com a estrutura curricular, totalizando noventa e quatro (94) discentes da graduação, determinada e especificada no ANEXO A.

O conjunto de dados I (CD1) foi processado pela instituição. O nome e a prova dos alunos ficaram em anonimato por pertencerem aos professores e ao banco de dados da Escola, respeitando a ética profissional. Inicialmente, os resultados do conjunto de dados do exame multidisciplinar foram descritivamente analisados, contemplando uma única “habilidade - a geral”, que visa à formação do acadêmico, composta por 30 itens de múltipla escolha, sobre conceitos dos conteúdos desenvolvidos e apreendidos no decorrer do semestre, apresentada na forma de questões elaboradas pelos docentes que ministraram as disciplinas.

Neste conjunto de dados, o exame multidisciplinar é composto por seis blocos de

disciplinas e, sem dúvida, todas fazem parte de uma habilidade geral, que é a formação do profissional, porém, cada bloco tem o objetivo específico de avaliar a habilidade do aluno. Essas análises estão no segundo momento desse estudo, na seção 4.1, em que se faz uma complementação das análises descritivas da avaliação multidisciplinar de acordo com o método da TRI e dividindo a prova em seis habilidades, cujas disciplinas, apresentadas na Tabela 1, abordaram cinco (5) itens.

Tabela 1: Disciplinas avaliadas e suas siglas

Disciplina	Sigla
Geometria Analítica e Álgebra Linear	GAA
Cálculo Diferencial e Integral	CDI
Física Geral e Experimental	FGE
Química Geral e Inorgânica	QGI
Estatística	EST
Introdução à Engenharia Química	IEQ

Averiguam-se as análises conduzidas de duas formas:

- a primeira discorre a prova como um todo, supostamente medindo uma habilidade e todas as questões foram avaliadas conjuntamente;
- a segunda considerou que cada disciplina é uma habilidade latente e sua prova foi analisada separada das demais.

O objetivo é verificar se as abordagens apresentam diferenças relevantes e qual é a mais adequada para a presente situação.

Como em toda análise estatística de dados, é interessante começar com análises descritivas simples, a fim de tomar afinidade com os dados. Os resultados apresentados neste Capítulo foram processados utilizando o *software R* (R Development Core Team, 2012), através do pacote Latent Trait Model (*ltm*) (Rizopoulos, 2006) modelos de variáveis latentes para dados dicotômicos. A Tabela 2 apresenta a proporção de acertos para cada uma das 30 questões que compõem o exame multidisciplinar para o curso e período em estudo. Para melhor compreensão, numeraram-se as questões dentro das habilidades de 1 a 5, e na prova, de 1 a 30. Esta codificação é utilizada neste conjunto para abordar com detalhes as análises.

De forma geral, o que chama atenção na Tabela 2 é que em todas as questões a proporção de acertos está abaixo de 50%. A disciplina de GAA apresenta duas das questões com mais baixa proporção de acertos: as questões Q12(10%) e Q15(13%). Por outro lado, a disciplina GAA também apresenta a questão com maior proporção de acertos

Tabela 2: Proporção de acertos por questões

Habilidade Geral	H.Q	acerto (%)	Habilidade Geral	H.Q	acerto (%)
	QGI			CDI	
Q1	Q 1	44 %	Q16	Q 1	30%
Q2	Q 2	20%	Q17	Q 2	27%
Q3	Q 3	18%	Q18	Q 3	37%
Q4	Q 4	17%	Q19	Q 4	40%
Q5	Q 5	30%	Q20	Q 5	24%
	IEQ			FGE	
Q6	Q1	18%	Q21	Q 1	15%
Q7	Q 2	39%	Q22	Q 2	23%
Q8	Q 3	24%	Q23	Q 3	39%
Q9	Q 4	49%	Q24	Q 4	31%
Q10	Q 5	39%	Q25	Q 5	15%
	GAA			EST	
Q11	Q1	47%	Q26	Q 1	27%
Q12	Q 2	10%	Q27	Q 2	37%
Q13	Q 3	16%	Q28	Q 3	26%
Q14	Q 4	27%	Q29	Q 4	35%
Q15	Q 5	13%	Q30	Q 5	38%

Nota: Habilidade por questão (HQ) e Questão (Q).

Q11(47%). Essa disciplina é a que apresenta maior distanciamento entre a proporção de acertos das questões. Nas demais disciplinas, os acertos foram mais homogêneos.

A Figura 1 e Tabela 3, mostram a porcentagem de acertos das questões. A figura apresenta o resultado referente às seis habilidades. Já a tabela envolve a habilidade geral.

Na Figura 1, observa-se que somente na disciplina QGI obteve-se quatro respondentes que acertaram as cinco questões, seguida pelas disciplinas IEQ e EST, com dois e um respondentes, respectivamente. Nota-se, também, que nas habilidades QGI e GAA, 34 alunos não acertaram nenhuma questão, seguidas pelas disciplinas FGE (31), CDI (27), EST (25) e IEQ (22).

Os respondentes do exame multidisciplinar nas disciplinas de IEQ, GAA e FGE obtiveram em torno de 50% de acertos, correspondendo entre uma e duas questões. Na habilidade de CDI, 13 alunos dos 94 respondentes acertaram quatro itens, e aproximadamente 75% dos respondentes acertaram duas, uma ou nenhuma das trinta questões.

É importante analisar o que aconteceria sob a suposição de que todos os alunos respondessem ao acaso às questões de uma prova de múltipla escolha. Para verificar qual seria esse comportamento, a Tabela 3 apresenta a frequência de acertos observados e o que é esperado sob a hipótese de que os alunos estão respondendo aleatoriamente a todas as questões.

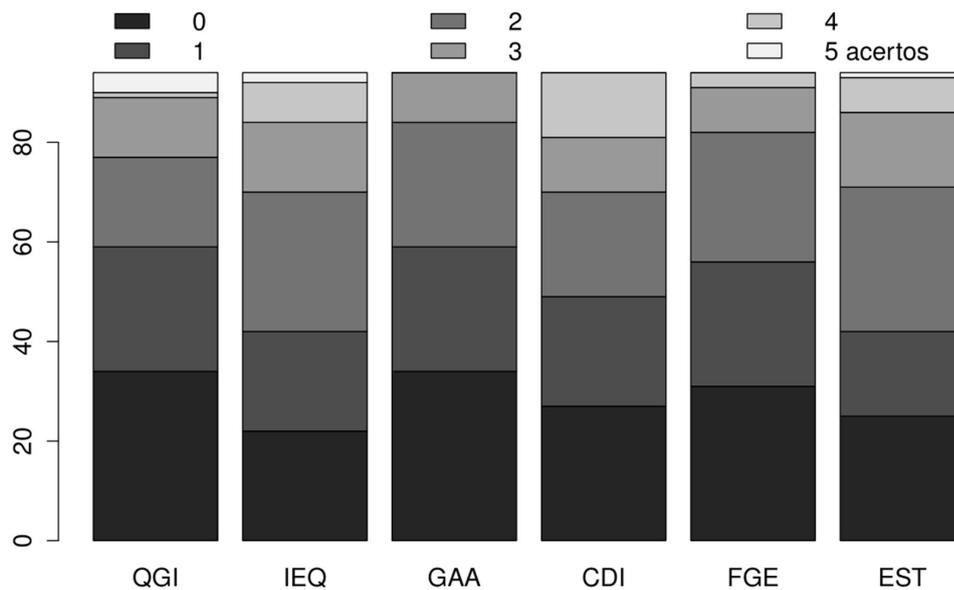


Figura 1: Porcentagem dos acertos das questões com as seis habilidades

Tabela 3: Frequência de acertos do valor observado e o número esperado de acertos sob a hipótese de respostas aleatórias

Acertos	Frequência	N. Esperado casual	Acertos	Frequência	N. Esperado casual
0	19	0,12	11	4	1,5
1	0	0,9	12	7	0,6
2	0	3,1	13	7	0,2
3	0	7,3	14	7	0,06
4	2	12,4	15	3	0,02
5	3	16,2	16	6	0
6	3	16,9	17	1	0
7	8	14,5	18	2	0
8	9	10,4	19	0	0
9	6	6,3
10	7	3,3	30	0	0

Nota: Entre 19 a 30 a frequências de acertos e o número esperado de acerto casual (NAC) foi zero.

Na Tabela 3, nota-se que a maior frequência de acertos concentrou-se entre 7 e 14 questões. Dos 94 alunos que fizeram a prova, esperava-se que ninguém zerasse, sob a suposição de acerto casual. Porém, 19 alunos resolveram a prova e não tiveram nenhum acerto, este é um resultado bastante relevante.

A segunda é que eles não objetivaram responder à prova de forma aleatória, mas, pelo contrário, tentaram respondê-la corretamente, porém, devido a sua baixa habilidade, erraram. De uma forma ou de outra, este é um resultado que indica que melhorias são necessárias, seja nas questões, seja no ensino ou mesmo na conscientização dos alunos

quanto a responder à prova de forma responsável.

Observa-se também que entre 4 a 8 acertos, o valor observado foi sempre menor do que o esperado em relação a acertos casuais. Nota-se que ninguém acertou da primeira à terceira (1 a 3) questão; em seguida, 75 alunos acertaram entre 4 a 18 questões e, dentre a décima nona e trigésima (19 a 30) questão, não houve acerto, ou seja, ninguém gabaritou a prova, conforme a Figura 2.

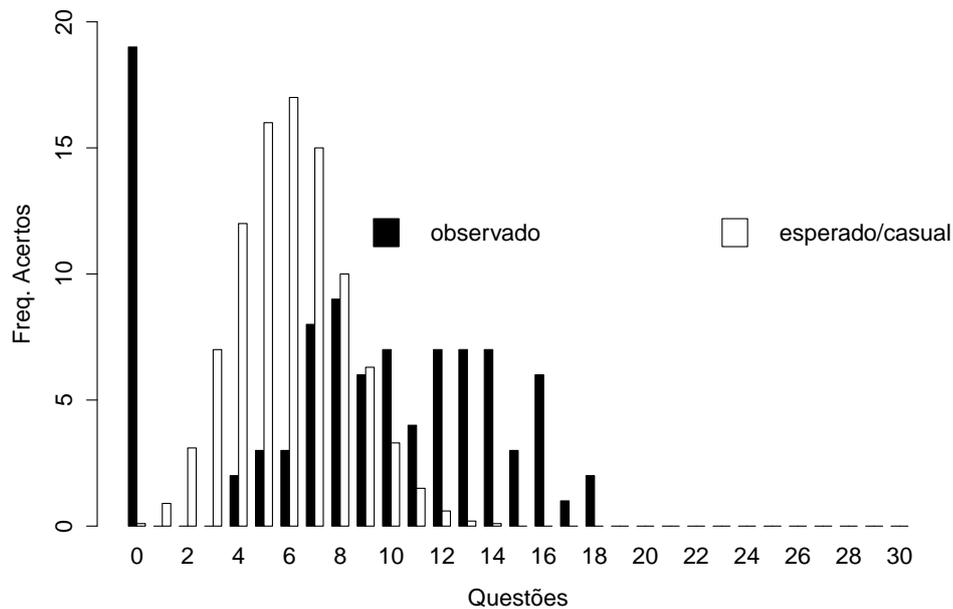


Figura 2: Frequência de acertos do valor observado da habilidade total e o número esperado de acerto, casual

Do ponto de vista da análise estatística, o conjunto de dados é desafiador para a aplicação de modelos da TRI. A base é pequena - apenas 94 respondentes - e ainda apresenta uma quantidade considerável de respondentes sem qualquer acerto.

A Figura 2, que representa os dados apresentados na Tabela 3, é de visualização imediata da frequência de acertos e do que era de esperar de acerto sob a hipótese de resposta aleatória. Há dois fatores visuais do gráfico, sendo o primeiro a quantidade de zeros no observado e em segundo o deslocamento dos valores observados com os do esperado de acerto casual, ou seja, o valor observado é do acerto casual.

A frequência contabilizada são de 2 respondentes que acertaram 18 questões, contra 19 respondentes que zeraram a prova e indica o oposto do esperado de acerto casual, conforme mostra a Figura 2.

Para verificar a quantidade de acertos por disciplina em contraste com o total

de acertos na avaliação, a Figura 3 tem como finalidade mostrar quantos respondentes também zeraram por disciplinas, mas que acertaram outras questões na prova.

Na Figura 3 a comparação é referente ao somatório da quantidade de acertos das questões, considerando as 6 disciplinas/habilidades com as 5 questões na razão proporcional da habilidade geral das 30 questões.

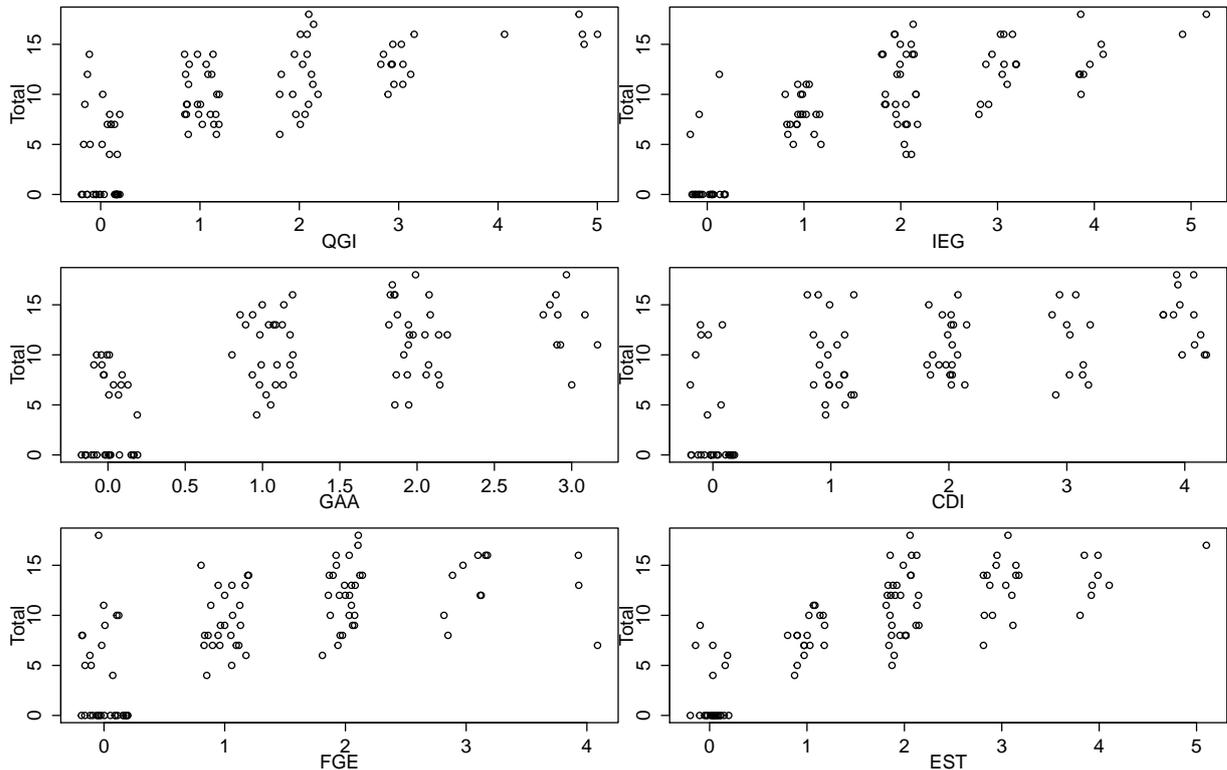


Figura 3: Somatório dos acertos das questões das seis habilidades com a habilidade geral

Observa-se, na Figura 3, que na habilidade FGE houve um aluno entre os que zeraram que acertou mais que a metade da prova considerando a habilidade geral. Na habilidade geral, considerando as 30 questões, o número maior de acertos não ultrapassou de 20 questões.

Para as disciplinas de CDI e FGE, não houve acerto nas 5 questões; para GAA, o número reduz para três questões certas. O indivíduo que respondeu a 3 questões certas das 5 em GAA, acertou em torno de 6 questões, na prova geral de 30 questões, ou seja, comparando com o total, obteve metade de sua habilidade na disciplina. Para QGI, IEQ e EST, existem alunos que gabaritaram a prova ao acertarem 5 questões, mas, em contrapartida, acertaram mais que a metade ou a metade da prova para habilidade geral.

A pesquisa retoma a estatística exploratória para abordar dois pareceres, a cor-

relação ponto bisserial e o *Cronbach's alpha*, para efeito de comparação da amostra e detalhamento das verificações proposta pelas bibliografias de análise da TRI. Conforme a Tabela 4, primeiro considerou-se uma habilidade geral e em segundo, a avaliação dividida em seis habilidades de estudos, definidas em: GAA; CDI; FGE; QGI; EST e IEQ. Cada disciplina abordou cinco (5) itens. Para a abordagem inicial do conjunto de dados foi utilizado o *ltm* (Rizopoulos, 2006).

Logo, para que se tenha uma compreensão sucinta, seguem as definições de ambos:

Correlação Ponto Bisserial : mede a capacidade de discriminação do item em relação ao resultado do teste, uma variável apresenta correlação significativa com o escore bruto produzido pelo conjunto das variáveis. A Correlação Ponto Bisserial fornece uma medida da relação entre uma variável contínua, como escores de testes, e outra variável com duas categorias ou dicotômicas, como aprovado ou reprovado. Esse passo é crucial para a escolha de itens que de fato apresentam consistência interna e se associam bem ao escore que é produzido. Segundo Souza e Tavares (2005), normalmente, aceitam-se valores para a correlação superior a 0.3. A correlação ponto bisserial pode ser derivada diretamente da correlação de Pearson e quanto maior o coeficiente de correlação, maior é a discriminação do item. É dado pela expressão:

$$r = \frac{(\bar{X}_p - \bar{X}_t)}{s_t} \cdot \sqrt{\frac{(\hat{\theta})}{(1 - \hat{\theta})}} \quad (2.1)$$

onde:

\bar{X}_p é a média dos escores dos examinados que responderam ao item corretamente;

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

s_t é o desvio padrão do teste;

$\hat{\theta}$ é a proporção de examinados que respondem ao item corretamente.

Na bibliografia, por Lee Cronbach (1951), *Cronbach's alpha* (AC), é uma forma de estimar a confiabilidade de um questionário aplicado em uma pesquisa. Ele mede a correlação entre respostas em um questionário através da análise das respostas dadas, apresentando uma correlação média entre as perguntas.

O coeficiente alfa é calculado a partir da variância dos itens individuais e da variância da soma dos itens de cada avaliador de todos os itens de um questionário, que utilizem a mesma escala de medição, cuja a expressão é:

$$\alpha = \left(\frac{I}{I-1} \right) \cdot \left(1 - \frac{\sum_{i=1}^I S_i^2}{S_t^2} \right) \quad (2.2)$$

onde:

I é o número de itens;

S_i^2 é a variância do i -ésimo item;

S_t^2 é a variância do total do questionário.

Para que sejam verificadas a consistência interna do instrumento de avaliação e a correlação entre os itens avaliados, na Tabela 4 estão apresentadas o *Cronbach's alpha* e a correlação ponto bisserial, para cada item, levando em consideração todas as questões e as habilidades específicas.

O coeficiente *Cronbach's alpha* é uma medida da consistência interna do instrumento de avaliação. Seu valor varia de 0 a 1, sendo que valores próximos de 1 indicam alta consistências e foi avaliado para cada habilidade com todas as questões e retirando-se uma a uma. Por exemplo, para a habilidade QGI, o AC apresentou o valor de 0,60, quando avaliado com todas as questões desta habilidade. Por outro lado, quando foi retirada a questão Q3 apresentou o valor de 0,57.

Quando trata-se as habilidade de forma independente, verifica-se, de forma geral, que a consistência das provas é baixa, variando de 0,60 para QGI até 0,31 para GAA. Destaca-se ainda na Tabela 4 as questões Q1 de GAA e Q4 de GAA, como as que têm maior influência no AC.

Com relação à Correlação Ponto Bisserial (CPB), ela é análoga ao coeficiente de correlação de Pearson, porém, adequada entre variáveis categóricas. Essa correlação é avaliada dentro do contexto de TRI para verificar a concordância da questão com a prova como um todo. Considere o escore em um teste como o total de acertos do indivíduo. A CPB mede a correlação entre a resposta em um determinado item com o escore total da prova. A ideia é de que se a questão apresenta boa aderência ao instrumento de medida, ela deve apresentar uma CPB acima de 0,3. O escore total pode ser calculado considerando o item ao qual se está testando, ou sem o item, o que for de maior interesse. Por isso, a Tabela 4 apresenta as colunas (incluir) e (excluir).

De forma geral, verifica-se que quando se avalia toda a prova, as duas medidas

Tabela 4: Correlação ponto bisserial e *Cronbach's alpha*

Habilidades	CPB		Cronbach's alpha	Habilidade	CPB		Cronbach's alpha
Questão (Q)	com Total	excluir	*Todas as questões e Excluindo a questão	Geral	com Total	excluir	*Todas as questões e Excluindo a questão
QGI			0,60*				0,82*
Q 1	0,61	0,29	0,59	Q1	0,59	0,53	0,81
Q 2	0,61	0,36	0,55	Q2	0,32	0,25	0,82
Q 3	0,56	0,31	0,57	Q3	0,42	0,36	0,82
Q 4	0,65	0,43	0,52	Q4	0,44	0,38	0,82
Q 5	0,68	0,42	0,52	Q5	0,47	0,40	0,82
IEQ			0,49*				
Q 1	0,40	0,12	0,51	Q6	0,20	0,13	0,83
Q 2	0,57	0,25	0,45	Q7	0,41	0,33	0,82
Q 3	0,59	0,32	0,40	Q8	0,41	0,34	0,82
Q 4	0,64	0,32	0,39	Q9	0,62	0,56	0,81
Q 5	0,62	0,32	0,40	Q10	0,52	0,45	0,82
GAA			0,31*				
Q1	0,68	0,26	0,14	Q11	0,59	0,53	0,81
Q 2	0,24	-0,04	0,39	Q12	0,23	0,18	0,82
Q 3	0,40	0,05	0,35	Q13	0,20	0,14	0,83
Q 4	0,66	0,29	0,12	Q14	0,36	0,29	0,82
Q 5	0,48	0,18	0,25	Q15	0,24	0,18	0,82
CDI			0,54*				
Q 1	0,53	0,23	0,53	Q16	0,23	0,15	0,83
Q 2	0,62	0,35	0,46	Q17	0,39	0,31	0,82
Q 3	0,61	0,32	0,48	Q18	0,41	0,33	0,82
Q 4	0,61	0,30	0,49	Q19	0,54	0,47	0,81
Q 5	0,58	0,32	0,48	Q20	0,36	0,28	0,82
FGE			0,34*				
Q 1	0,53	0,24	0,24	Q21	0,21	0,15	0,83
Q 2	0,54	0,18	0,28	Q22	0,30	0,23	0,82
Q 3	0,52	0,09	0,38	Q23	0,48	0,40	0,82
Q 4	0,58	0,20	0,26	Q24	0,36	0,29	0,82
Q 5	0,45	0,14	0,32	Q25	0,35	0,29	0,82
EST			0,43*				
Q 1	0,58	0,28	0,34	Q26	0,41	0,34	0,82
Q 2	0,58	0,24	0,36	Q27	0,54	0,47	0,81
Q 3	0,41	0,08	0,47	Q28	0,33	0,25	0,82
Q 4	0,56	0,22	0,38	Q29	0,46	0,38	0,82
Q 5	0,62	0,29	0,32	Q30	0,46	0,38	0,82

Nota: No *alfa de Cronbach*, que verifica a confiabilidade dos itens, a notação (*) indica questões que são consideradas e sem o asterisco questão excluída; Correlação ponto bisserial (CPB).

descritivas tendem a indicar melhores resultados, em relação a quando avalia-se a prova por disciplinas. A disciplina de GAA foi a que chamou mais atenção, pois apresenta

diversas questões com baixa aderência e pouca consistência interna.

Depois de aferir as análises e verificar os resultados, necessitou-se segmentar os estudos que são vistos no Capítulo 4. O conjunto com os 94 respondentes, e a partir desse, o conjunto de dados com amostras dos alunos que não zeraram a prova com 75 (setenta e cinco) respondentes.

A partir dos resultados observados na Tabela 3 e 4, é analisada, a pesquisa, em 04 (quatro) etapas, sendo que a primeira e a segunda são referentes aos 94 respondentes com seis habilidades e a habilidade geral, enquanto a terceira e quarta abordagens de análise estão relacionadas a 75 respondentes para as seis habilidades e a habilidade geral. Vale ressaltar que não era esperado um conjunto de dados inflacionados em zero. O excesso de zeros é sob a suposição de chute ou da baixa habilidade dos alunos ou ainda, de outro fator externo. Por esse motivo, a pesquisa aborda esse estudo e também verifica as influências nos modelos, todavia, não utiliza somente análise exploratória pela TCT.

O interesse principal reside, principalmente, nos parâmetros das questões, nas habilidades latentes dos alunos e no modelo da TRI que venha a representar as informações da realidade da avaliação. Portanto, abordadas as análises exploratórias, os resultados foram verificados e diferentes modelos da TRI foram utilizados, conforme segue no Capítulo 4, seção 4.1 dos modelos logísticos.

O conjunto de dados em estudo gerou questionamentos para desenvolver e dar continuação à tese, os quais estão abaixo listados:

- a quantidade de indivíduos que responderam ao exame influencia nas análises utilizando a TRI?
- o exame multidisciplinar possuiu quantas competências ao ser respondido pelos alunos?
- a quantidade de itens influencia nas análises utilizando a TRI?
- a partir do que foi avaliado na aplicação do exame multidisciplinar, é possível fazer uma análise utilizando a TRI unidimensional ou multidimensional?
- a quantidade de alunos que zeraram a prova, inflacionando em zero, influência no resultado?
- pela quantidade de zeros e os valores esperados e observados, como ficou o acerto ao acaso (o popular “chute”)?
- quais são os resultados diferentes que a TRI pode proporcionar ao ser aplicada nestes dados?

Quando se analisou com mais detalhes os resultados, surgiram tantos questio-

namentos que ficou inviável abordar somente essas análises. Tornou-se imprescindível buscar outros resultados para detalhar esse exame, inclusive, compará-lo a outro exame multidisciplinar. Essa abordagem, no entanto, trata-se somente de uma avaliação, mas multidisciplinar, em busca de resultados da qualidade do ensino e das competências adquiridas de apenas um período do curso. Com esse cenário, sente-se a necessidade de que sejam norteados e classificados alguns passos sobre como abordar e analisar uma avaliação multidisciplinar, em busca do desenvolvimento educacional com tomadas de decisão baseadas em resultados estatísticos.

2.1.2 Conjunto de dados II

O conjunto de dados II tem por objetivo dar continuidade à pesquisa, visto que o conjunto I gerou questionamentos, conforme citados na seção 2.1.1; bem como assegurar a consistência das análises e para isto, leva-se em conta a pertinência das indagações em outros exames multidisciplinares, mesmo em outra escola, com outras características distintas. O conjunto de dados II(CD2), da Escola de Negócios da PUCPR, refere-se também ao segundo semestre de 2011. Para essa análise, a escolha foi pelo curso de Administração, por ter mais respondentes e itens na prova. Apesar de CD2 ser maior que o CD1, ainda se trabalha em pequena escala, mas em busca de respostas na utilização da TRI em avaliações multidisciplinares internas.

O conjunto de dados abordado nesta seção possui 189 alunos do segundo período, que realizaram o exame multidisciplinar no segundo semestre. A prova contém 40 questões. Para maiores detalhes do CD2 consulte o ANEXO B. As disciplinas cobradas fazem parte da grade curricular deste período, a seguir:

- Métodos de Pesquisa em Administração;
- Ética;
- Teorias da Administração;
- Raciocínio Analítico;
- Fundamentos de Marketing;
- Matemática Aplicada a Negócios.

Para esse conjunto de dados, não são abordadas as análises como no CD1, porque das 6 disciplinas, nem todas possuíam a mesma quantidade de questões entre elas e ainda, as questões específicas de Administração estavam interligadas na própria área, fato que dificultou a separação por disciplinas, exceto as questões que envolviam cálculos da área de

Exatas. Portanto, é tratada como uma única análise, utilizando os mesmos procedimentos computacionais do pacote *ltm*.

Conforme denota a Tabela 5, os resultados são compostos pelas estatísticas descritivas do conjunto de dados II, da proporção de acertos por questões, do ponto bisserial e *Cronbach's alpha*.

Tabela 5: Resultados descritivos do exame multidisciplinar do conjunto de dados II

189 respondentes									
Q	P.A%	Ponto Bisserial com Total		Cronbach's alpha	Q	P.A%	Ponto Bisserial com Total		Cronbach's alpha
		incluir	excluir	*Todas as questões e Excluindo a questão			incluir	excluir	*Todas as questões e Excluindo a questão
				0,9012*					
Q1	51%	0,53	0,48	0,898	Q21	34%	0,43	0,38	0,899
Q2	35%	0,3	0,24	0,9014	Q22	55%	0,6	0,56	0,865
Q3	26%	0,36	0,32	0,9	Q23	17%	0,3	0,26	0,9
Q4	52%	0,59	0,55	0,897	Q24	57%	0,55	0,51	0,897
Q5	67%	0,71	0,68	0,895	Q25	7%	0,23	0,2	0,9011
Q6	58%	0,67	0,64	0,895	Q26	13%	0,15	0,11	0,902
Q7	55%	0,55	0,5	0,9	Q27	40%	0,45	0,4	0,899
Q8	22%	0,31	0,26	0,895	Q28	48%	0,55	0,5	0,897
Q9	66%	0,68	0,65	0,895	Q29	26%	0,4	0,35	0,899
Q10	60%	0,59	0,55	0,897	Q30	14%	0,24	0,2	0,9013
Q11	40%	0,43	0,38	0,899	Q31	57%	0,63	0,59	0,896
Q12	34%	0,45	0,4	0,899	Q32	42%	0,51	0,46	0,898
Q13	42%	0,35	0,29	0,901	Q33	3%	0,01	-0,01	0,902
Q14	32%	0,44	0,39	0,899	Q34	67%	0,62	0,59	0,896
Q15	20%	0,36	0,32	0,9	Q35	11%	0,16	0,12	0,902
Q16	45%	0,42	0,37	0,899	Q36	11%	0,21	0,18	0,9014
Q17	50%	0,6	0,56	0,896	Q37	9%	0,23	0,19	0,9012
Q18	55%	0,64	0,6	0,896	Q38	18%	0,31	0,26	0,9
Q19	62%	0,68	0,65	0,895	Q39	7%	0,17	0,14	0,9015
Q20	48%	0,47	0,42	0,899	Q40	21%	0,34	0,3	0,9

Nota: Questão (Q). Proporção de acerto (P.A%). No *Cronbach's alpha* verifica-se a confiabilidade dos itens com anotação (*) e significa, que todas questões são consideradas e sem o asterisco a questão é excluída. As questões em negrito são as questões que apresentaram valores inadequadas.

De forma geral, o que chama atenção na Tabela 5 é que em todas as questões a proporção de acertos varia entre 3% a 67%, ou seja, as porcentagens de acertos são heterogêneos. As questões que os respondentes mais acertaram - com índice de acerto superior a 60% - foram Q5 e Q34 (67%), Q9 (66%) e Q19 (62%), lembrando que existem duas interpretações para estas questões: ou os alunos possuíam a habilidade de resposta ou a questão era muito fácil. Por outro lado, as questões com a menor proporção de acertos - abaixo de 10% - foram Q25 e Q39 (7%), Q33 (3%) e Q37 (9%), e pode-se pensar que ou foram mal formuladas ou que eram muito difíceis. Todavia, nessa prova havia 4 questões de índice de acerto superior a 60%, e também 4 questões para um índice abaixo de 10%.

A medida de consistência interna do instrumento de avaliação, quando o valor do coeficiente *Cronbach's alpha*, representado pelo valor de 0,9012, indica alta consistência significativa, para todas as questões, visto a aproximação em 1, detalhes em Cronbach (1951). A consistência interna de todas questões é representado pelo valor em asterisco, portanto para cada questão o valor respectivo indica a consistência do item excluído e compara-se com o valor de 0,9012. Comparado esses valores e caso o valor do item seja acima, indica que, ao excluí-lo, há um aumento da consistência, então vale a pena retirar a questão analisada. O contrário seria que se o item fosse excluído e o valor estivesse abaixo então não faria diferença significativa do considerado.

A diferença significativa dessa exclusão ocorreria nas questões que estão em negrito: Q26, Q33 e Q35, mas os valores são quase imperceptíveis, pois são apenas 0,0008 um pouco acima do valor de 0,9012. Assim, houve a indicação de que esses itens necessitam de uma análise pedagógica dos fatores observados, antes de excluí-los totalmente.

Com relação à correlação ponto bisserial, adequada para verificar a correlação e a concordância das questões da prova como um todo, as questões que apresentaram baixa aderência ao instrumento de medida, ou seja uma correlação PB abaixo de 0,3 foram: Q25, Q26, Q30, Q33, Q35, Q36, Q37 e Q39, ou seja, dos 40 itens, 9 apresentaram baixa correlação, tanto incluindo a questão na prova como excluindo-a. Na Q33 a correlação de 0,014 é ao incluir a questão na prova e ao excluir, o valor reduz para $-0,01$, ou seja, essa questão apresentou nenhuma concordância com a prova.

A questão com maior índice de acerto foi Q34 com 67%. As médias de acerto das questões do CD2 estão em torno de 37%. Deve-se isto ao fato de que na análise do CD2 existem quatro questões abaixo de 10% na proporção de acertos. Pode-se dizer que os alunos do CD2 tiveram um índice melhor de acertos nas questões em relação à probabilidade de acerto ao acaso.

Na Tabela 6 e Figura 4, mostram-se a frequência de acertos das questões, para verificar qual seria o comportamento dos resultados observados no exame, sob a hipótese de que os alunos estão respondendo aleatoriamente a todas as questões.

Na Tabela 6, nota-se que a maior frequência de acertos concentrou-se entre as questões 15 e 21. Dos 189 alunos que fizeram a prova, a frequência esperada de 0 é 0,03, mas mesmo dobrando os respondentes, os que obtiveram zero acertos neste CD2, sob a suposição de acerto casual, totalizaram 35 alunos. Ainda, ninguém acertou acima de 29 questões, logo, ninguém gabaritou a prova. Este é um resultado bastante relevante, pois, além de se fazer presente na avaliação anterior do CD1, persistiu também nesta

Tabela 6: Frequência de acertos do valor observado e o número esperado de acertos casuais do conjunto de dados II

Acertos	Frequência	N. Esperado casual	Acertos	Frequência	N. Esperado casual
0	35	0,03	16	17	0,37
1	0	0,25	17	20	4,50
2	0	1,22	18	14	2,17
3	0	3,88	19	14	0,94
4	0	8,97	20	9	0,37
5	0	16,14	21	13	0,13
6	0	23,54	22	9	0,04
7	2	28,59	23	5	0,01
8	1	29,48	24	5	0
9	2	26,20	25	7	0
10	3	20,31	26	0	0
11	1	13,85	27	5	0
12	4	8,37	28	0	0
13	5	4,50	29	1	0
14	6	2,17
15	11	0,94	40	0	0

Nota: Entre 30 a 40 a frequências de acertos e o número esperado de acerto casual foi zero.

avaliação, indicando pelo menos duas situações que merecem atenção. A primeira é que nenhum aluno mostrou habilidade em todas as questões, e a segunda é que muitos alunos continuam zerando a prova, com a perspectiva de acerto, pois caso marcassem o item aleatoriamente na prova, a probabilidade de acerto casual seria acima do valor observado. Este resultado indica que melhorias são necessárias, seja nas questões, seja no ensino ou mesmo na conscientização dos alunos quanto a responder à prova de forma responsável.

Observa-se também que os acertos observados ficaram muito mais distantes para o CD2 do que o valor observado e vice-versa, ou seja, entre os acertos nas Q13 à Q25 o valor observado dos acertos na prova foi melhor do que se esperava. O mesmo foi constatado para os acertos nas Q27 e Q29, ou seja, mais do que a esperança sob acerto casual. Nota-se que ninguém acertou da primeira à sexta (1 a 6) questão.

Do ponto de vista da análise estatística, o conjunto de dados da avaliação multidisciplinar interna é bastante desafiador para a aplicação de modelos da TRI. A base é relativamente pequena, mesmo dobrando os respondentes e ainda apresenta uma quantidade considerável de respondentes sem nenhum acerto, como mostra a Figura 4, possuindo uma característica própria de inflação de zeros na prova.

Observa na Figura 4 que o valor do acerto casual “o esperado ” sob uma distribuição binomial, revela que o valor observado teve resultados melhores do que o esperado,

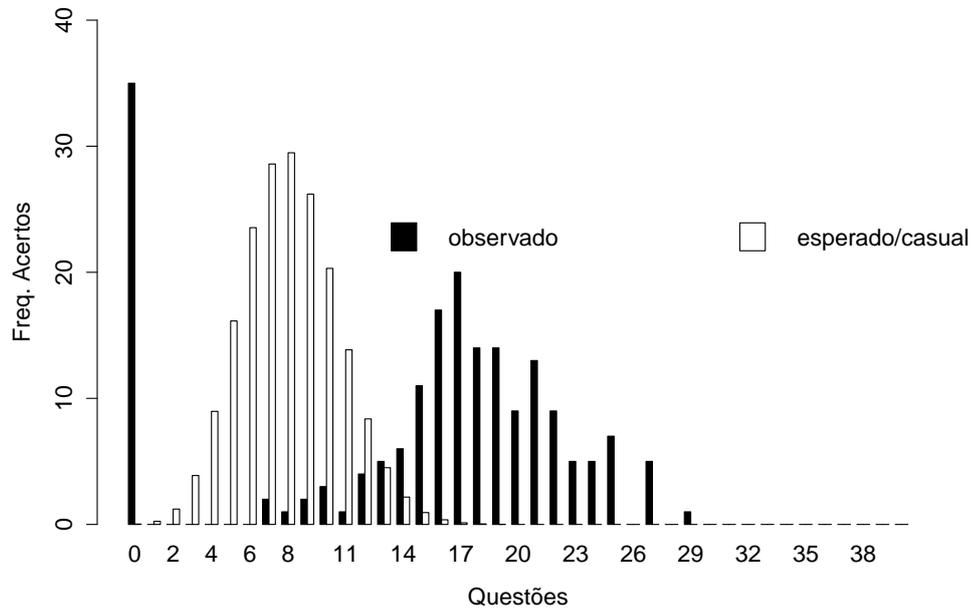


Figura 4: Frequência de acertos do valor observado da habilidade total e o número esperado de acerto casual

ou seja, neste exame multidisciplinar os alunos acertaram mais questões do que na probabilidade do acerto casual levando também em conta que aproximadamente 18% dos respondentes zeraram a prova.

A TRI é um método de análise que melhor descreve os fenômenos que é responder dado conhecimento, mas, é atraente por considerar uma gênese mais realista que permite caracterizar as questões por parâmetros.

No entanto, é abordado o exemplo da avaliação de larga escala, que não utiliza também a TRI, para tratar outra forma de exame, que a partir da qualidade de ensino, classifica os cursos de graduações nas IESs, para avançar nas análises necessárias na obtenção de respostas, abordando e abrangendo uma parcela significativa de avaliações multidisciplinares.

Apesar destas características apresentadas, os dois conjuntos de dados analisados preparam os alunos para o Enade e ambos não utilizam a TRI. O uso da TRI facilita as análises na avaliação multidisciplinar por abordar e abranger os itens na habilidade avaliada. Para isto, será verificada a coerência dos resultados e a estabilidade estatística.

2.2 DADOS DE AVALIAÇÃO EXTERNA

Frequentemente, estudos científicos em avaliações educacionais são de indivíduos, oriundos de diferentes grupos ou população, exemplo SAEB e ENEM. Estes, por sua vez, comumente, apresentam diferentes características individuais, próprias de cada um desses grupos e são avaliações, também, para seleção, qualificação e certificação. Visando à aplicação desses exames, disponibilizados pelo INEP, abordou-se apenas um conjunto real, Enade/2009. O motivo é devido à gama de possibilidades existentes de análises com diferentes grupos e por serem avaliações de larga escala, onde os dados são tratados como microdados. O grupo do curso escolhido é de Estatística, já que esta tese aborda esta área.

Outro ponto fundamental, na escolha do Enade para a análise da avaliação externa, é que no Congresso Internacional “Educação: uma Agenda Urgente” de 2011 o Conselho Assessor das Metas 2021 da Organização dos Estados Ibero-Americanos (OEI), no qual o Brasil é um dos países integrantes, e a presidente de 2011 do INEP, Malvina Tuttman, levantaram questões fundamentais que justificam os estudos do conjunto que é abordado nesta subseção, com base no Enade:

1. afirmaram que a avaliação externa é um importante elemento para a equidade;
2. conclamaram os pesquisadores a trabalharem com os dados da Educação divulgados pelo INEP, segundo Malvina Tuttman: “Temos que olhar e, efetivamente, enxergar e deixar de responsabilizar e de ficar de braços cruzados. Que País é esse que queremos? Vamos avaliar, diagnosticar. Nós já temos muitos dados. Não precisamos ficar esperando os microdados, as informações mais detalhadas das avaliações. Estão parados porque não têm os microdados? O que estamos fazendo com eles? Vamos usá-los. Podem não ser os mais atuais, mas não são tão diferentes de um ano para o outro. Vamos aprimorar o que temos”;
3. terceiro, na opinião do chefe da Unidade de Educação do Banco Interamericano de Desenvolvimento (BID), Marcelo Cabrol, a avaliação deve ser utilizada para entender melhor o que há para ensinar e aprender.

Neste enfoque, o Enade que é um dos procedimentos de avaliação do SINAES, realizado pelo INEP, por meio da lei nº10.861, de 14 de Abril de 2004, autarquia vinculada ao MEC, possui quatro fatores fundamentais e importantes para este exame: a melhoria da qualidade da educação superior; a orientação da expansão de sua oferta; o aumento permanente da sua eficácia institucional a efetividade acadêmica e social; o

aprofundamento dos compromissos e responsabilidades sociais das instituições de educação superior, por meio da valorização de sua missão pública, da promoção dos valores democráticos, do respeito à diferença e à diversidade, da afirmação da autonomia e da identidade institucional. Detalhes em INEP (2004 a 2013).

Conforme disposição do *art.5º*, §5º, da Lei nº. 10.861/2004,

[...] o Enade constitui-se componente curricular obrigatório dos cursos de graduação, sendo inscrita no histórico escolar do estudante somente a situação regular com relação a essa obrigação. A periodicidade máxima de aplicação do Enade em cada área é trienal e os estudantes selecionados que não comparecerem ao Exame estarão em situação irregular junto ao Enade. Todavia estão habilitados ao Enade os estudantes em final de primeiro ano (ingressantes) e de último ano (concluintes) das áreas e cursos a serem avaliados. No entanto, conforme estabelece a respectiva Portaria Normativa MEC de cada ano. A cada ano o MEC, por Portaria Normativa, define os cursos que participarão da prova e estabelece as regulamentações complementares relacionadas, ou seja, dependendo do ano podem ser convocados os ingressante e concluintes, ou somente os estudantes concluintes. [...]

As provas são elaboradas e aplicadas por instituição, ou consórcio de instituições, contratada por processo licitatório que, comprove capacidade técnica em elaboração de prova, preparo de instrumentos, aplicação e avaliação, e que atenda aos requisitos estabelecidos no Projeto Básico, conforme detalhes na página do INEP. As provas possuem questões de avaliação da formação geral, comuns aos cursos de todas as áreas, e questões específicas para cada curso. Nas duas partes, as questões foram discursivas e de múltipla escolha, envolvendo situações-problema e estudos de casos. Contudo, é abordado apenas as questões de resposta objetiva, pois as mesmas contemplam diferentes momentos da vida acadêmica do estudante, dividindo o nível de complexidade entre baixo, médio e alto. Os resultados do Enade são divulgados por meio da página na internet, conforme as referências Brasil/INEP-MEC, sendo de responsabilidade dos dirigentes e coordenadores de cursos de graduação divulgar amplamente, junto ao seu corpo discente e a todo o centro acadêmico, os resultados do exame. [...]

A finalidade da avaliação externa do Enade é tencionar, promover o aprofundamento dos compromissos e das responsabilidades sociais do cidadão em conjunto com as instituições de educação superior de formação, na busca pela

valorização de sua missão pública, da promoção dos valores democráticos, do respeito à diferença e à diversidade, da afirmação da autonomia e da identidade institucional. (INEP, 2004 a 2013)

A fim de melhorar a compreensão das análises desenvolvidas neste estudo, a subseção 2.2.1 traz como exemplo um exame multidisciplinar, aplicado em larga escala nacional por meio do Enade, em que o grupo populacional observado possui a mesma característica.

2.2.1 Conjunto de dados III

O conjunto de dados III (CD3) complementa o estudo visando a gama de possibilidades de análises das avaliações multidisciplinares. Nesta pesquisa, a abordagem abrange apenas um grupo da população e os três tipos de dados abordados são de avaliações multidisciplinares. O CD1 e CD2 representam os exames internos e o CD3 o externo. Porém, tanto a quantidade de questões como principalmente os respondentes realmente diferem entre si. Assim, como o conjunto de dados III é de uma avaliação externa, a grande escala se faz presente nesta análise, logo o dimensionamento amostral é garantido por existir maior número de padrões de resposta do que quantidades de parâmetros a serem estimados.

Esse conjunto é composto pelos alunos da graduação de estatística que realizaram a prova do Enade do ano de 2009. Como o exame do Enade acontece de três em três anos, ainda pelo portal do INEP os dados de 2012, até o presente momento, não estão disponíveis no “*link*” INEP (2012) das informações estatísticas dos microdados.

Em INEP (2009), disponibiliza a prova que foi aplicada em 25 cursos de estatística e 22 cursos, equivalentes a 88,0% desse total, os quais eram vinculados a universidades. As faculdades participaram com 2 cursos (8,0%) e os centros universitários, com 1 curso (4,0%). Não houve participação de cursos ministrados em centros de educação tecnológica e institutos superiores. Logo, há predominância das instituições federais de ensino, que concentraram 76,0% dos cursos avaliados. As instituições estaduais participaram com 16,0% dos cursos e as instituições privadas, com 8,0% do total nacional. Não houve participação de instituições municipais de ensino em nenhuma região brasileira.

No Exame de 2009 em todo o Brasil, a população estatística inscrita era de 1557 e participaram da avaliação 1153 estudantes, dos quais 710 eram ingressantes que se encontrava no final do primeiro ano, e 443 concluintes, que estava cursando o último

ano. Conforme INEP (2009) a maioria dos estudantes era oriunda de instituições federais (73,3%), em menor quantidade, estudantes da rede estadual (21,9%) e da rede particular (4,8%).

A avaliação teve duração total de 4 (quarto) horas e foi composta de 40 questões, no total: 10 são de formação geral e 30 de formação específica da área. As duas partes têm questões discursivas e de múltipla escolha, a de formação geral corresponde a 25% da nota e a parte específica, a 75%.

Para as análises da TRI, considerou-se somente a prova objetiva, no total: 8 questões da formação geral. Depois das análises da comissão a questão 1 foi anulada pelo critério do coeficiente de correlação ponto bisserial; 27 de formação específica, sendo que 3 questões foram anuladas (14, 22, 35) pelo SINAES assim que divulgaram o gabarito. Na formação específica, uma questão, o item 23, também foi excluída, depois do relatório geral pelo critério do coeficiente de correlação ponto-bisserial. Portanto neste estudo analisa-se 32 questões totais e verifica-se, ainda, o motivo de duas serem canceladas pelo coeficiente de correlação.

A parte de avaliação de Formação Geral do Enade/2009 foi composta de questões com abordagens de estudos de caso, interpretação de textos, análise de gráficos e charges, simulações, dentre outros, além de verificar o domínio de conhecimentos e de diferentes níveis de habilidades e competências. Foi também avaliada a compreensão de temas que poderiam transcender ao seu ambiente próprio de formação e que eram importantes para a realidade contemporânea, bem como essenciais na formação de todos os estudantes de qualquer área da Educação Superior. Para a Formação Específica da área de Estatística, as questões foram elaboradas a partir das Diretrizes Curriculares Nacionais do Curso, aprovada e instituída pelo Conselho Nacional de Educação (CNE) do MEC.

O conjunto de dados III analisado refere-se aos respondentes presentes, concluintes, que entregaram o cartão-resposta, totalizando 443. Todavia, 7 dentre esses entregaram em branco, e para aqueles que rasuraram alguma questão ou deixaram o cartão ou a questão em branco, utilizou a sigla N.A (nenhuma alternativa). Na primeira análise descritiva, trata-se do dado real com 443 respondentes, com as proporções de acertos por questão, considera-se (G) a formação geral os itens de 1 a 8 e formação específica (Q) os itens de 11 a 37 e somente as questões dicotômicas. Os detalhes da prova e do gabarito constam no portal do INEP.

Conforme INEP (2009), as questões foram classificadas segundo índice de facilidade (IF) e índice de discriminação (ID), para todas as áreas, inclusive para o curso em

estudo afim. A classificação é transcrita na Tabela 7.

Tabela 7: Classificações das questões objetivas segundo índice facilidade e o índice de discriminação - Enade/2009

Índice de Facilidade (IF)	Classificação	Índice de Discriminação (ID)	Classificação
$IF \geq 0,85$	Muito Fácil	$ID \geq 0,40$	Muito bom
$0,85 > IF \geq 0,60$	Fácil	$0,40 > ID \geq 0,30$	Bom
$0,60 > IF \geq 0,40$	Médio	$0,30 > ID \geq 0,20$	Médio
$0,40 > IF \geq 0,15$	Difícil	$0,20 > ID$	Fraco
$0,15 > IF$	Muito difícil	-	-

Fonte: INEP (2009).

Os dados da Tabela 7 são utilizados na Tabela 8, depois de tomar afinidade com os dados na estatística descritiva, utilizando o *software R* (R Development Core Team, 2012), através do pacote *ltm* (Rizopoulos, 2006).

Assim, na Tabela 8, estão os resultados das 8 questões de formação geral e 21 da formação específica, da proporção de acertos para cada uma das questões, o ponto bisserial, *Cronbach's alpha*, ID e IF. Para uma melhor compreensão e comparação, seguem agrupadas todas as análises descritivas na Tabela 8.

A Tabela 8, no que se refere ao índice de facilidade, duas das oito questões de formação geral foram difíceis; 1 delas foi muito difícil, 3 foram médias e, por fim, apenas 1 fácil e 1 muito fácil. Esses resultados são conforme a proporção de acertos da Tabela 7 de classificação, proposta pelo Enade/2009. Utiliza-se a mesma classificação para o índice de discriminação, o ponto bisserial informa que ao incluir as questões, todas apresentaram uma boa distribuição, com ressalva à G1, que apresentou um índice de discriminação fraco.

De acordo com os índices obtidos da proporção de acertos, as questões objetivas da prova foram assim avaliadas para a área de Estatística: das oito questões, somente a G1 teve o percentual abaixo de 15%, ou seja, somente uma questão foi muito difícil na avaliação. As questões G7 e G8 foram fácil e muito fácil. Quanto aos acertos na formação específica, dentre as 24 questões, 17 apresentaram valores em torno de $0,40 > IF \geq 0,15$, e somente 7 entre $0,60 > IF \geq 0,40$. Logo, na formação geral, a prova foi composta por questões difíceis e médias.

O coeficiente *Chonbach's alpha*, na formação geral, com o valor de 0,484 para as oito questões, indica uma média consistente, mas ao retirar as duas questões em negrito, a G1 e G4, a consistência interna entre as questões aumenta. Ao se observar o AC na formação específica, as 21 questões possuem uma consistência interna de 0,741 e, ao se

excluir a questão G23, aumentaria a confiabilidade entre as questões de 0,744. Maiores detalhes podem ser vistos em INEP (2009).

Tabela 8: Índices de facilidade conforme a proporção de acertos, índices de discriminação de acordo com o ponto bisserial e o *Cronbach's alpha* para as questões da formação geral e específica

443 respondentes						
Questões	Proporção % acertos	Ponto Bisserial		<i>Cronbach's alpha</i> *Todas as questões e Excluindo a questão	Índice de Facilidade	Índice de Discriminação
		incluir	excluir			
Formação Geral				0,484*		
G1	10,30%	0,24	0,04	0,497	Muito difícil	Fraço
G2	58,30%	0,43	0,12	0,476	Médio	Bom
G3	54,50%	0,51	0,22	0,435	Médio	Muito Bom
G4	34,90%	0,33	0,03	0,521	Difícil	Médio
G5	39,90%	0,49	0,2	0,456	Difícil	Bom
G6	53,50%	0,51	0,31	0,413	Médio	Bom
G7	84,90%	0,51	0,31	0,394	Fácil	Muito Bom
G8	87,90%	0,49	0,31	0,398	Muito fácil	Muito Bom
Formação Específica				0,741*		
Q11	51,80%	0,36	0,25	0,736	Médio	Bom
Q12	56,40%	0,29	0,17	0,74	Médio	Médio
Q13	17,30%	0,44	0,37	0,73	Difícil	Muito bom
Q15	38,50%	0,36	0,25	0,734	Difícil	Bom
Q16	36,90%	0,43	0,33	0,73	Difícil	Muito bom
Q17	27,00%	0,38	0,28	0,733	Difícil	Bom
Q18	15,50%	0,32	0,24	0,735	Difícil	Bom
Q19	27,20%	0,43	0,34	0,729	Difícil	Muito bom
Q20	16,20%	0,28	0,19	0,739	Difícil	Médio
Q21	29,20%	0,45	0,35	0,729	Difícil	Muito Bom
Q23	26,80%	0,18	0,08	0,744	Difícil	Fraço
Q24	30,60%	0,29	0,18	0,739	Difícil	Bom
Q25	18,50%	0,41	0,32	0,731	Difícil	Muito bom
Q26	23,20%	0,45	0,37	0,727	Difícil	Muito bom
Q27	48,20%	0,49	0,39	0,725	Médio	Muito bom
Q28	31,80%	0,38	0,28	0,731	Difícil	Bom
Q29	30,20%	0,28	0,17	0,74	Difícil	Médio
Q30	49,60%	0,3	0,19	0,738	Médio	Bom
Q31	41,50%	0,47	0,37	0,727	Médio	Muito bom
Q32	51,50%	0,27	0,16	0,739	Médio	Médio
Q33	32,40%	0,5	0,41	0,724	Difícil	Muito bom
Q34	19,10%	0,43	0,34	0,73	Difícil	Muito bom
Q36	20,80%	0,42	0,34	0,732	Difícil	Muito bom
Q37	44,40%	0,46	0,35	0,726	Médio	Muito bom

Nota: Questão (Q). No *Cronbach's alpha*, para verificar a confiabilidade dos itens, a notação (*) significa que todas as questões são consideradas, excluindo as que não apresentam asterisco. As questões em negrito são as questões que apresentaram valores inadequados para a proporção de acerto, ponto bisserial e *Chonbach's alpha*.

Recomenda-se que sejam verificadas as questões G1, G4 e Q23 da prova do Enade/2009 do curso de Estatística, além de analisá-las pedagógica e metodologicamente,

juntamente com profissionais da área, a fim de que se saibam os motivos que levaram a estatística descritiva a apresentar esses valores.

Para que as análises descritivas sejam exauridas, é preciso prosseguir considerando os 443 concluintes, mas sem realizar a divisão estabelecida pelo Enade, ou seja, a prova será totalmente considerada, as 32 questões, já que todos os itens foram respondidos pelos alunos de estatística.

Os resultados da estatística descritiva seguem no Apêndice A, devido as proporções de acertos permanecerem com os mesmos resultados, logo o índice de facilidade também continua o mesmo da Tabela 8. Os resultados do ponto bisserial nas questões G1 a G8 diferem entre a Tabela 8 e o Apêndice A, pois os valores reduziram ao serem incluídas e excluídas as questões, mas houve pouca diferença de valores para os outros itens.

O AC apresentou valor de 0,767 quando as 32 questões foram consideradas no Apêndice A, uma consistência interna, cujo valor se aproximou quando foram analisados somente os 24 itens da formação específica da Tabela 8. Mas, nesse caso, os resultados da formação geral foram diferentes, pois a questão G4 não foi indicada para ser excluída e sim a questão G2. Assim, permaneceram as indicações das questões G1 e G23 em ambas as Tabelas e no relatório do INEP (2009).

Os objetivos da pesquisa são analisar a prova e verificar quantas habilidades o aluno utilizou para responder cada item, portanto, a prova geral será detalhada, ou seja, sem dividir a prova em formação geral ou específica para verificar se a TRI relata e capta essas observações. As análises descritivas estão no Apêndice B, que aborda os 436 respondentes, ou seja, foram excluídos os 7 respondentes que zeraram a prova, com todos os 32 itens, sem dividir por formações, para melhor compreensão, a fim de que se garantam e englobem-se os resultados do conjunto de dados III. Numeraram-se as possibilidades de como tratá-las:

1. 436 respondentes e 32 itens;
2. 436 respondentes, com a exclusão das questões G1 e Q23;
3. 436 respondentes, com a exclusão das questões G1, G2, e Q23.

Conforme o Apêndice B, que corresponde ao item 1, as estatísticas descritivas indicaram as mesmas questões (G1,G2 e Q23) observadas no Apêndice A e a G1 e Q23 na Tabela 8, a partir do item 2 e quando as questões indicadas foram excluídas, nenhuma outra questão foi observada para ser excluída, a indicação permanece entre os itens G1,

G2 e Q23.

Para dar continuidade nos estudos do CD3, são abordados os 436 respondentes considerando as 32 questões da prova, visto que os sete que entregaram a prova em branco não influenciaram nos resultados e também sem dividir em formação geral e nem específica. Dado que na estatística descritiva, conforme a Tabela 8, Apêndice A e B, os resultados não tiveram alterações consideráveis, mas que devem ser analisados para confirmar se realmente nenhuma outra questão merece detalhamento dos valores descritivos, a partir dos estudos descritivos.

Para se verificar o desempenho de acertos nas questões, em relação à probabilidade de acerto ao acaso, a Tabela 9 e a Figura 5 mostram a frequência de acertos das questões. Assim, é possível verificar qual seria o comportamento dos resultados observados no exame, sob a hipótese de que os alunos responderam aleatoriamente todas as 32 itens. Ressalva-se que a proporção de acerto, ou seja, probabilidades de ambas as formações que compõem a prova e ao se considerar a habilidade geral, os valores observados e esperados respondidos pelos concluintes são os mesmos.

Tabela 9: Frequência de acertos do valor observado e o número esperado de acerto casual do CD3

Acertos	Frequência	N. Esp. casual	Acertos	Frequência	N. Esp. casual
0	0	0,35	16	19	0,05
1	0	2,81	17	14	0,01
2	1	10,88	18	12	0
3	3	27,2	19	9	0
4	5	49,3	20	7	0
5	13	69,02	21	1	0
6	9	77,65	22	8	0
7	27	72,1	23	3	0
8	30	56,33	24	3	0
9	34	37,55	25	2	0
10	40	21,59	26	2	0
11	27	10,8	27	2	0
12	35	4,72	28	1	0
13	18	1,82	29	1	0
14	23	0,62
15	16	0,18	32	0	0

Nota: Entre 30 a 32, a frequência de acertos e o Número Esperado (N.Esp) de acerto casual é zero.

Conforme a Tabela 9, pode-se observar que o acerto esperado casual dos 436 respondentes em acertar 18 ou mais questões é um evento raro sob a hipótese, e que foram observados 12 alunos com 18 ou mais acertos.

Na Tabela 9, somente o valor de acertar 9 questões que aproxima entre o acerto

observado do esperado, pois se esperava em torno de 37 alunos no acerto casual das 9 questões, mas na realidade, no Enade, 34 concluintes de estatística acertaram. Além de que, ninguém acertou uma questão e esperava no acerto casual, pelo menos duas questões corretas. Para entender melhor o desempenho dos concluintes, segue a Figura 5.

A Figura 5 mostra que na prova do Enade/2009 os alunos de estatística tiveram um desempenho melhor do que era de se esperar no acerto casual, ou seja, a partir de 10 acertos existe mais alunos acertando as questões. Quando considerou os 443 respondentes, sem excluir os sete alunos, o desempenho esperado foi o mesmo dos 436, essa diferença pouco muda as probabilidades dada pela binomial.

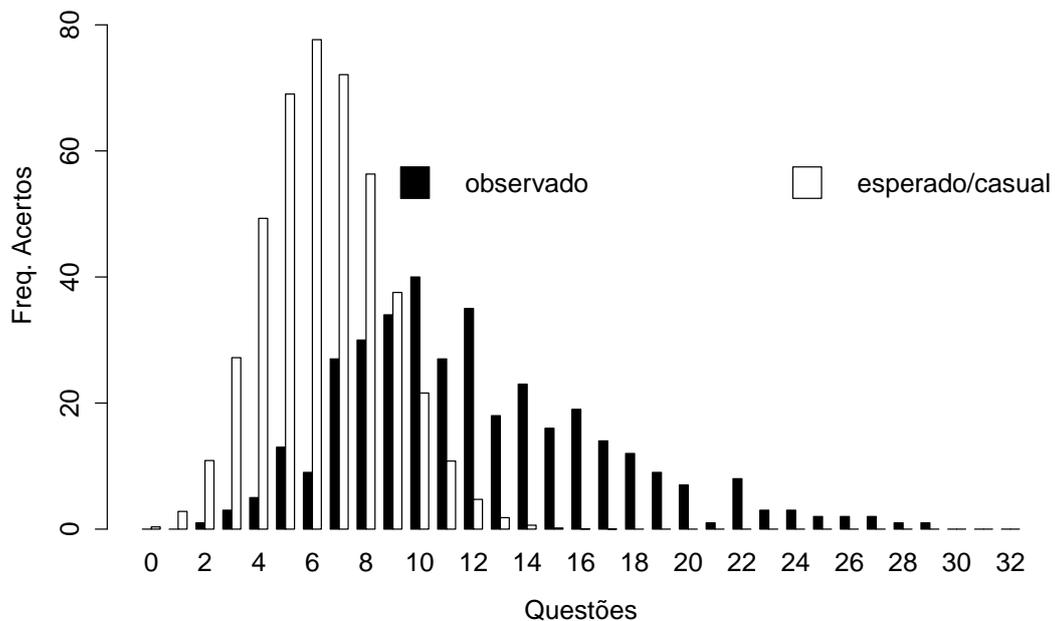


Figura 5: Frequência de acertos do valor observado e o número esperado de acerto casual dos 436 alunos que responderam o Enade/2009

Outro fator importante, na Figura 5 e neste conjunto de dados, é que nesta prova existem itens que os alunos deixaram de responder e foram considerados como dados faltantes. Observa-se que este CD3 não apresenta valores inflacionados na tabela com zeros. Mesmo considerando os dados faltantes substituídos por zero, não observou-se valores que apresentaram irregularidade no desempenho.

Para que sejam desenvolvidos estudos e haja avanço quanto às perspectivas de avaliar os resultados nas avaliações multidisciplinares, segue, no Capítulo 3, a metodologia

para aplicar a TRI nos exames educacionais e, no Capítulo 4, as análises dos resultados quando foram retirados os que zeram para verificar as influências nos modelos da TRI. Isso ocorreu porque para as abordagens descritas aqui não existem alterações desconsiderando dados faltantes e os zeros. Desse modo, segue o Capítulos 3 onde é abordada a metodologia da teoria de resposta ao item para os três conjuntos de dados propostos neste Capítulo.

3 METODOLOGIA

O principal objetivo deste Capítulo é apresentar os métodos estatísticos que serão utilizados e, ao refletir e instigar sobre os exames multidisciplinares, indagar e aferir sobre a Teoria Clássica do Testes e a Teoria da Resposta ao Item. Baseado neste planejamento, as reflexões sólidas alicerçadas em conhecimento já existentes são citadas no decorrer do texto referenciando alguns autores, para concretizar novos propósitos que tragam desenvolvimento na educação ao abordar a estatística. As seções 3.1 e 3.2 juntamente com as subseções trazem detalhes e particularidades quanto à metodologia abordada.

As duas teorias abordadas nesta pesquisa tiveram suas origens nos anos 50. A TCT foi axiomatizada por Gulliksen (1950) e a TRI foi inicialmente elaborada por Lord (1952) e por Rasch (1960) e, finalmente, axiomatizada por Birnbaum (1968) e por Lord (1980).

Somente no século XIX, houve registros quanto à preocupação em se medir fenômenos psicológicos. Assim, filósofos, psicólogos e estatísticos, dentre outros, preocupavam-se em encontrar um meio de mensurar traços latentes como sensações, emoções e comportamentos humanos (CHACHAMOVICH, 2007). De acordo com Pasquali e Primi (2003), tal medição deu origem à ciência denominada psicometria, cuja origem encontra-se nos trabalhos do estatístico Spearman, o qual lançou os fundamentos da teoria clássica.

Assim, a psicometria utiliza estatística e procura explicar o sentido das respostas dadas pelos indivíduos. No entanto, Pasquali (2003) descreve que é uma área da psicologia com uma concepção estatística, que explica os comportamentos e aptidões por meio de testes, cuja mensuração é feita a partir das respostas que os indivíduos fornecem aos itens. Neste sentido, segundo Pasquali (2007), Pasquali (2009), a psicometria moderna tem duas vertentes: a teoria clássica dos testes (TCT) e a teoria da resposta ao item (TRI).

A TCT se preocupa em explicar o resultado final total, isto é, o que a soma dos itens diz sobre o sujeito. Já a TRI tem o propósito de medir a habilidade do sujeito de acordo com as respostas dadas a cada item. Ou seja, a TCT analisa o resultado final e

a TRI analisa partes e probabilidades que geram o resultado final (PASQUALI; PRIMI, 2003).

Dessa forma, a TCT tem interesse em produzir testes de qualidade, enquanto a TRI se interessa por produzir tarefas (itens) de qualidade. No final, então, existem testes válidos (TCT) ou itens válidos (TRI), itens com os quais podem ser construídos tantos testes válidos quantos se quiser ou o número de itens possa permitir (PASQUALI, 2011).

Quando se assumem os itens como elementos centrais da análise, é possível:

- comparar populações (grupos de respondentes) submetidos a provas;
- captar aqueles candidatos que mais acertam ao acaso;
- valorizar aqueles que acertam as questões de forma mais coerente, ou seja, aqueles que acertam mais questões fáceis do que as difíceis.

Devido a esses fatores, a atenção está voltada para o método da TRI e aborda-se a TCT, porque as avaliações aqui estudadas como exemplos no conjunto de dados I e II utilizam essa teoria para medição do desempenho.

O questionamento a respeito da demora tanto para discutir como para abordar a TRI é inevitável, já que ambas as teorias surgiram juntas na década de 50. Ainda hoje, a TCT é a mais empregada tanto nas avaliações externas como internas, para auferir resultados na educação. A busca de melhorias é constante, e a TRI poderia ser utilizada nas avaliações internas e externas, com o propósito de realmente beneficiar a metodologia e pedagogia. Se existem tantas pesquisas e *softwares* estatísticos para a TRI, então porque ainda não foi suficiente para a utilização nas avaliações? Detalhes são observados em Moreira (2010), o qual relata as aplicações e o crescimento da TRI no Brasil e as pesquisas e artigos publicados. Por esse motivo, neste Capítulo, são abordados os conceitos e a estrutura probabilística geral do modelo e a partir do modelo geral, serão introduzidos os dois métodos de avaliação.

3.1 TEORIA CLÁSSICA DOS TESTES

A TCT tem seu foco direcionado ao próprio instrumento de medida - o teste - e emprega, para tal, uma escala macroscópica. É exatamente nesse aspecto em que a TRI fornece instrumentos adicionais não apenas para avaliar o indivíduo, mas também para o item em que reside a mais significativa distinção entre as teorias (Hambleton, Swaminathan e Rogers (1991) e Andriola (2009)).

Surgida nas primeiras décadas do século XX, começou a estruturar-se, após os trabalhos de Spearman (1907), Spearman (1913) seguido por Gulliksen (1950) e posteriormente Lord e Novick (1968) que apresentaram uma exposição completa da teoria.

Embora TCT tem servido a comunidade da educação de medição para o conhecimento, as principais vantagens são os pressupostos teóricos relativamente propensos, tornando a TCT fácil de aplicar em muitas situações de teste (HAMBLETON; JONES, 1993). As aptidões são medidas, em geral, pela soma dos itens, conforme é descrita brevemente na subseção 3.1.1.

3.1.1 Modelo da Teoria Clássica dos Testes

A Teoria Clássica dos Testes preocupa-se em explicar o resultado final total, isto é, a soma das respostas dadas a uma série de itens, expressa no chamado escore total. De acordo com Pasquali e Primi (2003), uma grande preocupação dos psicometristas era com o erro contido nesta soma. Quando o sujeito responde uma questão, ele recebe uma pontuação por essa resposta. No final da aplicação do teste tem-se um escore final, que é a soma dos pontos. O fato deste escore representar uma magnitude é porque trata-se de uma operação empírica e que está sujeita a erros.

Para Campbell e Stanley (1973), os erros são devidos a fatores estranhos; essas variáveis não controladas decorrem de várias causas tais como: defeito do próprio teste, estereótipos e vieses do sujeito, fatores históricos e ambientais aleatórios.

A grande tarefa da TCT, segundo Muñiz e Hambleton (1992, p. 17), consiste em elaborar estratégias estatísticas, ou seja, o propósito de Spearman, era “encontrar um modelo estatístico que fundamentasse adequadamente as pontuações dos testes e permitisse estimar os erros de medida associados a todo processo de medição”.

A TCT, segundo Lord e Novick (1968, p. 34), consiste em um modelo relativamente simples, que se fundamenta na observação dos escores, definidos por:

- i : item, $i = 1, 2, \dots, I$;
- j : indivíduo, $j = 1, 2, \dots, n$;
- X_{ij} : escore bruto do avaliado j no item i ;
- μ_j : escore verdadeiro do avaliado j ;
- ε_{ij} : erro de observação na resposta do avaliado j no item i .

Logo o modelo da TCT é:

$$X_{ij} = \mu_j + \varepsilon_{ij} \quad (3.1)$$

O modelo matemático, de forma inferencial, tenta entender o comportamento da realidade, muito mais por sua limitação prática do que pelo seu desejo teórico. Isso ocorre em razão do grande número de variáveis não controladas que influenciam a resposta dos candidatos, em um teste qualquer, e que um modelo teórico não poderia comportar toda sua abrangência e complexidade.

O modelo de Spearman, segundo Requena (1990, p. 26), continua sendo bastante influente na atualidade, não só pela facilidade de sua aplicação como também pelo entendimento de seus resultados. Detalhes e a prova dos escores observados dos indivíduos encontra-se em Lord e Novick (1968, p. 48). Ademais, sua conceituação deve estar fundamentada sob hipóteses que lhe dão sustentabilidade teórica e que possibilitarão ao modelo descrever, da melhor forma possível, o comportamento de uma realidade. Tais hipóteses são fundamentadas na proposta de Spearman, como se pode ver em Requena (1990, p. 26), Muñiz e Hambleton (1992, p. 28), Pasquali (2003, p.69) e Arias, Lloreda e Lloreda (2006, p. 38).

No nível do item, o modelo TCT é relativamente simples e não invoca um modelo teórico complexo para relacionar a capacidade do respondente com o sucesso em um determinado item. A dificuldade do item na TCT é definida em termos de percentagens de acertos. Quanto mais próxima a taxa de acertos estiver de 100% mais fácil é o item (PASQUALI, 2003).

Assim, segundo Pasquali (2011), existem testes válidos de TCT ou de itens válidos de TRI, itens com os quais se poderão construir tantos testes válidos quantos se quiser ou o número de itens puder permitir. Essa abordagem é feita na seção 3.2.

3.2 TEORIA DA RESPOSTA AO ITEM

Em termos estatísticos, o avanço da TRI como uma metodologia de melhor validação de análises das respostas parte do princípio de que os itens constitutivos do instrumento avaliativo assumem a característica de “elementos centrais da análise”, pois quebram o paradigma, cuja centralização faz referência a “prova” como um todo (VALLE, 2000).

Esta seção, cujo foco é a TRI, aborda com detalhes gerais os modelos da teoria na subseção 3.2.2 enquanto a subseção 3.2.3 particulariza os modelos em k -dimensões. Para análise desses modelos, a subseção 3.2.4 averigua os modelos exploratórios compensatórios embasados na análise fatorial e sugere-se como fundamento as referências bibliográficas

referenciadas ao longo do capítulo.

Com relação aos métodos de estimação e inferência para os parâmetros, na necessidade de se escolherem e /ou desenvolverem novos modelos e métodos, segue a subseção 3.2.5. A subseção 3.2.6 apresenta um breve detalhamento dos métodos computacionais, aplicados para desenvolver o Capítulo 4.

3.2.1 Considerações

De acordo com Andrade, Tavares e Valle (2000), Lord é o precursor dos estudos da TRI e na maioria das obras pesquisadas apontava que, Lord, em 1952, desenvolveu o modelo unidimensional de dois parâmetros, baseado na distribuição normal acumulada (ogiva normal). Esse modelo foi aplicado com as limitações computacionais da época e após algumas aplicações, o pesquisador sentiu a necessidade da incorporação um parâmetro que tratasse do problema do acerto casual (no caso o “chute”), então, no decorrer dos estudos, surgiu o modelo de três parâmetros.

Segundo Linden e Hambleton (1997, p. 13), por volta de 1958, Birnbaum substituiu a função distribuição da normal do modelo de Lord pela função logística. Isso facilitou o tratamento matemático nos programas desenvolvidos naquela época.

Para Pasquali e Primi (2003), foi somente após os anos 50 que os psicometristas começaram a descobrir a solução para o problema, baseados na teoria do traço latente de Lazarsfeld (1959), nos trabalhos de Lord (1952) e do dinamarquês Rasch (1960), os quais se tornaram as bases da moderna Teoria da Resposta ao Item, inclusive conhecida como a Teoria do Traço Latente, esboçada por Lord em 1952 e finalmente axiomatizada por Birnbaum em 1968 e por Lord em 1980.

A pesquisa de Rasch (1960) é baseado em um modelo, na concepção, de uma distribuição de Poisson, para o número de erros em leituras de textos. Com aprimoramentos e o uso da função logística, tornou-se conhecido como Modelo Logístico de um parâmetro (ML1) e é chamado de Modelo Rasch. Assim, o trabalho, impulsionou e motivou o desenvolvimento do Modelo Logístico com dois parâmetros (ML2) ao modelo inicialmente já proposto por Lord.

Conforme Linden e Hambleton (1997, p. 13), Birnbaum implementou, em 1968, um terceiro parâmetro no modelo de Rash, conhecido como Modelo Logístico de Três Parâmetros (ML3). Deve-se também a Birnbaum a descrição da Função de Informação do Item (utilizando um modelo de mensuração de Fisher) e também a primeira técnica

utilizando o método da máxima verossimilhança para estimar os parâmetros dos itens.

No ano de 1977, Bock e Zimowski apresentaram os modelos logísticos de 1, 2 e 3 parâmetros para duas ou mais populações de respondentes. A introdução desses modelos trouxe novas possibilidades para as comparações de rendimentos de duas ou mais populações submetidas a diferentes testes com itens comuns (ANDRADE et al., 2000).

Hambleton, Swaminathan e Rogers (1991) salientam cinco avanços e impacto que apontam porque a TRI, supera as limitações na psicometria clássica:

- os cálculos dos parâmetros clássicos dos itens (dificuldade e discriminação);
- o cálculo do nível de aptidão do sujeito independente do teste;
- a possibilidade de emparelhar itens com a aptidão do sujeito;
- definição de modelo e não precisa fazer suposições que aparentam serem improváveis, tais como os erros de medida serem iguais para todos os testandos;
- não necessitam trabalhar com testes estritamente paralelos, como exige a TCT.

A razão do atraso na aplicação desta teoria está no fato da complexidade de manipulação de seus modelos matemáticos, que apresentava algoritmos de tal dificuldade que a tecnologia computacional da época era incapaz de resolver de uma maneira útil e prática. Com o surgimento dos modelos mistos, a dificuldade não está mais na integração que requeria otimização, mas sim agora nos itens que compõem a prova.

Este problema vêm sendo solucionado, detalhes na subseção 3.2.5 e 3.2.6, com o avanço da tecnologia da informática, da disponibilidade de *softwares* apropriados que entraram no mercado nos anos de 1980 e, com os procedimentos de estimação dos parâmetros. Assim, a TRI vem se tornando a técnica predominante no campo de estudos para testes. O primeiro software para as análises da TRI surgiu somente em 1979 com o BICAL de Wright, Mead e Bell (1979), seguidos depois pelo LOGIST de Wingersky, Barton e Lord (1982) e pelo BILOG de Kim (1997). Para Pasquali (2011), atualmente, com o avanço tecnológico, a TRI veio para ficar e inovar grande parte da teoria clássica da Psicometria. Para Reckase (2009), ainda são necessários avanços educacionais e estudos porque o teste é construído para o resultado total sem centrar-se nos resultados dos item que compõem o teste. É preciso detalhar e fornecer uma descrição das habilidades dos respondentes, ou seja, modelar o item respondido. Portanto, os modelos da TRI estão apresentados na subseção 3.2.2.

3.2.2 Modelos da TRI

Visto que ainda buscam-se medidas avaliativas em instrumentos de medição de desempenho, muitas questões em educação com o teor avaliativo permanecem sem respostas. Portanto, quando se assumem os itens como elementos centrais da análise, é possível: comparar populações (grupos de respondentes) submetidos à provas; captar aqueles candidatos que mais acertam ao acaso; valorizar aqueles que acertam as questões de forma mais coerente, ou seja, aqueles que acertam mais questões fáceis do que difíceis.

Entretanto, Moreira (2010) relata que aqueles que acertam mais questões difíceis do que fáceis teriam a sua pontuação prejudicada, uma vez que, de acordo com a lógica da TRI, esses indivíduos deveriam acertar as mais fáceis já que sabem as mais difíceis. Assim, é possível que candidatos com mais itens acertados possam ter uma nota inferior quando comparados a outros candidatos com menos itens.

Os modelos de respostas ao item existentes se diferenciam quanto à forma matemática da função característica do item ou ao número de parâmetros especificados no modelo. Os modelos podem conter um ou mais parâmetros relacionados aos itens/questões e ao indivíduo. A distinção principal entre modelos da TRI refere-se à suposição sobre o relacionamento entre a seleção de opções de uma resposta e o nível do traço latente. Existem dois tipos de processo de resposta: o acumulativo e o de desdobramento. Para esse trabalho serão utilizados os modelos acumulativos da TRI em que a probabilidade de um indivíduo dar ou escolher uma resposta correta ao item aumenta com o aumento do seu traço latente, isto é, níveis maiores de traço latente conduzem a valores mais altos de probabilidade de resposta correta. Os modelos citados seguem detalhados em Muraki (1992) e de Araujo, Andrade e Bortolotti (2009).

Para tanto, a TRI tem utilizado duas funções matemáticas para caracterizar os parâmetros dos componentes de itens em um teste: a função logística e a função distribuição da normal padronizada também conhecida como ogiva gaussiana (MUÑIZ; HAMBLETON, 1992). Ambas tem contradomínio em $(0,1)$ e, nesta escala, situa-se a probabilidade de um examinado acertar um item específico, de modo que, os modelos usados pela TRI procuram se adequar a essas funções.

Os modelos com relação à quantidade de traços latentes ou habilidades podem ser classificados em: modelos unidimensionais, quando consideram uma habilidade e multidimensionais quando se examinam duas ou mais habilidades num mesmo instrumento.

Os modelos unidimensionais são os mais utilizados na prática e toda a teoria

matemática da TRI foi fundamentada com este enfoque, porém encontra-se, por exemplo, em Linden e Hambleton (1997) referências sobre os modelos multidimensionais e aplicações recentes em artigos de Chalmers (2012), em que o modelo unidimensional está inserido como caso particular do modelo multidimensional, pois é tratado como modelo que apresenta uma dimensão.

O estudo da dimensionalidade (k -dimensionalidade) nos exames multidisciplinares (k -multivariado) está diretamente relacionado à validade do instrumento. Nesta tese, trata-se como modelos de k -dimensões. Todos os exemplos são de respostas dicotômicas e a investigação da dimensionalidade é feita a partir da análise de componentes principais sobre a matriz de correlação tetracórica, análise paralela e, principalmente, pela abordagem da análise fatorial de informação completa, que se baseia em ajustes de modelos da TRI. Para esta análise, utilizam-se métodos computacionais descritos na seção 3.2.6. Nesta última abordagem, estuda-se a capacidade do instrumento em produzir novas dimensões interpretáveis, a fim de que se conduzam e se analisem algumas relações estruturais nos itens da prova multidisciplinar.

Segundo Reckase (1997), um dos modelos de resposta ao item que tem recebido bastante atenção é o probito k -dimensional de 2 parâmetros, o qual é adequado para a análise de exames em que cada questão exige do respondente dois ou mais traços latentes para que uma resposta correta seja obtida. Pormenor em: Adams, Wilson e Wang (1997), Nojosa (2002) Marvelde (2006), Matos (2001) e Reckase (2009) encontram-se detalhes para referência.

Nos exames multidisciplinares das avaliações internas dos conjuntos de dados I e II, essa metodologia é aplicada com base nas referências que abordam a k -dimensionalidade, mas não há evidências de trabalhos publicados em exames multidisciplinares de pequena escala. Logo é uma das propostas de análise e discussão do Capítulo 4. Nos estudos por Nojosa (2002) de como utilizar a aplicabilidade da dimensionalidade em exames externos de larga escala, em teste do Enem do Inep/MEC e recentemente nos estudos de Barbetta, Trevisan e al. (2013), na prova geral do vestibular da UNESP/2012.

Em ambos trabalhos, os estudos da dimensionalidade apontavam mais do que uma habilidade para responder o item corretamente, ou seja, cada item valoriza o conhecimento interdisciplinar e contextualizado; e o teste é dito ser do tipo multidimensional, pois leva à aplicabilidade dessa metodologia nos exames multidisciplinares de larga escala, como é o caso do exemplo do conjunto de dados III e que estende-se os estudos também para as avaliações internas.

Mais detalhes sobre os modelos matemáticos utilizados na TRI podem ser encontrados em Baker (2001), Baker e Kim (2004), Boomsma, Duijn e Snijders (2000), Embretson e Reise (2000), Hambleton e Jones (1993), Hambleton, Swaminathan e Rogers (1991), Lord (1980), Matos (2001), Reckase (2009), Linden e Hambleton (1997).

3.2.3 Modelos de k dimensões

As teorias do traço latente (que vêm dos anos 30) afirmam que existe um conjunto de habilidades variáveis não observáveis diretamente que estão por detrás de um desempenho comportamental qualquer. Assim, um sujeito qualquer se situa num espaço de “ k ” dimensões, agora representado por θ , e seu desempenho depende do tamanho que ele possui em cada um desses θ , de sorte que o seu comportamento pode ser expresso como um vetor de escores ou pesos nos vários θ , isto é, desempenho = $f(\theta_1, \theta_2, \theta_3, \dots, \theta_n)$.

Entretanto, em sua maioria, os modelos da TRI postulam que há apenas uma aptidão responsável pela realização de um conjunto de tarefas (itens de um teste). Parece pacífico que qualquer desempenho humano é sempre multideterminado ou multimotivado, dado que mais de um traço latente entra na execução de qualquer tarefa, como o modelo geral do traço latente afirma. Contudo, para satisfazer o postulado da unidimensionalidade, é suficiente admitir que haja uma aptidão dominante (um fator ou traço dominante) responsável pelo desempenho num conjunto de itens de um teste, assim, supõe-se que tal fator é medido pelo teste.

O postulado da unidimensionalidade continua importante dado que a TRI ainda não possui soluções muito usadas para modelos multidimensionais, embora esses sejam pesquisados já de longa data conforme Mulaik (1972), Samejima (1974), Bock e Aitkin (1981) e Reckase (1997). Isso também não quer dizer que os autores se entendem completamente sobre o que seja dimensionalidade e como ela deve ser estimada (PASQUALI; PRIMI, 2003). Por ainda não estar bem definido nem determinado quando devem ser usado a dimensionalidade e as vantagens de uma ou mais dimensões, é de grande valia a abordagem dos conjuntos de dados adotadas nesta pesquisa, como análise de referência e proposta de estudo, para verificar os resultados de um avaliação em pequena/larga escala e quando deve ser tratada de exame multidisciplinar de k -dimensão.

Para o desenvolvimento do trabalho, o modelo proposto é o logístico com itens dicotômicos que dependem dos itens, já que as provas são multidisciplinares. Aborda-se os traços latentes, conforme a dimensionalidade em que ela foi inserida. Assume-se a independência local, conforme Hambleton, Swaminathan e Rogers (1991, p. 16), em que,

para os exames de múltipla escolha, é possível o examinando responder acertadamente um item sem conhecer o assunto do outro item.

Uma pessoa avaliada com habilidade elevada para a elaboração do tema que está em avaliação pode extrapolar este princípio no exame multidisciplinar. Logo, ele tem elevada probabilidade de responder corretamente ao item. Ao mesmo tempo, itens que melhor avaliam a habilidade de um sujeito correspondem ao nível de traço necessário para a correta resolução. Green et al. (1984) afirmam que itens demasiadamente fáceis ou difíceis para um participante, contribuem pouco para estimar a sua habilidade e dificultam até para verificar a dimensionalidade na prova multidisciplinar como habilidade do conhecimento.

Os modelos Multidimensionais da Teoria de Resposta ao Item (TRIM) são casos gerais da TRI, construídos sobre a premissa de que a função matemática inclui como parâmetros um vetor de características de várias pessoas que descrevem habilidades e conhecimentos que a pessoa traz para um teste. Há também um vetor de características do item que descreve a dificuldade e a sensibilidade para diferenciar as características dos respondentes, ou seja, agora na TRIM, as questões são organizadas em grupos e os indivíduos tem como escore um vetor ao invés de um escalar.

Para Reckase (2009), os modelos Unidimensionais da Teoria de Resposta ao Item (TRIUI), compreendem um conjunto de modelos da TRI que têm como premissa básica as interações de uma pessoa com os itens do exame e podem ser adequadamente representados por uma expressão matemática que contém um único parâmetro descrevendo a ação das características do indivíduo.

Assim, a representação básica de um modelo é pela TRIM que é um caso geral tanto da análise fatorial como na TRIUI

Há dois propósitos para descrever os modelos da TRI neste Capítulo. O primeiro é apresentar conceitos básicos sobre a modelagem da interação entre pessoas e itens de teste a partir de modelos simples que permitem uma explicação dos conceitos. O segundo objetivo é identificar quando e quanto vale a pena para um teste multidisciplinar utilizar os modelos da TRIUI e TRIM. Tal como acontece com todos os modelos científicos de fenômenos observados, os modelos são úteis apenas na medida em que eles proporcionam aproximações razoáveis para relações do mundo real. Além disso, o uso de modelos mais complexos só se justifica quando eles proporcionam maior precisão ou novas percepções. Assim, inicia-se com a descrição do modelo de k -dimensão que permeia a TRIUI e a TRIM.

Inicialmente para descrever o modelo da TRI, considere:

\mathbf{Y} : um vetor aleatório;

i : item, $i = 1, 2, 3, \dots, I$;

j : indivíduo, $j = 1, 2, 3, \dots, J$;

P_{ij} : probabilidade de um indivíduo j responder corretamente ao item i ;

Y_{ij} : variável aleatória correspondente à resposta de um indivíduo j ao item i em um teste qualquer;

y_{ij} : valor observado da resposta dada a probabilidade condicionada;

\mathbf{Y}_j : $(Y_{1j}, Y_{2j}, \dots, Y_{Ij})$ vetor total de respostas do respondente j submetido ao um teste e composto por itens;

$\boldsymbol{\theta}$: vetor do traço latente;

k : dimensionalidade do instrumento de medida ou número de habilidades envolvidas;

$\boldsymbol{\theta}_j$: um vetor k -dimensão do traço latente do indivíduo j ; logo, $\boldsymbol{\theta}_j = (\theta_{1j}, \theta_{2j}, \dots, \theta_{Kj})$;

$P(Y_i|\boldsymbol{\theta}_j)$: probabilidade de resposta ao item i , dado uma certa habilidade do indivíduo j .

$P(Y_{ij} = 1|\boldsymbol{\theta}_j)$: probabilidade de o respondente j , condicionado no seu traço latente $\boldsymbol{\theta}_j$, responder corretamente, ou concordar, satisfazer as condições do item i , denominada de Função de Resposta do Item TRI;

α_i : é o parâmetro relacionado à discriminação do item;

$\boldsymbol{\alpha}_i$: é o vetor de parâmetros relacionados à discriminação do item i , $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ik})$;

β_i : é o parâmetro que representa a dificuldade do item;

c_i : é o parâmetro que representa a probabilidade de um indivíduo com baixa habilidade responder corretamente ao item i ;

$\boldsymbol{\eta}_i$: vetor de parâmetros correspondentes às características do item;

d_i : é o parâmetro relacionado à dificuldade do item, medido na mesma escala da habilidade, ou seja, é o escalar que representa a dificuldade de cada item $d_i = -\alpha_i\beta_i$;

D : é um fator escala, introduzido para tornar a função logística mais próxima possível da função da ogiva normal ao utilizar $D = 1,702$, caso contrário, utiliza a constante $D = 1$.

Ao assumir a independência condicional sob a forma geral do modelo da TRI, temos:

$$P(Y_{ij} = 1|\boldsymbol{\theta}_j) = \prod_{i=1}^I P(Y_{ij} = y_{ij}|\boldsymbol{\theta}_j, \eta_i). \quad (3.2)$$

Portanto:

$$Y_{ij}|\boldsymbol{\theta}_j \sim \text{Bernoulli}(P_{ij}) \quad (3.3)$$

sendo (C_o) a matriz de covariância dos traços latentes, tem-se,

$$\boldsymbol{\theta}_j \sim NM(0, C_o) \quad (3.4)$$

ou considerando cada dimensão individualmente assume-se que

$$\theta_{kj} \sim N(0, 1). \quad (3.5)$$

Com o caso particular, tem-se para $k = 1$ o traço latentes conforme:

$$\theta_{1j} \sim N(0, 1). \quad (3.6)$$

Assume-se o modelo logístico com um traço latente/unidimensional (MLD1):

$$P_{ij} = P(Y_{ij} = 1|\boldsymbol{\theta}_j) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-D(\alpha_{i1}\theta_{1j} + d_i)}}. \quad (3.7)$$

Outro caso particular considera-se a dimensão onde $\theta_{kj} \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$, para $k = 2$ os traços latentes são independentes entre si, conforme:

$$\begin{cases} \theta_{1j} \sim N(0, 1), \\ \theta_{2j} \sim N(0, 1). \end{cases}$$

Logo, o modelo logístico com dois traços latente/multidimensional (MLD2) é:

$$P_{ij} = P(Y_{ij} = 1|\boldsymbol{\theta}_j, \alpha_i, d_i, c_i) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-D(\alpha_{i1}\theta_{1j} + \alpha_{i2}\theta_{2j} + d_i)}}. \quad (3.8)$$

Considera-se ϕ a função de distribuição acumulada da normal padrão e que o modelo completo descrito na forma hierárquica é:

$$P_{ij} = P(Y_{ij}|\boldsymbol{\theta}_j, \eta_i) = \phi(\boldsymbol{\theta}_j \boldsymbol{\alpha}_i + d_i). \quad (3.9)$$

Vale ressaltar que como o expoente $\boldsymbol{\alpha}_i(\boldsymbol{\theta} - \beta_i)$, o que é idêntico a $\boldsymbol{\alpha}_i \boldsymbol{\theta} - \boldsymbol{\alpha}_i \beta_i$ e ao se substituir $-\boldsymbol{\alpha}_i \beta_i$ por d_i , a expressão anterior é dada por $\boldsymbol{\alpha}_i \boldsymbol{\theta} + d_i$.

Assim, a extensão do modelo multidimensional, é obtida ao se considerar $\boldsymbol{\theta}_j$ como vetor ($1 \times k$) referente ao j examinado; enquanto, $\boldsymbol{\alpha}_i$ é um vector ($1 \times k$) relativo aos

parâmetros de discriminação do item i , em que k é o número de dimensões do fator latente e o termo d_i é um escalar que representa a dificuldade de cada item.

Portanto, considerando o modelo probabilístico,

$$P_{ij} = F_i(\eta_{ij}) = P(Y_{ij} = y_{ij} | \boldsymbol{\theta}_j, \boldsymbol{\alpha}_i), \quad (3.10)$$

em que $i = 1, \dots, I$, $j = 1, \dots, J$, para representar a probabilidade (condicional) de uma determinada resposta y_{ij} , dado um vetor k -dimensão com traços latentes

$$\boldsymbol{\theta}_j = (\theta_{j1}, \dots, \theta_{jk}, \dots, \theta_{jK})' \quad (3.11)$$

e um vetor de parâmetros correspondentes às características do item i , α_i . A quantidade $\eta_{ij} \in \mathbb{R}$ é uma função adequadamente escolhida dos parâmetros $\theta_j \in \mathbb{R}^k$ e $\alpha_i \in \mathbb{R}^{l+k}$ com $l = 1$ ou $l = 2$ com possíveis restrições no espaço paramétrico de α_i que depende da função da resposta do item, ver detalhes em Matos (2008).

Para entender as análises propostas da dimensionalidade, primeiro, serão abordados os modelos da TRI e dentre os modelos para itens com resposta dicotômica ou itens de múltipla escolha (corrigido como certo/errado) acumulativos. Destacam-se: o modelo logístico de um parâmetro $ML1$, o modelo logístico de dois parâmetros $ML2$ e o modelo logístico de três parâmetros $ML3$, descritos nas Equações 3.7 e 3.8. Detalhes desses modelos sugerem consultas em Andrade, Tavares e Valle (2000) e Reckase (2009).

A partir do $ML3$, o parâmetro de acerto casual (c_i) refere-se à probabilidade de um respondente com baixa habilidade em acertar o item (ao acaso). Quando não se necessita desse parâmetro na análise, ele pode ser retirado e tem-se o modelo logístico de dois parâmetros $ML2$. Ao se retirar o parâmetro de discriminação (α_i), tem-se o modelo de um parâmetro $ML1$ chamado de modelo de Rasch para a unidimensionalidade e para a multidimensionalidade, cujo valor é considerado constante ou 1 e que considera que todos os itens discriminam da mesma forma.

Segundo Chalmers (2012), os modelos $ML3$, $ML2$ e $ML1$ conhecidos na TRI são análogos da TRIM e tratados como $3PL$, $2PL$ e $1PL$ para avaliar três, dois e um parâmetro logístico, respectivamente.

Seguem pontualmente os modelos conforme a hierarquia:

Modelo logístico multidimensional de 3 parâmetros - $ML3/3PL$ com $3 + (k - 1)$

parâmetros:

$$P_{ij} = P(Y_{ij} = 1|\theta_j, \eta_i) = c_i + (1 - c_i) \frac{1}{1 + e^{[-(\sum_{k=1}^k \alpha_{ik} \theta_{kj} + d_i)]}}. \quad (3.12)$$

Modelo logístico multidimensional de 2 parâmetros - $ML2/2PL$ com $2 + (k - 1)$ parâmetros:

$$P_{ij} = P(Y_{ij} = 1|\theta_j, \eta_i) = \frac{1}{1 + e^{[-(\sum_{k=1}^k \alpha_{ik} \theta_{kj} + d_i)]}}. \quad (3.13)$$

Modelo logístico multidimensional de 1 parâmetro - $ML1/1PL$ com $1 + k$ parâmetros:

$$P_{ij} = P(Y_{ij} = 1|\theta_j, \eta_i) = \frac{1}{1 + e^{[-(\sum_{k=1}^k \alpha_k \theta_{kj} + d_i)]}}. \quad (3.14)$$

A partir do modelo logístico $ML3$ da TRIU e por ser o modelo com mais parâmetros da TRI, abordam-se os demais modelos que são casos particulares com restrições especiais dos parâmetros. Assim, para as habilidades em estudo detalham-se 09 (nove) modelos logísticos (ML), que são aplicados para verificar, comparar e analisar os resultados dos parâmetros discriminação, dificuldade e acerto ao acaso (α_i, β_i, c_i). Portanto, os ML unidimensionais conforme os parâmetros são:

Modelo Logístico Rasch - M_1 ($1, \beta_i, 0$). Conhecido pela bibliografia como ML_1 de Rasch, o parâmetro de dificuldade, estimado para cada questão e considerando: $\alpha_i = 1, \beta_i = \beta_i, c_i = 0$. Logo o modelo proposto é:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-(\theta_j - \beta_i)}}. \quad (3.15)$$

Modelo Logístico Rasch - M_2 ($\alpha, \beta_i, 0$). Do modelo de Rasch com parâmetros de discriminação diferente de 1, mas igual entre as questões e o parâmetro de dificuldade estimado para cada item, considerando: $\alpha_i = \alpha, \beta_i = \beta_i, c_i = 0$. Logo o modelo proposto é:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-\alpha(\theta_j - \beta_i)}}. \quad (3.16)$$

Modelo Logístico - M_3 ($1, \beta_i, c$). Consideram os parâmetros de discriminação com valor igual a 1 e o acerto ao acaso, o mesmo valor para cada questão, conforme: $\alpha_i = 1, \beta_i = \beta_i, c_i = c$. Logo o modelo proposto é:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) = c + (1 - c) \cdot \frac{1}{1 + e^{-(\theta_j - \beta_i)}}. \quad (3.17)$$

Modelo logístico - M_4 (α, β_i, c). Consideram os parâmetros de discriminação e o acerto ao acaso, os quais são estimados em um único valor para todas as questões e a dificuldade é estimada para todas as questões, conforme: $\alpha_i = \alpha, \beta_i = \beta_i, c_i = c$. Logo o modelo proposto é:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) = c + (1 - c) \cdot \frac{1}{1 + e^{-\alpha(\theta_j - \beta_i)}}. \quad (3.18)$$

Modelo Logístico - M_5 ($\alpha_i, \beta_i, 0$). O modelo é o de dois parâmetros conhecido na bibliografia como ML_2 , consideram-se os parâmetros de discriminação e o acerto ao acaso, sendo estimado um valor para todas as questões e para dificuldade, é estimado um valor para cada questão, conforme: $\alpha_i = \alpha_i, \beta_i = \beta_i, c_i = 0$. Logo o modelo proposto é:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-\alpha_i(\theta_j - \beta_i)}}. \quad (3.19)$$

Modelo Logístico - M_6 ($1, \beta_i, c_i$). O modelo tem dois parâmetros em que dificuldade e acerto ao acaso devem ser estimados para cada questão, ou seja, considerando $\alpha_i = 1, \beta_i = \beta_i, c_i = c_i$. Logo o modelo proposto é:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-(\theta_j - \beta_i)}}. \quad (3.20)$$

Modelo Logístico - M_7 (α_i, β_i, c). Apresenta parâmetros de discriminação e dificuldade, os quais são estimados para cada questão e o acerto ao acaso possui um valor para todas as questões, conforme: $\alpha_i = \alpha_i, \beta_i = \beta_i, c_i = c$. Logo o modelo proposto é:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) = c + (1 - c) \cdot \frac{1}{1 + e^{-\alpha_i(\theta_j - \beta_i)}}. \quad (3.21)$$

Modelo Logístico - M_8 (α, β_i, c_i). Os parâmetros de dificuldade e acerto ao acaso são estimados para cada questão e a discriminação possui um valor para todas as questões, ou seja, considera $\alpha_i = \alpha, \beta_i = \beta_i, c_i = c_i$. Logo o modelo proposto é:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-\alpha(\theta_j - \beta_i)}}. \quad (3.22)$$

Modelo Logístico - M_9 (α_i, β_i, c_i). O modelo é conhecido na bibliografia como ML_3 de Birnbaum. Também, consideram-se os parâmetros de discriminação, dificuldade e o acerto ao acaso, e um valor é estimado para cada questão, conforme: $\alpha_i = \alpha_i$, $\beta_i = \beta_i$, $c_i = c_i$. Logo o modelo proposto é:

$$P_{ij} = P(Y_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-\alpha_i(\theta_j - \beta_i)}}. \quad (3.23)$$

Os diferentes modelos descritos seguem a hierarquia conforme o número de parâmetros do modelo (df). As quantidades de parâmetros propostos na sequência dos ML possibilitaram a organização dos resultados das habilidades em estudos. Assim, o organograma proposto na Figura 6 mostra que todos os modelos são ramificações do modelo principal que é o M_9 .

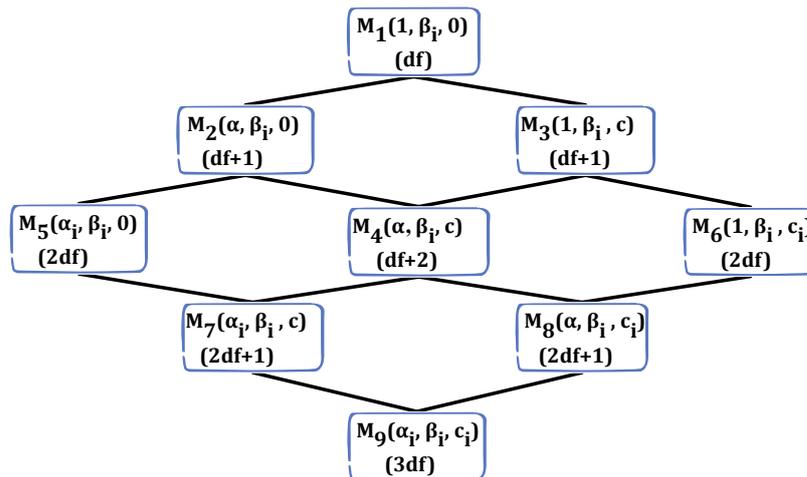


Figura 6: Diferentes modelos da TRIU de acordo com a hierarquia

A idéia do diagrama é fornecer, seguindo os ramos, sequência de modelos encaixados que podem ser testados estatisticamente pela qualidade de ajuste.

Aplica-se tal metodologia dos diferentes modelos da TRI de acordo com a hierarquia, ou seja, o organograma, no caso do conjunto de dados I, abrem-se possibilidades de análises e verificações de quais modelos detalharam e representaram o melhor ou os melhores resultados da avaliação em estudo.

Segundo Embretson e Reise (2000), os modelos utilizados na TRI requerem dois pressupostos relevantes: a Curva Característica do Item (CCI), pois há uma forma específica para cada mecanismo do processo de resposta utilizado, e a independência local ou dimensionalidade. A forma de uma curva característica do item descrever com a altera-

ção do traço latente relaciona-se à mudança na probabilidade de acerto de uma resposta específica. A independência local é obtida quando, condicionada aos níveis do traço latente, os itens do teste são independentes. Assim, a probabilidade de responder a um item é precisamente determinada pelo nível do traço latente do respondente e não por suas respostas aos outros itens do conjunto.

Entretanto, para Andrade, Tavares e Valle (2000), a independência local é vista como consequência da correta determinação da dimensionalidade dos dados. A questão da dimensionalidade de um teste/instrumento consiste no número de fatores responsáveis para expressar o traço latente e pode ser verificada por Análise Fatorial apropriada para dados categorizados. A realização deste tipo de análise permite, por exemplo, a verificação do pressuposto de que existe um fator latente dominante responsável pelo desempenho em um conjunto de itens de um teste e deve ser verificado em cada teste para que se possa utilizar qualquer um dos modelos de resposta ao item.

Nos modelos unidimensionais, a curva relacionada à probabilidade de acerto de cada item com a habilidade do indivíduo difere da CCI dos modelos multidimensionais da TRI, porque esses são representados por uma superfície, fato esse por ter mais que uma dimensão.

Para a comparação e ilustração, o mesmo item, avaliado na TRIM para o caso de um e dois traço latente, cujo primeiro modelo possui $\alpha_1 = 2,31$; $d_i = -2,56$ ou $\beta_i = 1,108$; e $c_i = 0$ e para o modelo multidimensional $\alpha_1 = 2,004$; $\alpha_2 = 1,102$; $d_i = -2,717$ e $c_i = 0$, em ambas dimensões utilizam o modelo *ML2*. Segue a Figura 7, com os valores estipulados para os parâmetros.

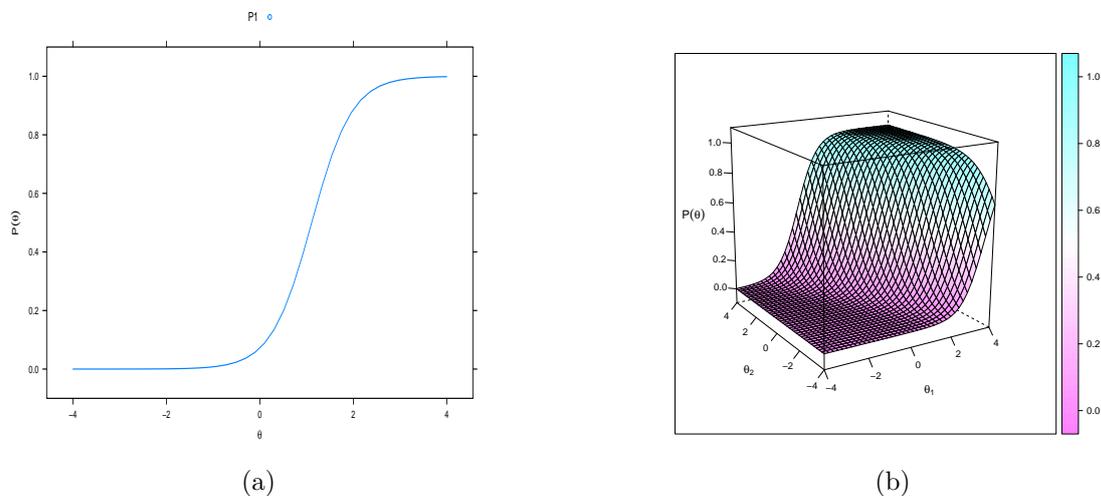


Figura 7: CCI do modelo unidimensional (a) e multidimensional (b) com o ML2 do mesmo item

Note que o parâmetro de acerto casual é a assíntota inferior da superfície, enquanto os parâmetros de discriminação representados no vetor α_1 são responsáveis pela sua inclinação em cada eixo dos traços latentes. Cada elemento do vetor indica a sensibilidade do item para avaliar cada traço latente.

Segundo Fragoso (2010), para medir o poder discriminativo do i – ésimos item para a combinação de todos os traços latentes, usa-se a estatística a Norma do Vetor Discriminações (MDISC $_i$),

$$MDISC_i = \sqrt{\sum_{k=1}^p (\alpha_{k_i}^2)}.$$

Essa estatística mede a intensidade da transição da região de baixa probabilidade de acerto ao item e para a região de alta probabilidade na superfície característica do item. Nos modelos unidimensionais, o parâmetro de dificuldade do item é medido na mesma escala do traço latente e significa o ponto de inflexão da CCI, ou seja, o ponto onde obtém uma probabilidade de acerto de

$$\frac{1 + c_i}{2}.$$

Nos modelos multidimensionais, o β_i não pode ser interpretado como o parâmetro de dificuldade da TRIU. Dessa forma, segundo Reckase (2009), para evitar confusão com os modelos, o β_i é frequentemente chamado Dificuldade (MDIFF) na literatura TRIM. O mesmo é análogo à unidimensionalidade para medir a distância de origem até o ponto de maior inclinação da superfície, dada por

$$MDIFF_i = \frac{\beta_i}{(MDISC_i)}.$$

Existem métodos clássicos usados para analisar a dimensionalidade, nomeadamente, a correlação bisserial e a correlação tetracórica, que são abordadas na análise dos conjuntos de dados propostos nesta tese, no Capítulo 4 e detalhes em D’Hainaut (1992).

Segundo Soares (2005), existem dois tipos de métodos para a análise da dimensionalidade associada a um conjunto de variáveis dicotômicas: métodos da informação restrita e os métodos de informação plena.

Nesta tese, trata-se do método de análise fatorial da informação plena Bock e Aitkin (1981); Bock, Gibbons e Muraki (1988); Muraki e Engelhard (1985), propostos a partir de uma adaptação do modelo tradicional de análise fatorial (Johnson e Wicherin, 1992), que considera a estrutura de dimensões associadas a variáveis contínuas. Contudo, o método de informação plena está, entre as técnicas da análise fatorial proposta como

ferramenta para a verificação da dimensionalidade das provas do ENEM. Para o método de informação restrita, seguem como referências para consultas, Soares (2005) e Divgi (1979), as quais não são abordadas.

Para Pasquali (2011), ao assumir a independência local dos itens, retoma-se o segundo pressuposto da TRI, que se baseia na ideia de que a resposta a um item qualquer não afeta as respostas fornecidas aos demais itens de um grupo maior.

Em palavras menos técnicas: os itens de um teste não podem apresentar pistas que permitam aos respondentes acertar outros itens, posteriormente apresentados. De acordo com essa ideia, a independência local dos itens pode ser matematicamente definida como o produto das probabilidades de acertar a cada um dos itens que compõem um teste unidimensional. Este postulado afirma que, mantidas constantes as aptidões que afetam o teste, as respostas dos sujeitos a quaisquer dos itens são estatisticamente independentes.

Assume-se a independência condicional, já que a resposta a um item para dado valor de θ é considerada prova de Bernoulli em geral. Como os itens são variáveis de Bernoulli independentes e para cada j os Y_{ij} são variáveis aleatórias independentes dado o θ_j , se e somente se, o θ_j for conhecido. Portanto a distribuição marginal de Y_{ij} é:

$$[Y_{ij}] = \int_R [Y_{ij}|\theta_j] [\theta_j] d\theta_j.$$

Sobre independência condicional no θ_j , tem-se a verossimilhança marginal do modelo TRI:

$$L(\alpha_i, \beta_i, c_i) = \prod_{j=1}^J \int \prod_{i=1}^I P_i(\theta_j)^{y_{ij}} [1 - P_i(\theta_j)]^{(1-y_{ij})} \Phi(\theta_j) d\theta_j. \quad (3.24)$$

É necessária o cálculo da verossimilhança marginal para se fazer inferência sobre os parâmetros do modelo proposto como no modelo representado pelas equações 3.7 ou 3.15, a qual é obtida após a integração dos efeitos aleatórios, neste caso, as habilidades latentes θ_j . O integrando desta verossimilhança marginal com várias questões é o produto de uma Bernoulli por uma gaussiana padrão, e não tem solução analítica, conforme em (3.17) e (3.18). Desta forma, podem-se usar métodos numéricos para integração. A aplicação da inferência sobre os parâmetros deste modelo é abordada com mais detalhes na subseção 3.2.5, e a expressão da função de verossimilhança é:

$$L(\alpha_i, \beta_i, c_i) = \prod_{j=1}^J \int_R \left\{ \prod_{i=1}^I [P_{ij}^{y_{ij}} (1 - P_{ij})^{(1-y_{ij})}] \right\} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta_j^2} d\theta_j. \quad (3.25)$$

Avaliam-se os parâmetros (α_i, β_i, c_i) e a probabilidade de uma resposta correta em determinado item, depende da proficiência (θ) do respondente e dos parâmetros que caracterizam o item. Em geral, tanto os parâmetros quanto às proficiências são desconhecidas; apenas as respostas dos indivíduos aos itens do teste são conhecidas. A estimação dos parâmetros dos modelos logísticos da TRI é feita ao se maximizar a máxima verossimilhança marginal (MVM), ver detalhes em Bock e Aitkin (1981, p. 443- 459)

Os modelos mais simples até o mais popular da TRI são aqueles que especificam um único traço latente (TRIU). Modelos unidimensionais da TRI têm sido predominantes em Ciências Sociais e pesquisa educacional, principalmente por causa das tradições históricas, mas também porque os procedimentos de estimação, de parâmetros em casos multidimensionais não foram totalmente desenvolvidos ou estudados (BAKER; KIM, 2004; RECKASE, 2009).

Enquanto os modelos unidimensionais são muitas vezes mais simples e podem ter várias propriedades de medida, interessantes e importantes, muitos constructos psicológicos são inevitavelmente um caráter multidimensional. Por exemplo, as construções não observadas podem ser entendida como uma combinação de componentes de subescala aninhadas para uma construção mais geral ou como fatores de compensação que se combinam para influenciar a probabilidade de resposta ao item.

Como o foco é o modelo multidimensional, basicamente há dois tipos de modelos que são utilizados para respostas dicotômicas: o compensatório e o não compensatório. Aborda-se para este trabalho, o modelo compensatório que são generalizações diretas do modelo logístico unidimensional, onde um valor baixo da habilidade em uma das dimensões pode ser compensado por um valor alto em outra dimensão.

Um dos principais obstáculos para se decidir e utilizar esses modelos, tem sido o fato de a estimativa dos parâmetros dos itens no espaço dimensional mais elevado (pelo aumento do número de fatores) ser computacionalmente dispendiosa para as técnicas padrão de integração numérica. No entanto, com os recentes avanços nas técnicas da estimação, juntamente com os avanços no poder computacional, a pesquisa TRIM está finalmente começando a florescer como uma metodologia de análise estatística viável (EDWARDS, 2010; RECKASE, 2009). Detalhes ver em Wirth e Edwards (2007).

3.2.4 Modelos Exploratório e Confirmatório

Na teoria de resposta ao item, é possível analisar os itens de maneira análoga à análise fatorial exploratória e confirmatória para variáveis contínuas (Mc Donald, 1999). Historicamente, segundo Andrade, Tavares e Valle (2000), os modelos da TRI começaram em um espírito de confirmação por modelar as probabilidades de resposta ao item em função de um único fator. A TRI também pode ser aplicada de um modo experimental, em que o número de dimensões não são assumidas e previamente conhecidos; neste caso, são explorados, ao invés de empiricamente estimados. Assim, Bock e Aitkin (1981), comparam-se os modelos aninhados e, em seguida, verificam-se, pela rotação da matriz, as cargas de fatores para encontrar uma estrutura mais parcimoniosa; ver detalhes em Bock, Gibbons e Muraki (1988).

A análise de itens pelo método de confirmação na TRI, pode ser adotado quando é razoável considerar-se mais de um fator que está presente nos dados, ou seja, quando suspeita-se que o modelo multidimensional é o modelo que melhor representa o conjunto de dados em uma prova, então faz-se o uso do método para verificar sua possível dimensão. Uma das primeiras abordagens é o método bifator explicitado por Gibbons e Hedeker (1992) para dados dicotômicos.

Portanto, neste trabalho, aborda-se primeiramente a análise exploratória para verificar a dimensionalidade da avaliação multidisciplinar e, em seguida, averiguar o modelo que melhor representa a prova, conforme a quantidade de parâmetros propostos. Caso a análise exploratória indique um modelo multidimensional, utiliza-se a análise confirmatória pela TRIM. Para avaliar a dimensão do traço latente, é empregado o pacote *ltm* e *mirt* para análise exploratória unidimensional e *mirt* para análise exploratória e confirmatória multidimensional. Os *softwares* mencionados estão detalhados na subsubseção 3.2.6.1. Segue na subseção 3.2.5 a estimação dos parâmetros dos modelos mais utilizados para a TRI.

3.2.5 Estimação dos Parâmetros - Inferência

Nesta subseção, são discutidos os problemas inerentes à estimação de parâmetros dos itens e dos traços latentes, θ_j , $j = 1, \dots, J$. Uma das dificuldades da TRI é a estimação dos parâmetros dos itens e das proficiências envolvidos nos modelos. O problema é contornado pelos métodos de inferência e neste estudo referenciado bibliograficamente, pelas discussões sob a abordagem por máxima verossimilhança e também, sem entrar aqui

em maiores detalhes, pela bayesiana.

Segundo Chalmers (2012), a estimação de parâmetros na TRI tem sido uma ciência progressiva ao longo desse últimos 60 anos, passando de técnicas de estimativa heurísticas até os mais avançados métodos bayesianos proposto por Baker e Kim (2004). Muitos autores, em várias décadas, desde Wringth e Panchapakesan (1969), Andrade, Tavares e Valle (2000) e Battauz (2011), analisaram formas e caminhos diferentes de contornar a estimação. Ademais, outros pesquisadores mais recentes como Edwards (2010), Sheng (2010), Chalmers (2012), lançaram, verificaram, implementaram e agilizaram computacionalmente a estimação dos parâmetros em problemas da TRI para modelos de respostas politômicas e dicotômicas.

A teoria da inferência estatística tem como objetivo principal fazer afirmações sobre características de uma população baseando-se em observações de uma amostra. Esta subseção aborda a inferência com referências na verossimilhança e modelos de regressão com efeito aleatório (BONAT et al., 2012).

A verossimilhança é uma medida de compatibilidade da amostra observada com um particular vetor de parâmetros. Desta forma, é natural definir como estimador para o vetor de parâmetros θ aquele particular vetor seja, $\hat{\theta}$, que tenha a maior compatibilidade com a amostra. Em outras palavras, o vetor que maximiza a função de verossimilhança ou compatibilidade. Não é importante apenas o valor particular assumido pela função de verossimilhança, mas também interessa para inferência os valores relativos de $L(\theta, y)$ para diferentes conjuntos de θ (BONAT et al., 2012).

O método de máxima verossimilhança fornece estimadores com propriedades desejáveis para os parâmetros desconhecidos de um modelo probabilístico. Este método é baseado na função de verossimilhança e fornece uma abordagem integrada para a obtenção de estimativas pontuais e intervalares, além da construção de testes de hipóteses (BONAT et al., 2012).

A expressão da verossimilhança $L(\theta, y)$ pode ser mais cuidadosamente definida considerando a natureza das variáveis. Para modelos discretos, como no caso da TRI não há ambiguidade e o valor da função de verossimilhança é a probabilidade do dado observado, para cada resposta de cada indivíduo (BONAT et al., 2012):

$$L(\theta) = P(\theta)[Y = y].$$

Em Bock e Lieberman (1970) introduziram o método da máxima verossimilhança marginal para a estimação dos parâmetros em duas etapas em problemas de TRI. Na

primeira etapa, estimam-se os parâmetros dos itens, portanto, assume-se certa distribuição para as proficiências. Na segunda etapa, assumem-se os parâmetros dos itens conhecidos, os quais predizem as proficiências. Apesar do avanço que esse método trouxe para a solução do problema, ele requeria que todos os parâmetros dos itens fossem estimados simultaneamente. Bock e Aitkin (1981) propuseram uma modificação no método acima e utilizaram o algoritmo Esperança e Maximização (EM) de Dempster, Laird e Rubin (1977), de modo a permitir que os itens pudessem ter seus parâmetros estimados em separado, a fim de reduzir o custo/despendido computacional.

A principal dificuldade em maximizar o resultado da expressão em 3.25 é a presença da integrais sobre o efeito aleatório, principalmente quando é suposto k -dimensional. Nessas situações, onde não é possível a solução analítica, são necessários métodos de integração numérica que podem ser usados quando são ajustados modelos de efeitos aleatórios de baixa complexidade na estrutura aleatória. O método de integração numérica usados nos pacotes *ltm* e *mirt* é o método de Gauss-Hermite e para maximizar a função quasi-Newton criado por Broyde, Fletcher, Goldfarb e Shanno (BFGS), mais detalhes nas referências em Bonat et al. (2012), Molenberghs e Verbeke (2005).

Segundo Baker e Kim (2004), as estimativas dos parâmetros de itens para os modelos unidimensionais, introduzidas por Bock e Aitkin (1981), na solução baseada na estimativa EM, foram limitadas para teste de larga escala. O algoritmo EM, por meio de Gauss-Hermite pela quadratura fixa, ou seja, pela aproximação da integral dentro da verossimilhança, apresentou solução razoável para os modelos da TRIM, sem comprometer a precisão numérica, (RECKASE, 2009). Infelizmente, esta técnica torna-se rapidamente ineficiente quando o número de dimensões aumenta, uma vez que o número de pontos de quadratura necessário também aumenta para estimar esses parâmetros. Uma solução parcial para um número moderado de dimensões foi descrita por Schilling e Bock (2005), cujos autores demonstraram que a quadratura adaptativa pode ser usada, quando um número menor de quadraturas por dimensão é utilizado, mas para o problema da elevada dimensão ainda não foi estabelecida uma solução (RECKASE, 2009).

Os modelos da TRI envolvem um grande número de parâmetros a serem estimados, tanto referentes a itens, quanto aos traços latentes dos indivíduos. Segundo Neyman e Scott (1948), o número de parâmetros referentes aos traços latentes aumenta com o tamanho da amostra. Bock e Aitkin (1981) sinalizam que esse fato gera estimadores de máxima verossimilhança, abordados nos parágrafos acima, que não necessariamente apresentam consistência e eficiência mesmo em modelos mais simples da TRI.

Como consequência, três métodos também são abordado no pacote *mirt* devido a estrutura flexível, para estimar o parâmetro de acerto ao acaso. O primeiro procedimento é estimar todos os parâmetros na máxima verossimilhança marginal.

O segundo é usar a verossimilhança perfilhada; dentro do modelo o acerto ao acaso e, em seguida, verificar o valor maximizado dentro da verossimilhança e fixar o acerto ao acaso no modelo estimado; portanto, ao se conhecer o c_i , é possível determinar os outros parâmetros na MVM.

O terceiro procedimento é utilizar o método bayesiano com uma distribuição “a priori”, considerando uma distribuição log-normal para a discriminância e uma distribuição normal para o acerto ao acaso, ambas estimadas no ML_3 na TRIU. Após serem estimados, escolhe-se somente o parâmetro do acerto ao acaso com “priori” e estimam-se novamente os modelos da TRIM pelo MVM, a fim de que se amenizem os problemas com escores nulos e perfeitos. Recomenda-se a leitura dos artigos de Patz e Junker (1999), Klein (2003), Fragoso (2010), Gonzalez (2010), Yanyan(2005), para detalhes da metodologia empregada. Seguem no Apêndice C as implementações dos três métodos de estimação dos parâmetros para os itens, com códigos do programa *R* (R Development Core Team, 2012) e desenvolvidos no pacote *mirt* (Chalmers, 2012).

Dentro desse contexto, a estimação é uma das ferramentas imprescindíveis para o modelo da TRI e sua aplicação depende da eficiência dos métodos de integração. Para o procedimento iterativo de estimação dos parâmetros, utilizaram-se o software *R* (R Development Core Team, 2012). Inicialmente, o pacote *ltm* (Rizopoulos, 2006) fornece uma estrutura flexível para analisar itens com estrutura dicotômica ou politômica, sob a abordagem de máxima verossimilhança marginal. No entanto, esse pacote têm limitações pois só pode analisar os modelos da TRIU. Assim utilizou-se o segundo pacote *mirt* (Chalmers, 2012) que estima os parâmetros para os modelos exploratórios e confirmatórios da TRIM, usando a máxima verossimilhança.

Os dois algoritmos de estimativa, segundo Chalmers (2012), para a TRIM, implementados no pacote *mirt* são:

1. o algoritmo EM de quadratura fixa, conforme Bock, Gibbons e Muraki (1988), Muraki e Carlson (1995);
2. o algoritmo Metropolis Hastings-Robbins Monro (MH-RM), segundo Cai (2010).

O algoritmo MH-RM oferece uma maneira flexível para especificar modelos de itens tanto confirmatórios e exploratórios. O MH-RM apresenta duas falhas principais:

a comparação de uma solução de EM, os dados observados para a log-verossimilhança, automaticamente, não são calculados. Recomenda-se utilizar o algoritmo MH-RM apenas quando a dimensionalidade aumenta ou quando um modelo de confirmação multidimensional é necessária, detalhes em Cai (2010), Chalmers (2012) e Reckase (2009).

As aplicações dos métodos de inferência conforme referências bibliográficas recentes apresentam solução para o problema da k -dimensionalidade e os métodos de estimação estocásticas têm sido empregados para análise exploratória e confirmatória de itens. Métodos Bayesian - Monte Carlo via Cadeias de Markov (MCMC), têm sido explorados por Edwards (2010) e Sheng (2010). Ambos os autores lançaram *software* para estimar os parâmetros para os modelos de resposta politômica e dicotômica. Estes métodos não foram implementados no pacote *mirt*, mas segue para futuras pesquisas, Baker e Kim (2004); Bock, Gibbons e Muraki (1988) que propõem técnicas de estimação dos parâmetros para TRI.

Outra abordagem relevente são os métodos de integração numérica porque podem ser usados quando se ajustam modelos de efeitos aleatórios de baixa complexidade na estrutura aleatória. O método de integração numérica que é usado no modelo k -dimensão é o método de Gauss-Hermite. Para maximizar a função, será usado quasi-Newton (BFGS); mais detalhes encontram-se em Bonat et al. (2012), Molenberghs e Verbeke (2005).

Existem outros métodos de integração como: aproximação de Laplace Robert e Casella (2010); Quase Monte Carlo usando pontos de Halton (Pan e Thompson, 2007) e Quadratura adaptativa (Cools e Haegemans, 2003), que não são trabalhados nesta pesquisa.

Na adequação, para a escolha dos modelos da TRI para o presente estudo, verifica-se o ajuste e utiliza-se medidas como estas para testes de hipóteses que são: a verossimilhança maximizada (logLik), o Critério de Informação de Akaike (AIC), o Critério Bayesiano de Schwarz (BIC) e o teste da razão da verossimilhança. O modelo é determinado por testes e critérios, que melhor representam a realidade proposta dos dados do exame multidisciplinar, sem desprezar a viabilidade existente para aferir o modelo.

Os valores da verossimilhança maximizada são observados conforme os resultados dos números de parâmetros e as diferenças dos modelos, que mostram o crescimento de como os dados se ajustam aos modelos, ou seja, quanto maior o número de parâmetros do modelo maior a combinação dos dados com o modelo pela verossimilhança maximizada.

Para Cohen, Henik e Linden (2008), a estimativa do AIC é empregada para as

análises de comparações em busca da escolha do modelo e admite-se aquele com menor valor considerando o melhor ajuste, assim

$$AIC = -2\hat{l} + 2n$$

de modo que \hat{l} é o valor da verossimilhança maximizada da amostra no ponto de máximo e n é o número de parâmetros do modelo.

Nos estudos dos modelos da TRI, também aplica-se o BIC. O critério é definido como a estatística que maximiza a probabilidade de identificar o modelo dentre os avaliados. Assim, o modelo com menor BIC é considerado o de melhor ajuste também, logo

$$BIC = -2\hat{l} + n \log(N), \quad (3.26)$$

e N é o número de indivíduos na amostra.

Os ajustes dos modelos AIC e o BIC dependem de valores para a função de verossimilhança da amostra, obtidos na convergência do algoritmo EM. Isso nem sempre ocorre, já que, frequentemente, interrompem-se o processo iterativo, quando observam-se mudanças pequenas o suficiente nos valores dos parâmetros. Porém, não necessariamente implica em mudanças insignificantes no valor da função verossimilhança (FRAGOSO, 2010).

Esta subseção mencionou alguns dos procedimentos que estão sendo utilizados para estimar os parâmetros para os modelos. Esses procedimentos são necessariamente incorporados dentro dos programas computacionais para realização de estimação. Uma metodologia de excelente propriedades de estimação pode apresentar uma má execução porque está mal programada ou vice versa.

Haja vista a estreita ligação entre a metodologia de estimação e os programas computacionais, segue na seção 3.2.6 uma visão geral projetada para dar uma estrutura conceitual para a compreensão dos procedimentos computacionais de estimação, sem entrar em detalhes técnicos, mas complementando as duas metodologias estimativas / programas.

3.2.6 Métodos Computacionais

As vantagens e o uso dos métodos computacionais em cálculo numérico crescem cada vez mais com a popularização dos computadores de alta capacidade de processamento, em todas as áreas e principalmente a de exatas, haja vista que solucionam problemas complexos. Quando as soluções manuais são impraticáveis, impossíveis ou im-

precisas, o computador é a mais ferramenta viável e resolutive. Quando a complexidade do problema impossibilita o avanço pela via analítica ou porque os cálculos numéricos são longos demais, fazem-se necessárias diversas ferramentas computacionais, a fim de que se automatizem os cálculos. Exemplo esse explorado na integração numérica via métodos de aproximação ou estatísticos de Gauss-Hermite.

Esta seção não entrará em detalhes de funcionamento de programas específicos, os quais podem mudar rapidamente, como aconteceu nas análises desenvolvidas nos conjuntos de dados. Em relação ao pacote computacional escolhido para desenvolver a tese para a TRI, pode-se afirmar que, em menos de dois meses, os resultados alteraram as análises observadas, na estimativa dos parâmetros. Assim, as análises precisaram ser refeitas. O cenário de desenvolvimento nos sistemas informáticos não é estático e muda rapidamente.

Conforme Reckase (2009), os métodos para determinar as estimativas dos parâmetros de diferentes programas são transformações simples do outro, ou seja, com o desenvolvimento e o aprimoramento dos métodos computacionais e dos métodos de estimação, são inevitáveis as mudanças em busca do crescimento e de novos modelos.

O uso do computador para manipular modelos matemáticos torna-se imprescindível, ou seja, indispensável, pois geralmente como já foi citado que é o caso da TRI, os modelos exigem algoritmos complexos para processá-los. Sendo assim, o processamento pelo computador permite investigar, analisar e justificar cenários em tempo e custo relativamente razoáveis, levando em conta que as soluções tornam-se seguras na área de exatas.

Os modelos matemáticos da TRI são utilizados para descrever a relação entre as respostas dadas pelos indivíduos aos itens em função de variáveis latentes. Na análise com a TRI, é necessário dispor de um método de integração para computar o ajuste dos dados ao modelo utilizado, este processo é chamado de estimação ou calibração que requer o processamento destes métodos e exigem a execução de cálculos numéricos complexos. Atualmente, existem poucas ferramentas computacionais para se trabalhar com a TRI, e as mais populares (Bilog SSI, 2007a), (Parscale SSI, 2007), (Multilog SSI, 2007b), (XCALIBRE Corporation, 2011) e *R* (R Development Core Team, 2012).

Nos últimos doze meses, para estimar os parâmetros em vários modelos da TRI, utilizava-se o *ltm* (Rizopoulos, 2006), ERM (Mair P.; Hatzinger, 2007), e MCMCpack (Martin, Quinn e Park, 2011) pacotes mais flexíveis. No entanto, esses pacotes têm limitações, por exemplo, o *ltm*, que só analisa modelos da TRIU de forma eficaz. Outro

exemplo são as extensões multidimensionais exploratórios disponíveis em MCMCpack que requer prévio entendimento de estimativa Bayesian para os diagnósticos de convergência e são computacionalmente intensivo, (CHALMERS, 2012). Por esse motivo Chalmers implementa o *mirt*, mas requer também cuidado, porque a última versão é recente e com atualizações constantes.

Essa pesquisa utiliza o *softwares R* (R Development Core Team, 2012) para as análises com a versão 2.15.3 e 3.0.3 (2013-03-01 e 2014-03-06), por ser dotado de linguagem de programação científicas, com bibliotecas matemáticas e estatísticas, de tal forma que é possível codificar os algoritmos necessários. Os pacotes do *R* utilizado para desenvolver as análises da TRI seguem enumerados e detalhados conforme os seus métodos de estimação.

1. O *ltm* de Rizopoulos criado em 2006, com atualização do pacote em setembro de 2012 e versão 0.9-9, até o momento, apresenta uma biblioteca com implementação dos modelos da TRI, com análise multivariada de dados dicotômicos e dados politômicos usando variáveis latentes. Inclui no pacote o modelo de Rasch, a logística de dois e três parâmetros também a resposta gradual e os modelos generalizados de crédito parcial. Por “Default”, o número de iterações no algoritmo EM é de 40; o número de iterações quasi-Newton 150 e; o número de pontos de quadratura Gauss- Hermite é de 15. Os parâmetros foram estimados pela abordagem de MVM;
2. O *mirt* de Chalmers (2012), com atualização em abril de 2014 e versão 1.3, o pacote *mirt* foi criado para análise de dados de resposta dicotômica e politômica usando modelos da TRI. O *mirt* segue a estratégia de análise fatorial, estimativa de máxima verossimilhança marginal descrito no Bock e Aitkin (1981), Bock, Gibbons e Muraki (1988), e Muraki e Carlson (1995). Modelos aninhados podem ser comparadas pelos teste do qui-quadrado, ou por AIC / BIC. Os modelos exploratórios podem ser estimados partir da quadratura ou métodos estocásticos, da análise bifatorial confirmatória enquanto os modelos confirmatórios podem ser enquadrados com o algoritmo de Metropolis - Hastings Robbins- Monro. Pode-se incluir o polinomial ou construir produto de traços latentes, além da análise de múltiplos grupos. Destina-se à análise dos modelos uni/multidimensionais para os modelos exploratórios e de confirmação usando métodos de máxima verossimilhança. O método de quadratura de Hermite-Gauss utilizado na estimativa EM tradicional, por exemplo, Bock e Aitkin (1981) é apresentado por modelos experimentais de resposta ao item, assim como para modelos de confirmação bifactor (GIBBONS; HEDEKER, 1992). Modelos exploratórios e de confirmação são estimadas por um algoritmo estocástico descrito por Cai (2010).

3. O *psych* de Revelle (2013), com última versão 1.3.10.12 e atualizada em 14 de outubro de 2013, o pacote serve para estudos da psicometria e psicologia. As funções são principalmente para construção de escala usando a análise fatorial e análise de cluster. Para a Teoria da Resposta ao Item, o pacote usa análise fatorial das correlações tetracóricas e policóricas. É possível simular funções para itens em particular, testar a estrutura do modelo e equalização. O pacote oferece resultados gráficos das análises feitas, como análise fatorial e modelos de equações estruturais Klem (2002), Thompson (2002) e Ullman (2007). A utilização desse pacote é somente para indicar, graficamente, quantas dimensões são trabalhadas, ou seja, o quanto compensa e explica a dimensão do fator, para os itens que compõem uma prova.

Alguns outros pacotes e trabalhos são citados para exemplificar os pacotes disponíveis existentes para a TRI, como *PlotIRT* (Hill Cheryl D.; Langer, 2007), especificamente para gerar gráficos da CCI e do Teste e Função de Informação do Item e do Teste. Seguem outros pacotes disponíveis como base de consulta como *irtoys* (Partchev, 2010); *Deducer* (Fellows, 2012) e *plink* (Weeks, 2010). As funções são aplicáveis para um determinado conjunto de modelos: função logística de 2 e 3 parâmetros, modelos de múltipla escolha, graduais e nominais. O pacote não lida com estimação dos parâmetros dos itens e da variável latente.

Os problemas computacionais exigem a execução iterativa de algoritmos numéricos, sendo comum a utilização de métodos matemáticos variados. Portanto, quanto mais rápida for a convergência para os valores reais, mais eficiente é o algoritmo. Por isso, os pesquisadores estão constantemente procurando desenvolver algoritmos mais otimizados. Bonat et al. (2012) implementaram o modelo logístico com três parâmetros e utilizam os comandos do *R* com princípios implementados de outros modelos de efeito aleatório. Nesta pesquisa, destacam-se estes pacotes pertencentes do *R* (R Development Core Team, 2012), por possuírem diversas técnicas estatísticas já implementadas na TRI bem como muitos outros recursos desenvolvidos para simplificar a manipulação de dados e métodos numéricos.

Para análise, comparação e abordagens da seção subsequente, utilizou-se o pacote *ltm* e *mirt*, o segundo por também estar implementado os modelos de TRIU e TRIM. Várias comparações dos programas são inevitáveis, as quais estão apresentadas e discutidas, a fim de detectar os funcionamentos diferenciais dos itens que adicionam novas possibilidades nas aplicações dos modelos da TRI na educação.

3.2.6.1 Pacotes *ltm* e *mirt*

Um dos principais objetivos da TRI é estimar os parâmetros do modelo proposto/assumido. Contudo, são necessárias técnicas de análise numérica. Para as estimativas dos parâmetros, obtidos computacionalmente, as dificuldades aumentam com complexidade do modelo. Todavia, as construções das futuras análises são inevitáveis para o caráter multidimensional. Contudo segue uma abordagem geral, para trabalhos futuros de pesquisa que como base de estudos está detalhada na bibliografia de Reckase (2009) e Chalmers (2012). Mesmo porque Reckase (2009) alerta sobre diferenças de estimativas dos parâmetros para diferentes *softwares*.

O *mirt* é um pacote novo com versões em pleno desenvolvimento, é necessário analisar o conjunto de dados proposto também nesta tese pelos pacotes *ltm* e *mirt*, as quais são realizadas no Capítulo 4. Essa comparação e as diferenças de estimativas dos parâmetros entre os pacotes que já ocorreram no conjunto de dados de Bock e Lieberman (1970), o motivo é devido às diferenças de implementação para estimar os parâmetros dos modelos.

Os programa disponíveis para TRI variam pelos procedimentos de estimação e nas formas diferentes de como abordar os modelos da TRI. Além disso, os programas diferem no número de dimensões que possam estimar e nos recursos computacionais necessário para implementar o processo de estimação. Alguns programas são muito rápidos, mas têm capacidade limitada. Outros tomam quantidades substanciais de tempo para convergir em estimativas.

Segundo Chalmers (2012), ao serem comparados os pacotes *ltm*, *TESTFAC*, *MCMCparck* com o seu pacote *mirt*, verificou-se que, para análise com um fator, o *mirt* é muito mais eficiente nas iterações, nas estimativas dos parâmetros e no tempo computacional. A questão aqui não é discutir um pacote ou outro, além de se obter essas comparações, é preciso analisar para detalhar a pesquisa a fim de que a mesma não se torne obsoleta com apenas um *softwares*. Embora ainda, não está previamente estabelecido com precisão, qual estimador é utilizado na obtenção dos melhores resultados para estimar os parâmetros, e sim tornar as comparações como ferramentas benéfica para o usuário.

Recomenda-se também que se investiguem novos *softwares* e sejam sempre atualizados. Tais atitudes são necessárias porque novos métodos de estimação poderão substituir aqueles apresentados já que cada programa tem um desempenho melhor do que o

outro para situações específicas. É difícil generalizar esses resultados com as análises de dados, onde o desenvolvimento e as atualizações são fatores recentes que ainda requerem estudos. Por esses fatores o Capítulo 4 aborda os resultados desses dois pacotes e as discussões para as devidas contribuições dessa tese, mas antes segue na subseção 3.2.7 a dimensionalidade dos dados.

3.2.7 Dimensionalidade dos dados

Na teoria de resposta ao item multidimensional, uma das utilizações mais importantes é a análise da estrutura dos dados de resposta, produto que resulta da administração do conjunto de itens de um teste com os respondentes. Segundo Reckase (2009), este tipo de análise pode ser feita pelos modelos exploratórios ou confirmatórios. Nos modelos exploratórios as análises de dados na TRIM é utilizada quando não há hipótese clara da dimensão dos dados. Para os modelos confirmatórios as análises requerem uma hipótese clara da dimensão dos dados para assegurar os resultados.

Para Reckase (2009), as medidas resultantes de ajuste do modelo aos dados são equivalente a um teste de uma hipótese sobre o número de dimensões. Como o número de dimensões das coordenadas é um componente crítico de ambas as análises exploratórias e de confirmação, esta questão é abordada em mais detalhes na subseção 3.2.7.1. Depois de uma análise do problema do número de dimensões, os procedimentos são descritos para a determinação da estrutura de dados de resposta ao item usando procedimentos exploratórios.

A determinação do número de dimensões em um conjunto de dados na TRIM está relatada na literatura, inicialmente com Holzinger e Harman (1941, p. 64-68). Nela aborda-se a análise fatorial e determina-se o número de variáveis n , necessárias para apoiar a estimativa das cargas fatoriais de m fatores independentes, com a expressão:

$$n \geq \frac{(2m + 1) + \sqrt{8m + 1}}{2}. \quad (3.27)$$

Esta expressão foi derivada ao se assumir que não há erro na estimativa de correlações. Thurstone (1947) sugeriu que o número necessários de variáveis para uma análise convincente, com fatores m deve ser “duas ou três vezes maior” do que o número de variáveis. Para uma solução com três eixos de coordenadas, é necessário um mínimo de seis variáveis, mas Thurstone recomenda 12 a 18 variáveis para produzir um resultado convincente.

Para Reise, Waller e Comrey (2000), a seleção do número de dimensões está resu-

mida em se concluir que é melhor superestimar o número de dimensões do que subestimá-los. Econometristas começaram a perceber esse problema e trataram a partir de índices de ajuste com base nas medidas de *AIC* e *BIC* (BAI; NG, 2002). É interessante que Bai e Ng (2002) vêem o problema como se determinar o “verdadeiro número de fatores”. Conforme Reckase (2009) relata, a posição aqui é que um verdadeiro número de dimensões não existe, mas que é necessário um número suficiente de dimensões para representar fielmente as principais relações dos dados de resposta ao item.

Ao especificar as dimensões dos dados na TRIM de resposta ao item de testes reais, essa interação é complexa, por serem respondentes e questões de teste. Os indivíduos variam em grande número de formas e os itens do exame exigem inúmeras habilidades, incluindo a decodificação da linguagem e outros símbolos especializados, tais como a notação por exemplos de matemática ou química. Logo, os resultados, são susceptíveis aos padrões de respostas, pois exigem um grande número de dimensões para representar com precisão as relações dos dados. Além disso, cada item de teste pode ser considerado para medir uma construção um pouco diferente e um teste de n-produto pode ser considerado como se necessitasse de n-dimensões para representar todas as relações, segundo Reckase (2009).

Em um sentido puramente matemático, tem-se $N \times n$ matriz de dados, onde N é o número de respondentes, n é o número de itens de teste, e $N > n$.

Quando o número de dimensões usadas para modelar uma matriz de resposta ao item é propositadamente minimizado para apenas identificar as principais características da matriz de dados, existe o perigo de que o significado das construções seja confuso ao invés de esclarecidas. Isso pode acontecer porque os respondentes e itens são projetados a partir de um espaço de alta dimensão para um espaço dimensional mais baixo. Isso pode acontecer porque a posição dos respondentes ou vetores de itens podem ser projetadas em cima uns dos outros quando eles realmente têm grandes distâncias entre eles no espaço dimensional superior.

Este é um fenômeno comum quando se olha para as estrelas no céu noturno. Duas estrelas podem parecer próximas uma da outra, porém, elas estão realmente distantes. A ilusão de proximidade resulta a partir das localizações das estrelas em estar na mesma linha de visão a partir da localização do observador. Suas localizações tridimensionais são projetadas em um plano bidimensional. As estrelas podem estar distantes uma das outras na terceira dimensão, embora suas projeções bidimensionais estejam próximas umas das outras, como explica Reckase (2009), nas especificações de dimensões.

Uma maneira de evitar o problema de projetar os respondentes ou itens de teste numa mesma dimensão é a utilização de mais dimensões para representar com precisão os dados. No entanto, há um preço a ser pago para a utilização de mais do que a necessidade e suficiência da dimensões. Como haverá mais parâmetros, é possível que haja aumento de imprecisão/instabilidade nas estimativas.

Outro desafio sobre as especificações do dimensionamento conforme Reckase (2009) e Reckase, Ackerman e Carlson (1988), foi quando o autor ao verificar as propriedades do modelo *mirt* compensatório que se confirmou empiricamente com os resultados da calibragem dos itens de um teste de matemática. Tal teste mostrou que, quando os itens foram selecionados, existiam questões que estavam na mesma dimensão e na mesma direção no espaço multidimensional, mas com os parâmetros com resultados diferentes, maiores detalhes podem ser encontrados em Reckase (2009).

Pelos fatores abordados nos parágrafos acima, na subsubseção 3.2.7.1 seguem questões relevantes ao se tentar determinar o número necessário de dimensões em um conjunto de dados.

3.2.7.1 Procedimentos para determinar o número necessário de dimensões na TRI

Os desafios que surgem quando se tenta determinar o número de eixos coordenados necessários para representar o comportamento em uma matriz e, neste caso, de respondentes versus itens. Essa matriz é dependente não só da sensibilidade dos itens de teste para diferenças das localizações dos respondentes no espaço paramétrico, mas também da quantidade de combinação que estão presentes na amostra dos indivíduos que respondem ao teste.

Dada a complexidade das causas de relacionamentos na matriz resposta do item, não é razoável pensar sobre o número de dimensões necessárias para modelar um teste, mas apenas no número de dimensões necessárias para modelar as relações da matriz de dados. O resultado advém da interação entre uma determinada amostra de respondentes e a amostra dos itens. Uma forma de conduzir o problema para determinar o número de dimensões requeridas a fim de modelar os dados na matriz resposta é que os itens do teste são sensíveis às diferenças de dimensões. Ou seja, se os de itens apontam na mesma direção, apesar de múltiplas habilidades cognitivas, a modelagem da matriz de dados vai exigir apenas uma dimensão e é necessária apenas uma resposta correta.

As diferenças em diferentes direções no espaço, possivelmente, existentes nos itens

e nos respondentes. O objetivo desta subseção é resumir os métodos que estão atualmente disponíveis e o que são utilizado para determinar o número de dimensões necessárias para modelar com precisão as relações em uma matriz pela TRI, detalhes dos outros/e métodos rever Reckase (2009).

O primeiro método utilizado é a análise paralela. Essa abordagem foi sugerida por Ledesma e Valero-Mora (2007) e a primeira, possivelmente, a trabalhar com a análise fatorial foi Horn (1965). O procedimento se inicia quando se utiliza uma maneira de determinar o número de fatores ou componentes em uma matriz de dados ou uma matriz de correlação dos autovalores sucessivos. Portanto, a quebra, ou seja, o decréscimo acentuada do número apropriado de componentes ou fatores a extrair sugere a dimensão avaliada.

A análise paralela é uma técnica alternativa que compara os dados observados com a de uma matriz do mesmo tamanho que o original, conhecida como dados aleatórios. Essa análise dimensional é realizada com o programa *R* com o pacote *psych*, mencionado na subseção 3.2.6, o qual fornece os primeiros n autovalores da matriz de correlações entre itens e baseia na decomposição em correlações tetracóricas/policórica. A norma proposta por Ledesma e Valero-Mora (2007) é que o número de dimensões necessárias para modelar os dados é o número de valores próprios que são maiores do que aqueles a partir dos dados aleatórios.

Portanto o primeiro procedimento na dimensionalidade da TRI é estudada com base na análise fatorial e dos componentes principais. A idéia básica da análise de componentes principais é decompor a matriz de correlação dos itens em componentes (fatores ou dimensões) e avaliar o quanto cada componente explica a estrutura correlacional dos itens. Como as respostas aos itens são dicotômicas, trabalha-se com a matriz de correlação tetracórica, detalhes Chalmers (2012), Bartholomew e Knott (1999).

Vale abordar para as análises que cada autovalor da matriz de correlação representa a porcentagem de variância explicada para cada dimensão. Assim, nos casos em que se tem um autovalor bastante superior aos demais, tem-se um indicativo de que o instrumento é unidimensional, ou, pelo menos, que há um fator dominante.

O segundo procedimento da dimensionalidade, conforme o pacote *mirt*, é observar os valores dos resultados das estatísticas diagnósticas (*AIC*, *BIC* e verossimilhança maximizada) da qualidade dos itens e respondentes. No entanto, verificam-se os valores dos ajustes dos modelos, ao realizar a TRIM para o conjunto de dados. A partir da análise dos valores, é possível indicar o quanto vale a pena mudar de uma dimensão para outra pelo ajuste dos modelos no conjunto de dados dicotômicos.

Juntamente com essa análise a quantidade de parâmetros é essencial porque quanto mais parâmetros existirem para estimar um maior número de dificuldades computacionais podem advir, portanto é preciso que as mesmas valham a pena de um modelo para outro para compensar a troca de modelo.

No pacote *mirt*, o autor considera o cálculo um pouco diferente do grau de liberdade, gl , da análise da TRIU, em que, $gl = r - p - 1$ e se trata no TRIM como o grau de liberdade, r é o número de padrões de resposta original disponível nos dados, os quais estão apenas indiretamente relacionados com o tamanho da amostra e p é a quantidade de parâmetros estimados do modelo da TRIM.

Ao considerar no pacote *mirt* o número de padrões de resposta original menos 1 é representado por DF , ou seja,

$$DF = r - 1,$$

que é o número de padrões de resposta únicos sem repetições nas respostas. Portanto, a razão é de evitar graus de liberdade totais negativos para o modelo, ou seja, não se devem estimar muito parâmetros em pequenos conjuntos de dados, principalmente quando os dados não variam muito os padrões de resposta no teste. Logo, se o conjunto de dados possuir vetores de resposta repetidas, diminui a efetividade do df total.

Além do mais é inviável que existam mais parâmetros para estimar do que a quantidade total de parâmetros. Contudo, o número de parâmetros calculados é geralmente apenas uma contagem de matriz de parâmetros livres. Exemplificando, para 32 itens devem existir em torno de 128 parâmetros estimados, ou seja, $n + (n \times m)$, onde m representa os 3 parâmetros da TRI.

Chalmers (2012) utiliza esta verificação no pacote *mirt* para precaução de todos os modelos da TRIM a fim de evitar que haja graus liberdade (gl), negativos. Um dos motivos de abordar a quantidade de parâmetros dessa forma é que se houver mais parâmetros para estimar do que graus de liberdade haverá, na estatística, modelos indefinidos.

O terceiro procedimento continua em relação à análise exploratória da TRIM e utiliza o pacote *mirt*, afim de se verifiquem os valores dos carregamentos das cargas fatoriais dos modelos, além da comunalidade e com os valores dos carregamentos das cargas.

3.2.8 Escores da TRI

A Teoria Clássica dos Testes, segundo Pasquali (2003), tem como principal objetivo analisar o escore bruto do que os itens dizem sobre o sujeito. Diante do resultado da TCT, os resultados apresentados pela dificuldade de comparar as habilidades e os conhecimentos de examinandos submetidos às provas diferentes não estão correspondendo de forma satisfatória. As discussões teóricas sobre a viabilidade de se comparar as habilidades com os conhecimentos de examinandos, levou Lord (1968) a defender as vantagens teóricas dos modelos da teoria da resposta ao item em procedimentos que estabelecem a equivalência dos resultados de provas. Para maiores detalhes, recomenda-se a leitura de artigos Lord (1968), Beaton e Allen (1992), Andrade (2000) e Valle (2001).

Segundo Andrade (2000), os valores da escala do traço latente (ou habilidade) podem assumir teoricamente qualquer valor real entre $-\infty$ e $+\infty$, diferentemente da TCT, cuja escala geralmente varia entre 0 e a quantidade total de questões do teste. No entanto, o importante nessa escala não é a sua magnitude, mas as relações de ordens existentes. Dessa forma, é preciso estabelecer valores para a média e para o desvio padrão que representem a escala de habilidade dos indivíduos na população.

A calibração dos itens, devido à facilidade computacional, geralmente é feita na escala (0,1), ou seja, numa escala com média igual a zero e desvio padrão igual a 1. Após essa etapa, é realizada a construção da escala do traço latente, que é colocada na mesma escala métrica dos itens. Esse procedimento, normalmente utilizado nas avaliações para aplicar a escala métrica para a TRI, possui referências de Valle (2001), Beaton e Allen (1992), que detalharam a construção e a interpretação para escala do conhecimento.

Foram utilizados dois procedimentos metodológicos para determinar e comparar os escores do TCT com a TRI para os conjuntos de dados estudados nesta pesquisa. O primeiro de posicionamento, ou seja, conforme os escores obtidos pelos modelo indicados da TRI, ordenaram-se tanto as notas da TCT como os escores da TRI, das habilidades menores para as maiores, os quais foram enumerados conforme os resultados. Os escores são estimados pela TRI utilizando o método da máxima verossimilhança. Também foi atribuído, a cada pessoa, o escore que melhor identifica seu posicionamento na escala do traço latente.

A segunda abordagem usa uma transformação conveniente e interpretáveis para os escores dos modelos, em uma escala métrica entre 0 a 1, dada a probabilidade do aluno zerar ou gabaritar a prova, que neste caso considera-se:

- θ_{kj} : variável k -dimensão do traço latente (habilidade) do individuo j ;
- EP_{kj} : escala padronizada de k -dimensão do indivíduo j ;
- EI_k : escore inferior k -dimensão, escore de todos os itens errados (0);
- ES_k : escore superior k -dimensão, escore de todos os itens corretos (1).

Para determinar a ponderação para a transformação das habilidades dos alunos na escala padronizada (EP) ou unitária, considera-se que:

$$EP_{kj} = \frac{(\theta_{kj}) - EI_k}{ES_k - EI_k}. \quad (3.28)$$

A escala padronizado na escala métrica entre 0 a 1 para a TRI está detalhada no Apêndice D, em código em R .

Para a abordagem dos escores da TRIM com mais de um traço latente para cada respondente, a habilidade avaliada tem um escore que é padronizado. A mesma possibilita analisar os resultados de como cada respondente respondeu à prova e qual foi a habilidade que sobressaiu. Outra abordagem seria calcular uma média ponderada e obter uma escala padronizada multidimensional que representa as habilidades avaliadas. Leva-se em consideração que, ao se elevar o escore ao quadrado, expressa-se a variância explicada pelo fator. Isso implica que o peso é proporcional à variância explicada pelo fator, em relação à variância total. Então, considera-se que:

- EPM_j : escala padronizado multidimensional dada a habilidade do indivíduo j ;
- SF_{ik} : soma dos escores ao quadrado dos itens é de k -dimensão .

Ao se generalizar a escala padronizada multidimensional com ponderação, tem-se:

$$EPM_j = \sum_{j=1}^k \left[\left(\frac{SF_{ik}}{\sum_{k=1}^n SF_{ik}} \right) \cdot EP_{kj} \right]. \quad (3.29)$$

Os procedimentos adotados objetivam avaliar o desempenho dos respondentes nos exames multidisciplinares, com a perspectiva de fornecer indicadores qualitativos e quantitativos para as instituições de avaliação e as possibilidades de como abordar a TRI, com as habilidades apresentadas que corresponda à escala de conhecimento. No Capítulo 4 seguem as discussões e resultados das metodologias adotadas.

4 RESULTADOS E DISCUSSÕES

Neste Capítulo o principal objetivo é expor os resultados de aplicações das metodologias e as investigações propostas nos capítulos anteriores. Apresenta-se os resultados na perspectiva de desvendar e entender uma prova multidisciplinar pela TRI, além das possíveis vantagens e/ou desvantagens quanto à análise dos exames multidimensional com a informação prévia de como o comportamento dos traços latentes de uma determinada população podem ser representados por um modelo logístico.

A partir dos resultados desta pesquisa, é possível apresentar novas contribuições baseadas nas análises do exame multidisciplinar e no dimensionamento das avaliações, propostas pelos modelos da TRI nesta tese, as quais são concluídas no Capítulo 5. É importante ressaltar que o comportamento das habilidades surge como um elemento necessário ao desenvolvimento da implementação do processo de estimação dos parâmetros de um modelo da TRI.

Nas próximas seções, os resultados das análises apresentarão as diferenças quanto à dimensionalidade e o modelo indicado, ou seja, quanto à quantidade e/ou intensidade de traços latentes exigidos dos respondentes para se obter uma determinada resposta ao item (e conjunto de itens).

Cada seção do Capítulo 4, incluídas as subseções, tem o intuito de se familiarizar com a execução do processo de estimação e interpretação dos modelos de resposta ao item com o uso de distribuições assimétricas para os traços latentes. Como exemplo, o conjunto de dados I, em que a aplicação é de um teste constituído também por itens dos tipos unidimensional e multidimensional. Tais itens são analisados sob o modelo k -dimensional propostos na subseção 3.2.3, Figura 6.

Em seguida, assume-se qual modelo que melhor descreve a avaliação multidisciplinar, com o ajuste baseados em critérios de convergência estabelecidos. Em cada seção do Capítulo 4 são abordados testes multidisciplinares de maneira análoga à primeira análise de dados. Os modelos aplicados são avaliados e comparados nas diversas situações para

investigar a distribuição dos traços latentes e assim ajustar o modelo da TRI.

4.1 MODELO LOGÍSTICO DA TRI NO CONJUNTO DE DADOS I

Este estudo fornece uma visão geral dos modelos da TRI e as implicações da avaliação, composta por um conjunto de dados de 94 alunos da graduação, os quais realizaram exame multidisciplinar composto por 30 itens - uma avaliação interna da PUCPR - para verificar o desempenho e a análise descritiva. Primeiro, é verificado, na subseção 4.1.1, a possível dimensionalidade deste exame multidisciplinar e conforme a importância dos assuntos das análises da TRI, a necessidade de subdividi-los em subseções dando continuidade e ordenando os estudos.

Do ponto de vista da análise estatística, o conjunto de dados I é desafiador para a aplicação de modelos da TRI, e também, foi esse conjunto de dados que gerou todos os questionamentos de avaliação multidisciplinar.

A base além de ser pequena - apenas 94 respondentes - ainda apresenta uma quantidade grande de respondentes sem nenhum acerto, no entanto, foi possível considerar bons ajustes de modelos da TRI, mantendo o conjunto em estudo motivou a investigar, elaborar e verificar por:

- ilustrar e pensar nos procedimentos;
- permitir que resultados e inconsistência fossem investigados;
- apontar um numero sem razão aparente ou conhecida;
- investigar passos e características dos procedimentos de inferência;
- apresentar, mesmo com numero pequeno de indivíduos, resultados coerentes e;
- motivar as análises multidimensionais.

A prova multidisciplinar, mesmo que pelas análises preliminares com a TRI, ilustre um conjunto de dados pequeno, apresentou uma curva característica dos itens que modelou a probabilidade de acerto de um item condicionada à habilidade, o conhecimento do examinado. Os resultados, individualmente, tanto pelas habilidades específicas quanto as características de zeros, apresentaram inferências e resultados de coerência, logo, a TRI foi condizentes com o CD1.

4.1.1 Dimensionamento do conjunto de dados I

A dimensionalidade da TRI no CD1 primeiramente é estudada com base na análise paralela da Análise Fatorial (FA) e dos Componentes Principais (PC). Seguindo a metodologia das subseções 3.2.6 e 3.2.7, primeiro aborda-se a dimensão no exame multidisciplinar, considerando a prova com todas as 30 questões, para verificar como é tratada a dimensão em toda a avaliação, pois espera-se, em um exame multidisciplinar, que exista mais de um fator dominante que justifique mais de uma medida de proficiência para a prova. Portanto, observando-se o gráfico da Figura 8, os autovalores, superiores aos demais, apresentam um indicativo de que o instrumento é unidimensional, ou, pelo menos, que há um fator dominante que destacou tanto na FA como no PC.

O gráfico da Figura 8 do estudo da dimensionalidade foi obtido pelo pacote *psych*, a partir da análise de componentes principais sobre a matriz de correlação tetracórica da análise paralela. O estudo da dimensionalidade está diretamente relacionado à validade do instrumento, assim, estuda-se a capacidade do instrumento em produzir dimensões interpretáveis, a fim de verificar algumas relações estruturais nos itens da prova.

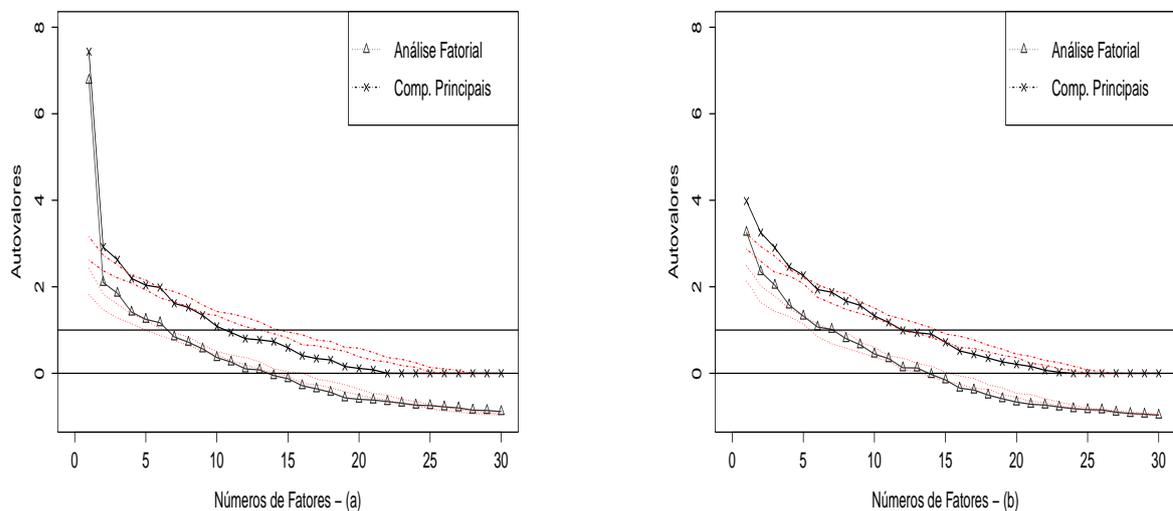


Figura 8: Diagnóstico da dimensão com: (a) 94 respondentes e (b) 75 respondentes.

Na Figura 8, observa-se que para os 94 respondentes, 6 fatores dos componentes principais (PC) acima da linha de corte, ou seja, a linha tracejada, são valores simulados dos PCs, em que fatores acima dessa linha representam algum fator explicativo e abaixo fatores que não são interpretados na dimensão. Para os PCs, dos 6 fatores que explicam a dimensão, há dois que explicam melhor os dados, porque os outros quatro aproximam-

se nas distâncias e deixam de ser representativos. A mesma verificação acontece para análise fatorial (FA), da Figura 8, onde há 9 fatores que estão acima da linha de corte, mas somente dois distanciam entre si.

De acordo com a Figura 8, para os 75 respondentes, 5 fatores estão acima da componente principal e 9, na análise fatorial. Assim, ao retirar os alunos que zeraram na prova, os fatores da análise paralela se aproximam e é possível indicar que há, não somente uma dimensão dominante, mas ganhos menores de dimensões, tornando-se necessário explorar a TRIM.

Formalmente, a prova não é unidimensional, já que ela é multidisciplinar, mas há um fator dominante, como observado na Figura 8, nos 94 respondentes. Pode-se, assim dizer, que a prova é unidimensional, o suficiente para a validade da aplicação da teoria de resposta ao item para os 94 e 75 respondentes. Vale ressaltar que há indicação do segundo fator para os respondentes.

De acordo com a análise, pode-se inferir que é possível usar até duas dimensões para os 94 respondentes, porque o terceiro fator, pela distância das cargas em relação ao segundo, não explica suficientemente a terceira dimensão, já que os fatores estão próximos tanto nos componentes principais como nas análises fatoriais. Contudo, ao serem retirados os 19 respondentes que zeraram, pode-se determinar que não há ganho ao se passar de 1 para 2 fatores, uma vez que as distâncias entres eles são menores do que a observada para os 94 indivíduos.

Do segundo para o terceiro fator, a distância é menor ainda e assim sucessivamente, mas como a prova é multidisciplinar, a segunda dimensão é abordada para verificar se alguma questão é explicada em outra dimensão, ou seja, se o respondente utilizou duas habilidades para responder as questões.

Para abordar os modelos TRIM, a quantidade de respondentes é maior do que itens, ou seja, $94/75 > 30$, basta agora verificar, no decorrer das subseções, se os parâmetros estimados são menores do que os padrões de resposta para garantir que não exista graus de liberdade totais negativos. Portanto, para iniciar as análises de uma e duas dimensões, conforme a metodologia do segundo procedimentos da subseção 3.2.7, é necessário abordar os modelos das dimensões indicadas pela análise paralela e essa análise de dimensão somente finaliza quando forem desenvolvidas todas as análises da TRIU na subseção 4.1.2 e em seguida, com o modelo exploratório que são detalhados na subseção 4.1.3.

4.1.2 TRIU no conjunto de dados I

Um pressuposto básico indicado na análise paralela é o de que a proficiência na área da prova possa ser representada por um único traço latente capaz de captar toda a estrutura correlacional entre os itens, a unidimensionalidade do instrumento.

Há necessidade de analisar a pesquisa em 4 (quatro) etapas, a fim de que se verifiquem as influências dos respondentes que zeraram a prova e também dos modelos da TRI.

A primeira e a segunda discussões referem-se aos 94 respondentes com 30 questões/habilidade geral, e as 5 questões/seis habilidades; a terceira e quarta abordagem de análise com 75 respondentes, com seis habilidades e com a habilidade geral.

Utilizam-se habilidades por ser uma prova multidisciplinar composta pela habilidade geral/ou o exame multidisciplinar na qual foi elaborada e constituída por 6 (seis) disciplinas/habilidades.

Para a escolha do modelo e para buscar o modelo geral, que descrevesse a habilidade da TRIU e o ajuste observado, foram analisadas e comparadas as influências e as possibilidades das habilidades verificadas no conjunto de dados, divididas em:

- habilidade geral com 94 respondentes;
- seis habilidades com 94 respondentes;
- habilidade geral com 75 respondentes; e
- seis habilidades com 75 respondentes.

Da mesma forma que a parte descritiva, o ajuste dos modelos TRIU é conduzido por disciplina e com a prova como um todo, para também verificar as influências do zeros. Como o CD1 foi a origem dos estudos pelo pacote *ltm* de análises unidimensionais foi possível a flexibilidade dos modelos com os parâmetros dos itens.

A metodologia descrita na subseção 3.2.2 possibilita a descrição dos modelos conforme a hierarquia. Para as habilidades em estudo, 9 (nove) modelos logísticos foram aplicados para verificar, comparar e analisar os resultados dos parâmetros discriminação, dificuldade e acerto ao acaso (α_i, β_i, c_i) , respeitando os modelos conforme a Figura 6.

Os modelos da hierarquia foram compostos para comparar e discutir os resultados do conjunto de dados, cuja origem está na habilidade geral, porém, em relação à análise exploratória da subseção 2.1.1, as possibilidades observadas no conjunto de dados

ampliaram tais análises. Então, para a investigação das influências inflacionadas em zero no modelo, foi necessário verificar as habilidades observadas conforme os resultados dos dados.

Conforme a metodologia da TRIU o modelo logístico é descrito por:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-\alpha_i(\theta_j - \beta_i)}}. \quad (4.1)$$

A Tabela 10 foi composta a partir do ajuste dos 09 modelos conforme os parâmetros de discriminância, dificuldade, e acerto ao acaso (α, β, c) constituído por: $M_1(1, \beta_i, 0)$; $M_2(\alpha, \beta_i, 0)$; $M_3(1, \beta_i, c)$; $M_4(\alpha, \beta_i, c)$; $M_5(\alpha_i, \beta_i, 0)$; $M_6(1, \beta_i, c_i)$; $M_7(\alpha_i, \beta_i, c)$; $M_8(\alpha, \beta_i, c_i)$ e $M_9(\alpha_i, \beta_i, c_i)$.

As habilidades dos alunos (HA) foram obtidas conforme os modelos ajustados e estimados nos referidos casos de 94 e 75 respondentes, sendo verificadas as seis habilidades (GAA, CDI, FGE, QGI, EST e IEQ) e a habilidade geral.

Para a comparação, de forma geral, busca-se o modelo com menor número de parâmetros, por ser de mais simples interpretação e estimação. Como recomendação geral, quanto menores o AIC e o BIC, melhor é o ajuste. A Tabela 10 apresenta as medidas de comparação para todos os modelos ajustados para a habilidade geral dos 94 respondentes e dos 75 respondentes.

Tabela 10: Resultados dos ajustes de diferentes modelos da TRIU considerado para a prova

HABILIDADES DOS ALUNOS (HA)	MODELOS (α, β, c)								
	M_1 $(1, \beta_i, 0)$	M_2 $(\alpha, \beta_i, 0)$	M_3 $(1, \beta_i, c)$	M_4 (α, β_i, c)	M_5 $(\alpha_i, \beta_i, 0)$	M_6 $(1, \beta_i, c_i)$	M_7 (α_i, β_i, c)	M_8 (α, β_i, c_i)	M_9 (α_i, β_i, c_i)
GERAL (94 alunos)									
parâmetros	df (30)	df (31)	df (31)	df (32)	df (60)	df (60)	df (61)	df (61)	df (90)
logLik	-1464,79	-1461,20	-1464,79	-1461,20	-1443,22	-1463,82	-1443,22	-1453,27	-1443,22
AIC	2989,57	2984,41	2989,57	2984,41	3006,44	3047,64	3006,44	3028,55	3066,44
BIC	3063,87	3063,25	3065,87	3063,25	3159,04	3200,24	3159,04	3183,69	3295,34
GERAL (75 alunos)									
parâmetros	df (30)	df (31)	df (31)	df (32)	df (60)	df (60)	df (61)	df (61)	df (90)
logLik	-1383,61	-1364,72	-1383,61	-1362,82	-1338,75	-1341,34	-1338,75	-1342,26	-1324,31
AIC	2827,21	2791,45	2827,21	2787,63	2797,49	2802,68	2797,49	2806,52	2828,61
BIC	2896,75	2863,29	2896,74	2859,48	2936,54	2941,73	2936,54	2947,88	3037,19

A Tabela 10 da habilidade geral continua na Tabela 11 das seis habilidades e ambas apresentam as medidas de comparações dos modelos ajustados.

Para os resultados da Tabela 10 e 11 utilizou-se o procedimento iterativo de estimação do parâmetro, abordando a máxima verossimilhança marginal e a integração pelo método de Gauss - Hermite com a quadratura verificada nos pontos para os nove modelos da TRI. Utiliza-se também o método de interação Quasi-Newton (BFGS) com

Tabela 11: Resultados dos ajustes de diferentes modelos da TRIU por áreas

HABILIDADES DOS ALUNOS (HA)	MODELOS (α, β, c)									
	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	
	(1, $\beta_i, 0$)	($\alpha, \beta_i, 0$)	(1, β_i, c)	(α, β_i, c)	($\alpha_i, \beta_i, 0$)	(1, β_i, c_i)	(α_i, β_i, c)	(α, β_i, c_i)	(α_i, β_i, c_i)	
6 HABILIDADES (94 alunos)										
parâmetros	df (5)	df (6)	df (6)	df (7)	df (10)	df (10)	df (11)	df (11)	df (15)	
HA. QGI logLik	-240,51	-238,44	-240,51	-238,44	-237,36	-240,51	-237,36	-237,69	-239,94	
AIC	491,01	488,89	491,01	488,89	494,72	501,02	494,72	497,38	509,88	
BIC	503,73	504,15	503,73	504,15	520,15	526,46	520,15	525,35	548,03	
HA. IEQ logLik	-278,36	-278,33	-278,36	-278,33	-276,66	-276,65	-276,66	-276,90	-276,66	
AIC	566,72	568,66	566,72	568,66	573,33	575,3	573,33	575,81	583,33	
BIC	579,44	583,92	579,44	583,92	598,76	600,73	598,76	603,79	621,48	
HA. GAA logLik	-223,32	-223,14	-223,32	-220,91	-218,92	-220,61	-218,02	-219,45	-218,02	
AIC	456,64	458,28	456,64	453,82	456,04	461,22	456,04	460,91	466,03	
BIC	469,36	473,54	469,36	469,08	481,48	486,66	488,88	469,36	504,18	
HA. CDI logLik	-276,87	-276,53	-276,87	-276,53	-275,26	-275,44	-275,26	-276,37	-275,26	
AIC	563,74	565,06	563,74	565,06	570,51	570,89	570,51	574,74	580,51	
BIC	576,46	580,32	576,46	580,32	595,95	596,33	595,95	602,72	618,66	
HA. FGE logLik	-248,02	-247,70	-248,02	-247,70	-245,52	-247,02	-245,53	-246,12	-245,39	
AIC	506,04	507,4	506,04	507,4	511,03	514,03	511,06	514,24	520,78	
BIC	518,76	522,66	518,76	522,66	536,47	539,47	536,5	542,21	558,93	
HA. EST logLik	-286,70	-286,60	-286,70	-286,67	-284,32	-285,53	-284,32	-285,32	-284,38	
AIC	583,4	585,21	583,4	585,34	588,64	591,07	588,64	592,64	598,77	
BIC	596,12	600,47	596,12	600,6	614,07	616,51	614,07	620,62	636,92	
6 HABILIDADES (75 alunos)										
HA. QGI logLik	-215,12	-215,10	-215,13	-214,08	-213,46	-214,45	-212,98	-212,18	-209,97	
AIC	440,25	442,2	440,25	440,15	446,92	448,91	445,96	446,36	449,94	
BIC	451,84	456,11	451,84	454,06	470,1	472,09	469,14	471,85	484,7	
HA. IEQ logLik	-243,92	-240,07	-243,92	-239,86	-238,30	-238,70	-237,99	-237,52	-237,99	
AIC	497,84	492,14	497,84	491,71	496,61	497,4	495,98	497,03	505,98	
BIC	509,43	506,05	509,43	505,62	519,78	520,58	519,15	522,53	540,74	
HA. GAA logLik	-199,80	-196,56	-199,80	-194,94	-191,87	-194,31	-191,87	-194,09	-191,81	
AIC	409,59	405,12	409,59	401,89	403,74	408,62	403,75	410,19	413,63	
BIC	421,18	419,03	421,18	415,79	426,91	431,8	426,92	435,68	448,39	
HA. CDI logLik	-246,48	-245,04	-246,48	-245,04	-246,07	-243,65	-242,07	-242,91	-242,07	
AIC	502,97	502,08	502,97	502,08	504,13	507,31	504,13	507,81	514,14	
BIC	514,56	515,99	514,56	515,99	527,31	530,49	527,31	533,31	548,9	
HA. FGE logLik	-223,76	-219,55	-223,76	-219,55	-215,91	-218,06	-215,92	-222,03	-215,65	
AIC	457,51	451,1	457,51	457,1	451,82	456,11	451,83	466,06	461,3	
BIC	469,1	465,01	469,1	465,01	475	479,29	475,01	491,55	496,07	
HA. EST loglik	-255,46	-249,89	-250,53	-249,32	-245,91	-247,92	-245,99	-247,60	-244,10	
AIC	520,92	511,8	511,07	510,64	511,87	515,86	511,99	517,2	518,2	
BIC	532,51	525,7	522,66	524,55	535	539,03	535,16	542,7	552,97	

Nota: Número de parâmetros (df)

máxima integração.

Considerou-se como exemplo a habilidade geral para escolha do modelo. Ao serem analisados os valores da verossimilhança marginalizada conforme o número de parâmetros, verifica-se que, com o aumento da complexidade do modelo (mais parâmetros sendo estimados), o valor da $\log Lik$ também cresce. Isso indica que o procedimento computacional está coerente.

Observa-se na Tabela 10 que os modelos $M_5 (\alpha_i, \beta_i, 0)$; $M_7 (\alpha_i, \beta_i, c)$ e $M_9 (\alpha_i, \beta_i, c_i)$ foram os que obtiveram os melhores ajustes, ressaltando-se que os valores foram iguais a $-1443,22$. O motivo de os modelos possuírem o mesmo valor é porque o valor do acerto casual (c) foi estimado igual a zero para ambos.

Ao se considerarem os critérios AIC e BIC, observa-se que os modelos com menor AIC e BIC são o M_2 e o M_4 . Porém, este último apresenta a mesma $\log Lik$ que o M_2 , uma vez que o valor do parâmetro c foi estimado em zero, recaindo assim ao M_2 . Desta forma, opta-se por escolher o modelo M_2 , já que ele apresenta um ajuste não diferente do modelo mais complexo M_9 , porém, com 59 parâmetros a menos.

Os mesmos critérios foram aplicados para a escolha dos modelos das habilidades específicas da Tabela 11. Os modelos escolhidos foram: para a habilidade QGI o $M_2 (\alpha, \beta_i, 0)$, para IEQ, CDI, FGE e EST o $M_1 (1, \beta_i, 0)$, e para a GAA, optou-se pelo $M_4 (\alpha, \beta_i, c)$.

O modelo para a habilidade geral com os 75 respondentes é o $M_4 (\alpha, \beta_i, c)$ com a quantidade de 32 parâmetros. Os mesmos critérios propostos foram aplicados nas 6 habilidades para a escolha do modelo, desta maneira foram obtidos: na habilidade QGI e H.CDI o $M_1 (1, \beta_i, 0)$; na H.IEQ, na H.GAA e na H.EST o $M_4 (\alpha, \beta_i, c)$ e somente na H.FGE o M_2 .

Verificou-se que os modelos propostos para as habilidades observadas e baseados na verossimilhança tiveram, conforme a hierarquia, bons ajustes nos dados, menos para habilidade com 94 da IEQ do M_8 para o M_9 , levando-se, desta forma, ao questionamento se foram somente erros numéricos.

Nesta subseção, mostrou-se o procedimento que indicou a escolha de um modelo, empregando a metodologia da subseção 3.2.7.1 que foi utilizada, quando analisou-se o AIC, BIC e verossimilhança maximizada, levando-se em consideração a quantidade de parâmetros.

4.1.2.1 Estimação dos parâmetros nos modelos da TRIU no conjunto de dados I

Para obter as estimativas dos parâmetros, é necessário indicar qual o modelo viável para o CD1. A Tabela 12 representa um quadro geral das habilidades dos 94 e 75 respondentes do modelo escolhido sobre os critérios e resultados da Tabela 10 e detalha quais modelos são os indicados.

Tabela 12: Modelos da TRI escolhidos conforme as habilidades

Modelos	df	$M_1(1, \beta_i, 0)$	$M_2(\alpha, \beta_i, 0)$	$M_4(\alpha, \beta_i, c)$
94 respondentes				
H. GERAL	31	3 ^o	1 ^o	2 ^o
H.QGI	6	3 ^o	1 ^o	2 ^o
H.IEQ	5	1 ^o	2 ^o	3 ^o
H.GAA	7	2 ^o	3 ^o	1 ^o
H.CDI	5	1 ^o	2 ^o	3 ^o
H.FGE	5	1 ^o	2 ^o	3 ^o
H.EST	5	1 ^o	2 ^o	3 ^o
75 respondentes				
H. GERAL	32	3 ^o	2 ^o	1 ^o
H.QGI	5	1 ^o	3 ^o	2 ^o
H.IEQ	7	3 ^o	2 ^o	1 ^o
H.GAA	7	3 ^o	2 ^o	1 ^o
H.CDI	5	1 ^o	2 ^o	3 ^o
H.FGE	6	2 ^o	1 ^o	3 ^o
H.EST	6	3 ^o	2 ^o	1 ^o

Nota: Número de parâmetros (df)

A Tabela 12 mostra as escolhas dos modelos para as habilidades sugeridas. Observou-se que para o traço latente dos modelos, não passou de $M_4(\alpha, \beta_i, c)$, em que a dificuldade e a discriminação foram estimadas juntamente com o acerto casual, o qual é apenas um valor comum e representativo de todas as questões.

Na Tabela 13 foram verificados os valores estimados dos parâmetros, conforme os modelos propostos para as habilidades, foram considerados tanto para os 94 como para os 75 respondentes, a habilidade geral e as seis disciplinas analisadas separadamente.

Na Tabela 13, para os 94 respondentes, a habilidade geral para o $M_2(\alpha, \beta_i, 0)$ mostrou que o modelo estimou o valor de 1, 4, ou seja, um único valor para a discriminação entre as questões. Ao se comparar o resultado da discriminação das seis disciplinas, esse valor está entre 1 a 2, 103 dos modelos escolhidos para cada estimação.

Verificou-se que o valor estimado foi de 0,7 pelo M_4 ao comparar a discriminação dos 75 respondentes da habilidade geral com os resultados de todas as outras habilida-

Tabela 13: Parâmetros estimados da TRI para a prova geral e por áreas

H. GERAL (94)	parâmetros (α_1) (β_1) (c_1)			6 Habilidades (94) Modelos	parâmetros (α_1) (β_1) (c_1)			H. GERAL (75) Modelos	parâmetros (α_1) (β_1) (c_1)			6 Habilidades Modelos	parâmetros (α_1) (β_1) (c_1)		
Questões	$M_2(\alpha, \beta_i, 0)$			H.QGI	$M_2(\alpha, \beta_i, 0)$			Questões	$M_4(\alpha, \beta_i, c)$			H.QGI	$M_1(1, \beta_i, 0)$		
Q1	1,4	0,3	0	Q1	1,5	0,2	0	Q1	0,7	0,2	0,2	Q1	1	-0,2	0
Q2	1,4	1,3	0	Q2	1,5	1,3	0	Q2	0,7	3,1	0,2	Q2	1	1,3	0
Q3	1,4	1,4	0	Q3	1,5	1,4	0	Q3	0,7	3,1	0,2	Q3	1	1,5	0
Q4	1,4	1,4	0	Q4	1,5	1,5	0	Q4	0,7	3,4	0,2	Q4	1	1,6	0
Q5	1,4	0,8	0	Q5	1,5	0,8	0	Q5	0,7	1,6	0,2	Q5	1	0,6	0
				H.IEQ	$M_1(1, \beta_i, 0)$							H.IEQ	$M_4(\alpha, \beta_i, c)$		
Q6	1,4	1,4	0	Q1	1	1,8	0	Q6	0,7	4,2	0,2	Q1	0,8	4,9	0,2
Q7	1,4	0,4	0	Q2	1	0,5	0	Q7	0,7	0,7	0,2	Q2	0,8	0,8	0,2
Q8	1,4	1,1	0	Q3	1	1,4	0	Q8	0,7	2,3	0,2	Q3	0,8	2,6	0,2
Q9	1,4	0,1	0	Q4	1	0,1	0	Q9	0,7	-0,3	0,2	Q4	0,8	-0,1	0,2
Q10	1,4	0,4	0	Q5	1	0,5	0	Q10	0,7	0,6	0,2	Q5	0,8	0,8	0,2
				H.GAA	$M_4(\alpha, \beta_i, c)$							H.GAA	$M_4(\alpha, \beta_i, c)$		
Q11	1,4	0,2	0	Q1	2,1	0,3	0,1	Q11	0,7	0,1	0,2	Q1	2,5	-0,1	0,2
Q12	1,4	2,0	0	Q2	2,1	6,7	0,1	Q12	0,7	19,7	0,2	Q2	2,5	6,1	0,2
Q13	1,4	1,5	0	Q3	2,1	2,2	0,1	Q13	0,7	5,0	0,2	Q3	2,5	2,5	0,2
Q14	1,4	1,0	0	Q4	2,1	1,1	0,1	Q14	0,7	2,1	0,2	Q4	2,5	1,0	0,2
Q15	1,4	1,7	0	Q5	2,1	2,4	0,1	Q15	0,7	6,6	0,2	Q5	2,5	5,7	0,2
				H.CDI	$M_1(1, \beta_i, c)$							H.CDI	$M_1(1, \beta_i, 0)$		
Q16	1,4	0,8	0	Q1	1	1,1	0	Q16	0,7	1,9	0,2	Q1	1	0,6	0
Q17	1,4	1,0	0	Q2	1	1,2	0	Q17	0,7	2,0	0,2	Q2	1	0,8	0
Q18	1,4	0,5	0	Q3	1	0,6	0	Q18	0,7	0,9	0,2	Q3	1	0,2	0
Q19	1,4	0,4	0	Q4	1	0,5	0	Q19	0,7	0,5	0,2	Q4	1	-0,0	0
Q20	1,4	1,1	0	Q5	1	1,4	0	Q20	0,7	2,4	0,2	Q5	1	1,0	0
				H.FGE	$M_1(1, \beta_i, 0)$							H.FGE	$M_1(1, \beta_i, 0)$		
Q21	1,4	1,6	0	Q1	1	2,1	0	Q21	0,7	6,5	0,2	Q1	1	1,7	0
Q22	1,4	1,1	0	Q2	1	1,4	0	Q22	0,7	2,7	0,2	Q2	1	1,0	0
Q23	1,4	0,4	0	Q3	1	0,5	0	Q23	0,7	0,6	0,2	Q3	1	0,0	0
Q24	1,4	0,8	0	Q4	1	1,0	0	Q24	0,7	1,6	0,2	Q4	1	0,5	0
Q25	1,4	1,6	0	Q5	1	2,0	0	Q25	0,7	4,2	0,2	Q5	1	1,7	0
				H.EST	$M_1(1, \beta_i, 0)$							H.EST	$M_4(\alpha, \beta_i, c)$		
Q26	1,4	1,0	0	Q1	1	1,2	0	Q26	0,7	2,0	0,2	Q1	1,6	2,7	0,3
Q27	1,4	0,5	0	Q2	1	0,6	0	Q27	0,7	0,8	0,2	Q2	1,6	1,3	0,3
Q28	1,4	1,0	0	Q3	1	1,3	0	Q28	0,7	2,3	0,2	Q3	1,6	18,1	0,3
Q29	1,4	0,6	0	Q4	1	0,7	0	Q29	0,7	1,1	0,2	Q4	1,6	1,3	0,3
Q30	1,4	0,5	0	Q5	1	0,6	0	Q30	0,7	0,7	0,2	Q5	1,6	1,0	0,3

des. Neste caso, a discriminância obteve o menor valor em relação às outras habilidades observadas. Assim, houve o questionamento se, ao serem retirados os alunos que zeraram as questões, a discriminação também seria reduzida? Verificou-se também que a discriminação dos 75 respondentes da habilidade GAA, que corresponde ao valor de 2,515, é a mais alta entre as questões. Portanto, observou-se que ao serem retirados os alunos que zeraram a prova, a discriminação entre as questões acentuou as diferenças entre as habilidades.

Para o parâmetro de dificuldade dos 94 alunos, tanto na habilidade geral como nas seis habilidades da Tabela 13, observou-se que os valores se aproximam ao serem comparados às questões entre as habilidades, sem tanta diferença na dificuldade, exemplo Q_9 da habilidade geral com a Q_4 da habilidade IEQ. Portanto, a dificuldade entre as questões, quando estimada tanto para a habilidade geral como para as seis disciplinas, os valores para esse parâmetro atenuaram as diferenças.

Este fator somente não aconteceu quando a prova foi tratada para habilidade

GAA porque, para essa disciplina, a dificuldade entre as questões do exame, tem valores diferentes, principalmente na Q_2 . Verificou-se esta circunstância, também, porque a discriminação estimada, do mesmo modo foi de 2,103 com um acerto ao acaso de 0,1. Essa observação pode ser mantida para os 75 alunos tanto para habilidade geral que corresponde à Q_{12} como para a habilidade GAA da Q_2 , em que a dificuldade também foi alta, comparada com as outras questões.

A dificuldade, da mesma forma, é alta para os 75 respondentes na H.EST na Q_3 com o valor de 18,079 em que a mesma questão não apresentou valores de dificuldade tão discrepantes quanto os observados as habilidades dos 95 respondentes.

O acerto ao acaso, representado na Tabela 13, não teve tanta significância, assim, o valor do “chute” maior representado para todo o conjunto de dados foi para os 75 respondentes na H.EST com (0,32) quando assumiu o modelo M4. Portanto, o acerto ao acaso somente teve algum valor representativo quando foram excluídos os alunos que zeraram o exame multidisciplinar.

Outro questionamento que pode ser observado é que, ao retirar os alunos que zeraram, as estimativas dos parâmetros retrataram melhor as questões ou estas foram somente mascaradas pela quantidade de respondentes, ou ainda, os zeros inflacionados desconfiguram os resultados.

4.1.2.2 Curva característica dos itens do conjunto de dados I - TRIU

O resultado do ajuste dos modelos pode ser visto através das Curvas Características dos Itens (CCI), as quais apresentam toda a informação relevante proveniente do modelo. A CCI relaciona a habilidade dos indivíduos, com a probabilidade de estes responderem corretamente a cada um dos itens. Ou seja, é um gráfico onde no eixo das abscissas está a habilidade, e no eixo das ordenadas, a probabilidade de acerto.

As curvas características dos itens representam as informações contidas na Tabela 13, conforme as Figuras 9, 10, 11 e 12.

A Figura 9 apresenta a CCI para o modelo ajustado com os trinta itens que compõem o exame multidisciplinar, considerando o modelo $M_2(\alpha, \beta_i, 0)$.

Na Figura 9, observa-se que a cauda da curva começa em zero para todas as questões conforme o $M_2(\alpha, \beta_i, 0)$. Também se verifica que a inclinação das curvas, que corresponde ao parâmetro de discriminância, é igual e foi estimado em 1,378. A curva representada pela Q_{12} que se aproxima do eixo da habilidade é o item mais difícil perante

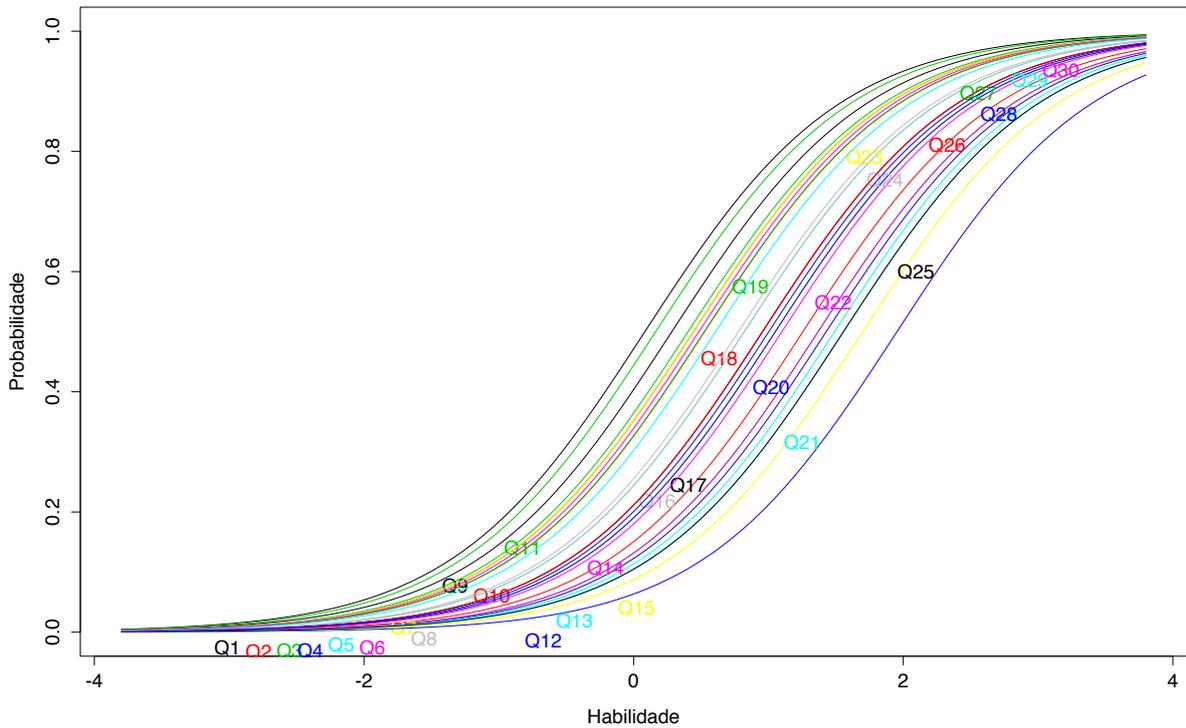


Figura 9: CCI para a habilidade geral dos 94 respondentes do $M_2(\alpha, \beta_i, 0)$

os outros e requer maior habilidade para ser respondida corretamente. Por sua vez, a curva que está mais distante do eixo das habilidades é a Q_9 , logo, é o item que no eixo das abscissas possui o menor valor, ou seja, é a mais fácil dentre as questões, por deter a menor habilidade.

A Figura 10 retrata as curvas características das trinta questões respondidas pelos 94 respondentes, considerando as seis habilidades, ao se ajustar o modelo conforme as disciplinas.

De acordo com as análises das CCI por disciplina, verifica-se, que para a maioria, não existe uma alteração relevante em relação a sua forma, já que os modelos escolhidos foram praticamente os mesmos. Porém, a disciplina de GAA apresenta uma questão que, mesmo um aluno com alta habilidade, tem baixa probabilidade de acerto, assim, esta probabilidade é praticamente toda atribuída ao acerto casual.

A questão Q_{12} foi indicada como a mais difícil quando se avaliou a prova como um todo. Ela apresentou baixa correlação PB, baixo AC e uma proporção de acerto de apenas 10%. Todas as medidas destacaram esta questão como fora do padrão das demais. Além disso, a disciplina de GAA foi sempre destacada como tendo itens não coerentes

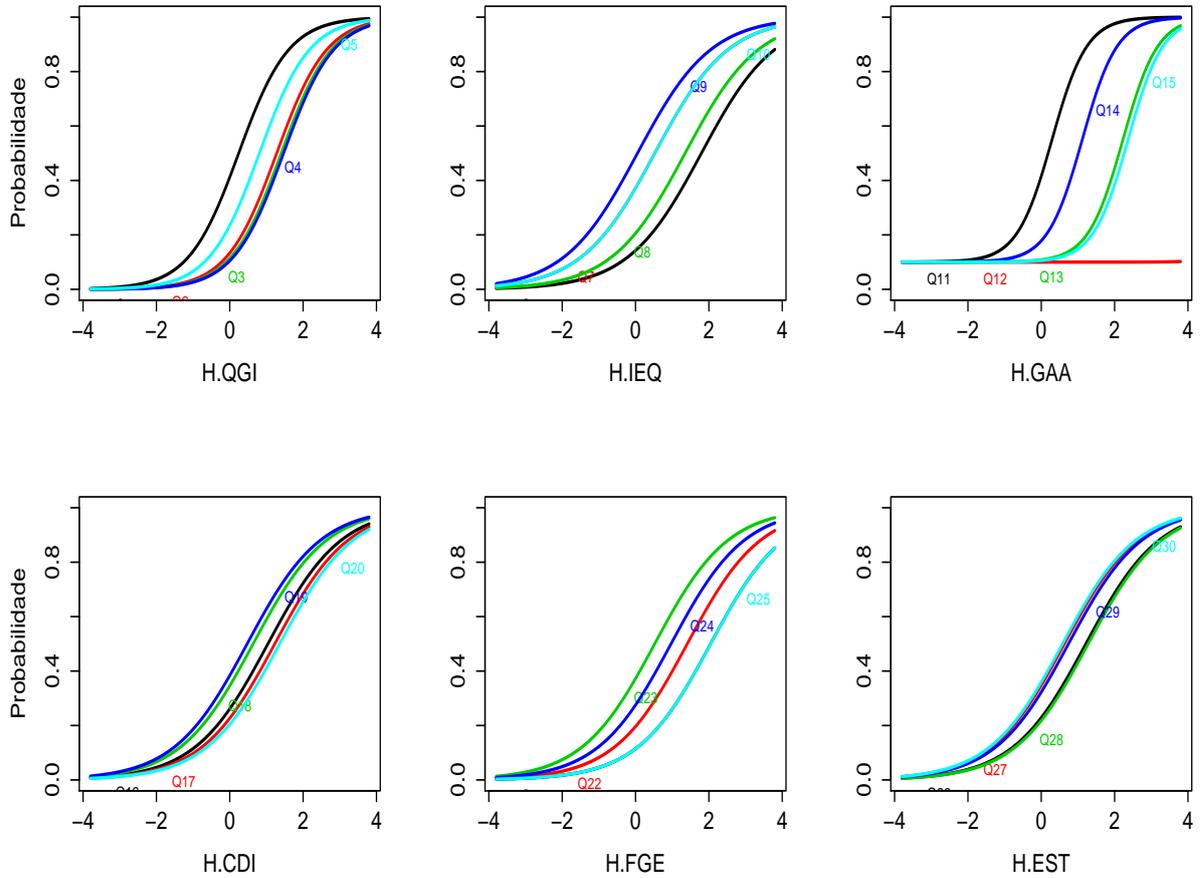


Figura 10: CCI para H.QGI - $M_2(\alpha, \beta_i, 0)$ H.IEQ - $M_1(1, \beta_i, 0)$, H.GAA - $M_4(\alpha, \beta_i, c)$, H.CDI, H.FGE e H.EST - $M_1(1, \beta_i, 0)$ dos 94 respondentes

e com baixa proporção de acertos. Isso pode indicar que a prova desta disciplina pode precisar de uma reformulação.

Na Figura 11 estão as CCI que correspondem às trinta questões respondidas pelos 75 alunos em que a habilidade geral ajustada é representada pelo $M_4(\alpha, \beta_i, c)$.

A Figura 11 mostra o acerto ao acaso, proposto de 0,15, conforme o modelo e uma discriminação pequena das questões. Por este motivo, a curva não apresentou, de imediato, o íngreme em formato de S , mas demonstrou que, para a dificuldade, o item que se destacou perante os outros foi também a Q_{12} .

A Figura 12 retratou as curvas características dos itens das trinta questões respondidas no exame multidisciplinar, pelos 75 respondentes que não zeraram a prova. Também considerou as seis habilidade conforme o ajuste do modelo de cada disciplina específica.

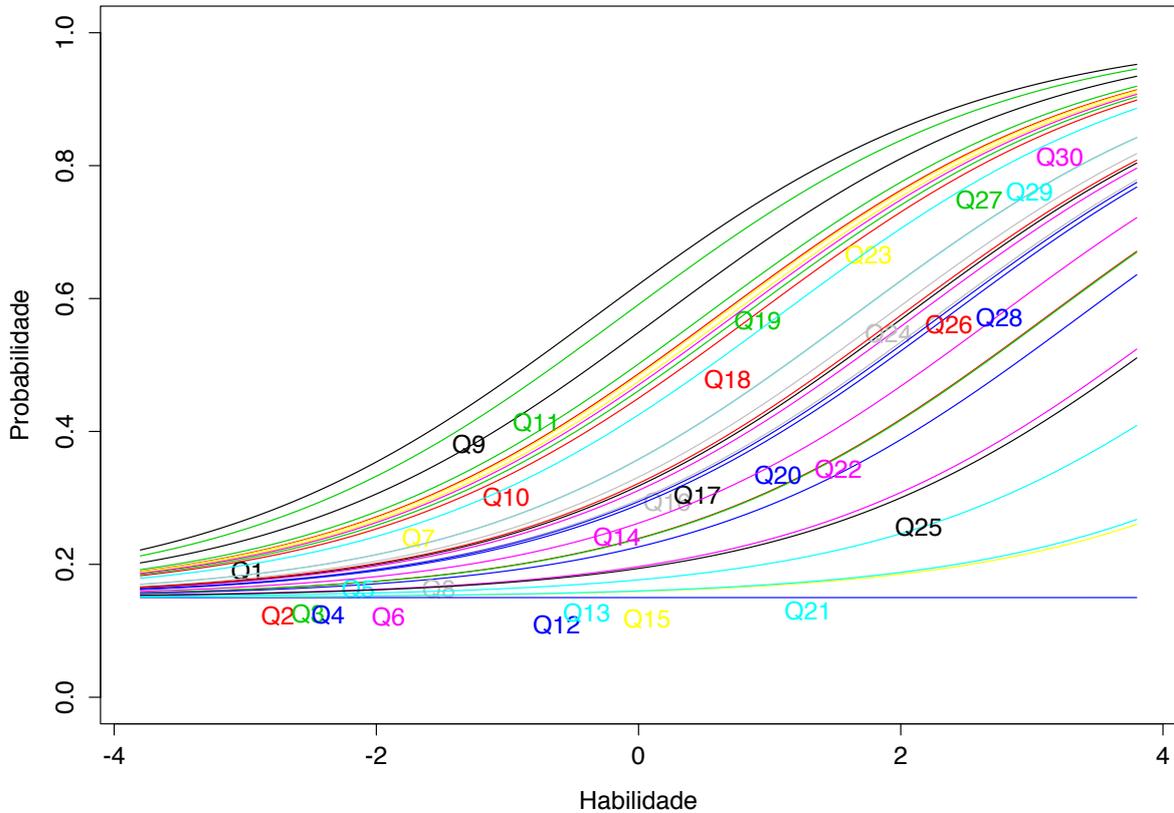


Figura 11: CCI para a habilidade geral dos 75 respondentes do M_4 (α, β_i, c)

Na Figura 12, para habilidade GAA e EST, notou-se que existem itens em que a curva linearizou-se para mostrar sua dificuldade, sem discriminação e com a probabilidade de acerto ao acaso. Os itens correspondem a H.GAA Q_2 e Q_5 que na habilidade geral equivale a Q_{12} e Q_{15} , na H.EST a Q_3 corresponde a Q_{28} , onde não foram apresentadas diferenças nas outras observações da CCI. Para H.IEQ a questão Q_1 equivalente a Q_6 para habilidade geral apresentou-se, entre as questões, como a mais difícil na comparação das questões e também não foi observada nas outras análises.

Tanto com as CCI das seis habilidades como para a habilidade geral dos 94 e 75 respondentes, foi possível a visualização imediata dos trinta itens e foram observados os valores e resultados que passaram despercebidos. Isso ocorreu quando os parâmetros de discriminância, dificuldade e acerto ao acaso foram observados e apresentados individualmente, pois representam os itens em cada curva. Assim, foram localizadas questões inesperadas a fim de apontar a dificuldade e discriminância entre elas.

De forma geral, os modelos permitiram indicar a dificuldade relativa de cada um

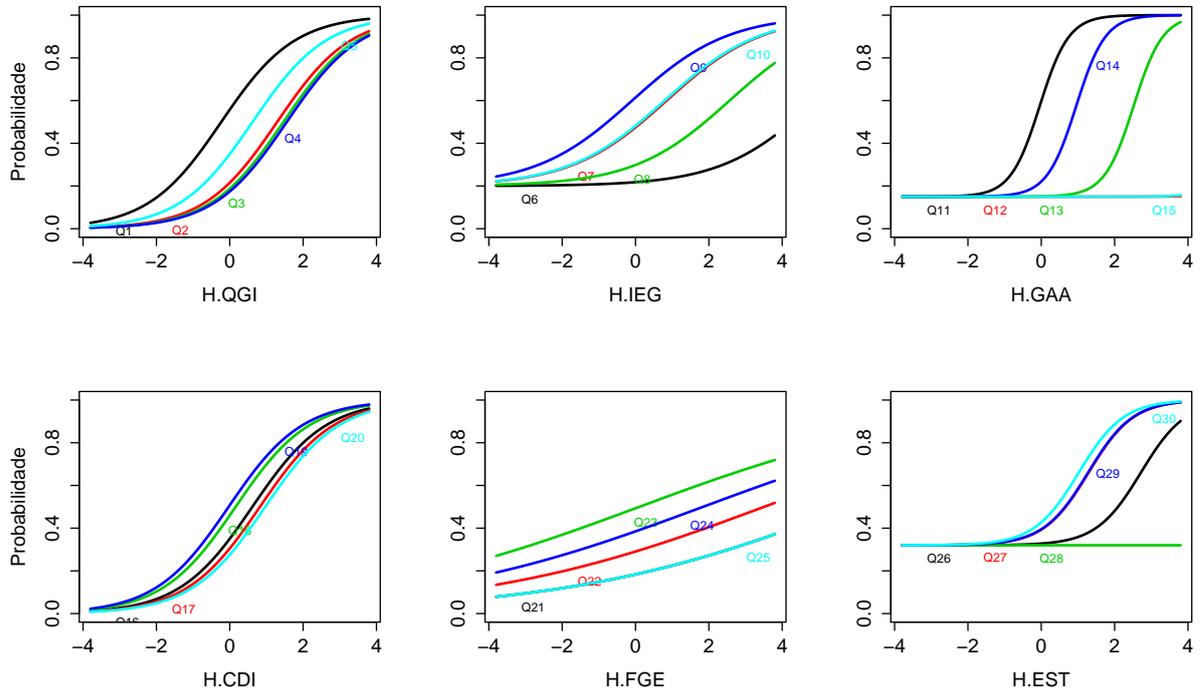


Figura 12: CCI para H.QGI e H.CDI com o $M_1 (1, \beta_i, 0)$, H.FGE $M_2 (\alpha, \beta_i, 0)$ e H.IEQ, H.GAA E H.EST com o $M_4 (\alpha, \beta_i, c)$ dos 75 respondentes

dos itens e também permitiram inferir que a discriminância dos itens é praticamente a mesma, além de mostrar que o acerto casual é pouco significativo para a maioria dos itens.

Para finalizar a análise, é interessante a comparação dos escores obtidos por alguns dos alunos, pelo TCT (número de acertos), pelos escores com o modelo geral e para cada uma das habilidades específicas. A comparação foi feita através da classificação dos alunos, para verificar se os resultados são diferentes. A subseção 4.1.2.3 verifica, de acordo com uma ou outra abordagem, por exemplo, a classificação dos alunos quanto à possibilidade de concorrer por posicionamento.

4.1.2.3 Escores das habilidades do conjunto de dados I - TRIU

Depois de estimados os parâmetros dos itens para as habilidades dos 94 e 75 respondentes, conforme a escolha de cada modelo, nesta seção, verifica-se os resultados das habilidades. Para cada habilidade dos 94 alunos avaliados pela proposta do modelo indicado da TRI, apresenta-se, na Tabela 14, alguns resultados, detalhados no Apêndice E e determina-se a localização ou posição de cada um dos discentes avaliados e o acerto por habilidade. Compararam-se os resultados pelos métodos TCT e TRI ao serem observados

os escores.

O procedimento é de posicionamento, conforme os escores obtidos por modelos. Na pesquisa, considera-se o desempenho dos avaliados, verifica-se a prova geral e também cada disciplina que compôs a avaliação. Foram observados, não somente os escores brutos padronizados na média total da prova, mas as observações com habilidades estimadas e a posição conforme a Tabela 14.

Tabela 14: Comparação dos escores de alguns alunos posicionados - TCT e TRI

A 1	A c	TCT		TRI																			
		Escore (0 a 10)	P o	H. Geral (94)		6 Habilidade (94)												A		P			
				M ₂	P o	A c	QGI M ₂	P o	A c	IEQ M ₁	P o	A c	GAA M ₄	P o	A c	CDI M ₁	P o	A c	FGE M ₁	P o	A c	EST M ₁	P o
1	12	4	7	0,547	7	1	0,001	5	2	0,207	4	1	-0,408	15	3	0,746	2	3	0,979	2	2	0,253	4
2	0	0	0	-1,566	0	0	-0,667	0	0	-0,855	0	0	-0,451	0	0	-0,780	0	0	-0,644	0	0	-0,799	0
3	11	3,7	8	0,442	8	3	0,958	3	1	-0,296	5	3	1,419	2	1	-0,216	4	2	0,492	3	1	-0,241	5
6	14	4,7	5	0,750	5	2	0,513	4	2	0,207	4	1	0,369	9	4	1,198	1	2	0,492	3	3	0,713	3
7	13	4,3	6	0,650	6	3	0,958	3	3	0,683	3	1	0,369	9	2	0,283	3	1	-0,041	4	3	0,713	3
9	15	5	4	0,850	4	5	1,855	1	2	0,207	4	1	0,369	9	2	0,283	3	2	0,492	3	3	0,713	3
12	16	5,3	3	0,949	3	2	0,513	4	5	1,635	1	2	-0,420	16	1	-0,216	4	4	1,448	1	2	0,253	4
13	15	5	4	0,850	4	3	0,958	3	4	1,152	2	1	0,369	9	4	1,198	1	1	-0,041	4	2	0,253	4
18	8	2,7	11	0,103	11	2	0,513	4	2	0,207	4	0	-0,451	0	3	0,746	2	0	-0,644	0	1	-0,241	5
26	18	6	1	1,149	1	2	0,513	4	5	1,635	1	2	1,008	4	4	1,198	1	2	0,492	3	3	0,713	3
27	17	5,7	2	1,049	2	2	0,513	4	2	0,207	4	2	1,008	4	4	1,198	1	2	0,492	3	5	1,630	1
31	10	3,3	9	0,334	9	0	-0,667	0	1	-0,296	5	0	-0,451	0	4	1,198	1	2	0,492	3	3	0,713	3
32	8	2,7	11	0,103	11	2	0,513	4	1	-0,296	5	0	-0,451	0	3	0,746	2	1	-0,041	4	1	-0,241	5
33	18	6	1	1,149	1	5	1,855	1	4	1,152	2	3	1,487	1	4	1,198	1	0	-0,644	0	2	0,253	4
48	9	3	10	0,221	10	2	0,513	4	3	0,683	3	0	-0,451	0	2	0,283	3	0	-0,644	0	2	0,253	4
61	4	1,3	15	-0,477	15	0	-0,667	0	2	0,207	4	0	-0,451	0	1	-0,216	4	1	-0,041	4	0	-0,799	0
68	6	2	13	-0,160	13	1	0,001	5	1	-0,296	5	1	0,369	9	1	-0,216	4	0	-0,644	0	2	0,253	4
71	7	2,3	12	-0,024	12	2	0,513	4	1	-0,296	5	0	-0,451	0	1	-0,216	4	1	-0,041	4	2	0,253	4

Nota: Alunos (A1); Acertos (Ac) por habilidades; Posição (Po) pelo escore obtido dos 94 respondentes.

Para melhor explorar os dados apresentados na Tabela 14, considera-se o aluno 1. Esse teve 12 acertos, ficou com um escore de 4, uma vez que respondeu corretamente a 40% da prova e foi o sétimo melhor aluno na classificação final. Quando avaliado pela TRI, seu escore ficou em 0,54, pontuação que o levou novamente ao sétimo melhor desempenho.

Para a construção da Tabela 14, observou-se que nas habilidades estimadas pelos modelos M_1 e M_2 , em que somente o parâmetro de dificuldade foi estimado, a habilidade estimada levou a um *ranqueamento* igual ao do TCT. Diferenças ocorreram somente quando as habilidades foram estimadas pelo modelo M_4 , que inclui o parâmetro de acerto ao acaso - comum a todas as questões - e a discriminância - estimada para cada questão. Apenas este modelo, entre os três escolhidos, possibilitou a obtenção de escores diferentes entre a TRI e a TCT. Como exemplos, os alunos 26 e 33, que tiveram mudanças em suas posições.

Na Tabela 14, observou-se que os escores da TCT e da TRI, para a habilidade geral que utilizou o M_2 , somente o parâmetro de dificuldade foi estimado, a discriminância

é igual para todos os itens. Verificou-se a quantidade de acertos e quando posicionados conforme o escore, averiguou-se que não houve diferenças entre os métodos da TRI e o da TCT. Essas diferenças de escores somente acontecem para a habilidade GAA com o M_4 quando se incluiu o parâmetro do acerto ao acaso e da discriminância, para ambos os casos, foi estimado um único valor, para todas as questões. Neste mesmo modelo, para o parâmetro dificuldade, estimaram-se valores para todas questões. Assim, o M_4 possibilitou ver escores diferentes para a TRI e para o TCT.

Para a habilidade QGI, que utilizou o M_2 , e para as habilidades IEQ, CDI, FGE e EST onde o modelo escolhido é M_1 , os alunos que tiveram a mesma quantidade de acertos também permaneceram na mesma posição, sem que fossem diferenciadas suas classificações. Isto ocorreu devido ao modelo proposto, em que somente a dificuldade foi estimada para todas as questões e a discriminação foi a mesma para todos os itens.

Quando foram retirados os alunos que zeraram a prova multidisciplinar, o modelo escolhido foi o M_4 e, neste caso, o escore da TRI difere da TCT. A Tabela 15 compara os diferentes modelos e verifica os escores dos 94 e 75 respondentes levando em consideração que o M_9 não é o modelo indicado, porque a quantidade de parâmetros estimados está acima do padrão de resposta. Segue a análise para conhecimento de resposta, que é o modelo mais utilizado e completo.

Na Tabela 15, o modelo escolhido, M_2 para os 94 respondentes, teve o mesmo desempenho do modelo M_4 , devido ao acerto ao acaso que foi estimado para cada questão, o valor foi zero, ou seja, não houve diferença entre os modelos do escores.

O aluno representado pelo número 1, que acertou 12 questões, possui a posição 7 tanto na TCT como para TRI na habilidade geral com 94 respondentes. Quando foram retirados os alunos que zeraram, no qual o M_4 foi proposto para a TRI, o aluno passou para trigésima primeira posição, mesmo sendo retirado somente 19 alunos, ou seja, sua habilidade é menor, por consequência sua nota apresentada também seria menor.

Verifica-se que a posição somente altera, para todos os alunos, ao comparar a TCT com a TRI na habilidade geral dos 75 respondentes do M_4 , mostrando a real habilidade que o aluno tem, conforme o grau de dificuldade e o conhecimento obtido pelo acerto das questões e não somente pelo acerto total de todos os itens.

Os alunos 9 e 13 possuem a mesma nota e classificação na TCT e na TRI com 94 respondentes, mas para a habilidade geral dos 75, o aluno 13 possui a habilidade maior do que seu colega e por consequência sua classificação é melhor. Caso se tenha somente

Tabela 15: Comparação dos escores da TCT e da TRI de alguns alunos posicionados com 94 e 75 respondentes

A l	A c	TCT				TRI									
		Escore (0 a 10)		H. Geral (94)		M ₉		M ₂		H. Geral(75)		M ₄ *		M ₉	
			Po	M ₂ *	Po	M ₄	Po		Po		Po		Po		Po
1	12	4	7	0,5473	7	0,5473	7	0,4518	35	0,2781	7	0,1860	31	-0,1958	45
2	0	0	0	-1,5657	0	-1,5657	0	-1,5134	0	-	-	-	-	-	-
3	11	3,7	8	0,4424	8	0,4424	8	0,4977	31	0,0806	8	-0,0054	37	-0,5473	58
9	15	5	4	0,8499	4	0,8499	4	1,1477	4	0,8614	4	0,8599	11	1,5282	3
12	16	5,3	3	0,9492	3	0,9492	3	0,6786	20	1,0536	3	0,6974	18	0,0109	37
13	15	5	4	0,8499	4	0,8499	4	0,7926	11	0,8614	4	1,0191	8	0,9302	14
18	8	2,7	11	0,1025	11	0,1025	11	0,2082	46	-0,5247	11	-0,2554	49	0,1180	34
26	18	6	1	1,1492	1	1,1492	1	1,1145	5	1,4364	1	1,6405	1	1,2603	6
27	17	5,7	2	1,0487	2	1,0487	2	1,0371	7	1,2452	2	1,5852	2	1,1087	9
32	8	2,7	11	0,1025	11	0,1025	11	0,1456	53	-0,5247	11	-0,3520	52	0,3327	29
33	18	6	1	1,1492	1	1,1492	1	1,2598	2	1,4364	1	1,5733	3	1,5044	4
39	16	5,3	3	0,9492	3	0,9492	3	1,2599	1	1,0536	3	1,1238	5	1,6742	2
61	4	1,3	15	-0,4767	15	-0,4767	15	-0,7581	75	-1,3733	15	-1,3121	75	-1,8843	73

Nota: Alunos (Al); Acertos (Ac); Posição (Po). O M_2^* e M_4^* são modelos de escolha.

10 vagas para um concurso, o aluno 13 seria classificado enquanto o aluno 9 não, diferente de quando são tratadas todas as questões com o mesmo grau de dificuldade, pois, neste caso, os mesmos teriam a mesma posição. Quando é comparado com os outros modelos que não foram escolhidos, ou seja, por exemplo, o modelo M_9 , verifica-se as diferenças de posição, pois existe o acerto ao acaso, mas mantém-se a posição do aluno 9 e difere da posição do aluno 13, assim, sua classificação para os 75 respondentes melhora.

Quando são comparados os alunos 26, 27 e 33, verifica-se a importância da discriminação entre as questões e a dificuldade, ou seja, os alunos 26 e 33 acertaram 18 questões e tiveram a mesma nota e classificação, tanto para TCT como para TRI do M_1 ; enquanto o aluno 27 acertou 17, logo ficou em segundo lugar. Comparando o aluno 27, no M_4 , esse possui mais habilidades do que o aluno 33, ou seja, ele acertou menos, mas a discriminação e a dificuldade entre as questões mostraram que ele acertou uma questão de dificuldade maior que o aluno 33. Logo, esse aluno possui mais habilidades, levando-o a segunda posição, mesmo com uma questão a menos do que seu colega.

Observa-se que para o modelo que não foi escolhido, M_9 , o aluno 33, que ocupava o primeiro lugar, passa para segunda posição. O aluno 26, que era também o primeiro colocado para os modelos escolhidos tanto para o 94 e 75 respondentes, passa para quinta posição e sexta posição do M_9 para ambos. Isso mostra a importância da escolha criteriosa para indicar um modelo da TRI, antes de lançar os escores. Nota-se que o aluno 39, que ocupava a terceira posição dos modelos escolhida dos 94 respondentes, ocuparia para o modelo M_9 , a primeira colocação.

Outra abordagem é usar a escala padronizado na escala métrica entre 0 a 1 para a TRIU, conforme a metodologia proposta, detalhada e escrita no Apêndice D, codificada em linguagem R . Segue, na Tabela 16, alguns resultados que estão complementados no Apêndice F. Reporta-se todo o produto dos escores do CD1 para comparar as notas pela TCT, os escores dos modelos escolhidos dos 94 e 75 respondentes e a escala padronizada.

Tabela 16: Comparação dos escores e escala padronizada da TCT e da TRIU de alguns alunos com 94 e 75 respondentes

Al	Total	TCT	H.Geral (94)/ $M_2-(\alpha, \beta_i, 0)$		H.Geral (75)/ $M_4-(\alpha, \beta_i, c)$	
	Acertos	Notas	Escore	Escala Métrica	Escore	Escala Métrica
1	12	4	0,547	0,47	0,18	0,29
2	0	0	-1,56	0	-	-
9	15	5	0,85	0,54	0,86	0,42
13	15	5	0,85	0,54	1,02	0,45
26	18	6	1,149	0,61	1,64	0,57
27	17	5,7	1,05	0,58	1,58	0,56
33	18	6	1,149	0,61	1,57	0,56
39	16	5,3	0,949	0,56	1,12	0,47

Nota: Alunos (Al); Habilidades (H).

A interpretação tornou-se conveniente pela escala métrica, atribuída de 0 a 1, com os escores dos modelos indicados para os 94 e 75 respondentes, e possibilitou a comparação das notas pela TCT.

Como exemplo, tem-se o aluno 1, que acertou 12 questões e possui a posição 7 na Tabela 15, tanto na TCT como para TRIU. Ao comparar com a 16, empregando a escala padronizada na TRIU no M_2 , a sua nota apresentou diferença. Portanto, essa diferença aumenta quando se considerou o M_4 , a nota passa de 0,47 para 0,29, mostrando que a penalização é maior quando retira-se os que zeraram e também por utilizar um modelo com o acerto ao acaso.

A mesma observação é feita para os avaliados 27 e 33, os quais diferem pela nota devido ao total de acertos. A nota é a mesma quando se utiliza o M_4 , motivo explicado pelos itens que acertaram. Nesse caso, o item que o aluno 33 acertou aproximam-se dos mesmos itens do respondente 27. A comparação entre os respondentes 33 e 26, quanto ao total de acertos é a mesma e nas notas também para a TCT e M_2 . A diferença acontece somente quando feita a retirada dos alunos que zeraram e o M_4 mostra os escores e a escala métrica diferentes. Isso indica que o aluno 26 teve um desempenho melhor na avaliação que os outros avaliados.

Estes resultados levaram aos seguintes questionamentos: isto ocorreu porque ha-

via poucos respondentes, ou por o exame multidisciplinar ter inflacionado em zeros ou ainda, pelo pacote *ltm* utilizado não conseguir reportar os resultados adequados? Esses fatores completaram a pesquisa desta tese com novas análises reportadas para responder as perguntas levantadas nestas seções.

Um fator importante é certo: a principal limitação relacionada à TCT, que não apresentou diferenças dos valores dos escores e nos posicionamentos dos modelos quando utilizadas a TRI, é que as características dos examinandos e as características dos testes não podem ser separadas. Uma só pode ser interpretada no contexto da outra e como no modelo da TRI havia somente a dificuldade a ser estimada e explicada, os valores dos posicionamentos foram iguais para os escores e posições entre a TCT e a TRI.

O CD1 respeitou a metodologia proposta na indicação da escolha do modelo, em que o número de respondentes foi maior que dos itens. Para a escolha do modelo, evitou-se graus de liberdade totais negativos, ou seja, modelos que se tenham mais parâmetros para estimar do que respondentes. Entretanto, os critérios de ajustes e a verossimilhança maximizada ($\log\text{Lik}$) foram as ferramentas principais de verificação dos ajustes dos modelos. Isso mostra que quando há 94 respondentes e 30 itens, é possível abordar a TRIU.

O valor da verossimilhança maximizada apresentou problemas numéricos na H.QGI do modelo M_8 para o M_9 . Com base nessa informação, esse fator requer uma análise metodológica para averiguar os itens e também garantir o emprego de outros conjuntos de dados para verificar o comportamento desses modelos, já que os valores aproximaram pelo critério de $\log\text{Lik}$, AIC e BIC.

Merece atenção um estudo no tamanho da amostra dos respondentes para o modelo da TRIU no interesse de responder: quantos indivíduos são necessários para se obter uma boa estimacão dos parâmetros? O número de respondentes influencia no modelo e qual é o valor que aciona um modelo? Quando vale a pena abordar um modelo mais complexo para possuir mais parâmetros a serem estimados? O modelo proposto por *Rasch* depende da quantidade de respondentes e quando deve ser empregado para fazer diferença entre a TRIU/TRIM e a TCT? Todas essas questões proporcionam discussões em busca de resultados que concedem novas abordagens.

O trabalho na escolha do modelo e na estimacão dos parâmetros possibilitou observar que o acerto ao acaso não representou de modo satisfatório a probabilidade de acerto casual. Por esses fatores, verifica-se na subseção 4.1.3 o modelo multidimensional para TRI a fim de avaliar, representar e generalizar as habilidades observadas, conforme o dimensionamento do conjunto de dados.

A tese busca ampliar os estudos e verificar como o CD1 se comporta no modelo de teoria de resposta ao item multidimensional TRIM e assim entender a capacidade e os desafios da TRI, na prova multidisciplinar, levando-se em consideração que uma prova seja composta e conseqüentemente respondida por múltiplas habilidades latentes.

4.1.3 TRIM no conjunto de dados I

Quando tratado o conjunto de dados I em mais de uma dimensão, foi necessário um pacote no *R* que atendesse às expectativas de se aplicar uma prova multidisciplinar na TRI, e não apenas em uma dimensão, como foi proposto e analisado pelo *ltm* nas subseções acima. O pacote *mirt* favoreceu essas análises e como se está trabalhando com a multidimensão, possibilitou trabalhar com uma dimensão abordada pela TRIU e TRIM. Portanto, ao aplicar a TRIM e abordar primeiramente apenas uma dimensão, não são aplicados os nove modelos conforme a hierarquia, como apresentado na subseção 4.1.2. Para esse estudo, segue o modelo multidimensional com a dimensão igual a um e, em seguida, com duas, para comparar as decisões do modelo da TRI no CD1.

O modelo proposto na metodologia da TRIM, quando aborda uma dimensão, é respectivamente o mesmo do modelo logístico da TRIU. A diferença é que, conforme aumenta a dimensão, o modelo também se adapta às dimensões propostas, detalhes da metodologia na subseção 3.2.5.

No *mirt*, como é possível abordar a TRIM uni/multidimensionalmente, são utilizados os seguintes modelos: para avaliar somente um parâmetro logístico-dificuldade (1PL) conhecido como *Rasch*, para avaliar dois parâmetros logístico (2PL) e para avaliar três parâmetros logístico (3PL). Outras abordagens dos modelos podem se vistas com detalhes em Chalmers (2012).

O pacote permite que o acerto ao acaso (c_i) seja um valor fixo diferente de 1, ou seja, determinado pelo usuário do sistema, onde uma das abordagens é perfilar um único valor para c_i . Quando se utiliza a verossimilhança ou simplesmente depois que avaliar 3PL, obtem-se os acertos ao acaso que foram estimados ao se calcular uma média aritmética. Essas abordagens analisam como o modelo se ajusta respeitando a hierarquia dos modelos e o comportamento da CCI. Outro ponto do modelo logístico é que, pelo pacote *mirt*, o valor de $D = 1$.

Primeiro, o modelo logístico com um traço latente, conhecido como unidimensional (MLD1) - equação 3.7 e o modelo com dois traços latentes/multidimensional (MLD2)

- equação 3.8 são abordados na prova pois ambos são multidisciplinares, mas há um fator amplamente dominante que é o *MLD1*, observado na Figura 8. Por consequência, analisou-se o modelo unidimensional em detalhes e foi escolhido para habilidade geral o modelo $M_2(\alpha, \beta_i, 0)$ para os 94 respondentes e $M_4(\alpha, \beta_i, c)$ para os 75 respondentes. Todavia, para o pacote *mirt*, a discriminação não pôde ser fixada em somente um único valor para α . Assim, são utilizados 3 modelos o $2PL(\alpha_i = \alpha_i, d_i = d_i, c_i = 0)$; $2PLf(\alpha_i = \alpha_i, d_i = d_i, c_i = c)$ fixando o c_i ; $3PL(\alpha_i = \alpha_i, d_i = d_i, c_i = c_i)$ e todos podem ser abordados para o uni/multidimensional, com a ressalva de que o caso descrito é para unidimensional. Para a multidimensional tem-se $\alpha_{i1}, \alpha_{i2}, \alpha_{i3} \dots \alpha_{iK}$

Nesta seção não foram detalhadas as habilidades específicas divididas em seis habilidades unidimensionalmente, a partir de agora, utiliza-se a prova geral. Verificam-se as perspectivas e os desafios de como a prova procede, sendo analisada uni/multidimensionalmente. Na TRIM, as análises são exploratórias e confirmatórias, então não faz sentido dividi-la em seis habilidades, porque seria inviável esta análise de seis dimensões, pelo fato de o número de parâmetros estimados ser maior que o número de respondentes.

A partir de duas dimensões, o valor correspondente ultrapassaria a quantidade de respondentes. Exemplificando, para a análise dos 94 respondentes, seriam, seis dimensões com 30 itens, assim o modelo teria 180 parâmetros de estimação, ou seja, o valor estimado seria maior que dos respondentes, inviabilizando as análises. Por esse motivo, trabalhou-se primeiramente com a TRIM compensatória no modelo exploratório e caso houvesse indicação de mais do que uma dimensão seria aplicada a análise confirmatória.

Outro fator importante é que, com as análises paralelas, a TRIM indicou, pelo menos, mais que duas dimensões para o CD1. O pacote *mirt* foi utilizado para validar a aplicação da teoria de resposta ao item e comparar os modelos, de acordo com a verossimilhança e pelos critérios de AIC e BIC e assim verificar os ajustes do MLD1 e MLD2, conforme a construção da Tabela 17.

Na Tabela 17, para os 94 respondentes, no modelo uni/multidimensional, os valores dos critérios da *log-verossimilhança* de 2PL e 2PLf foram iguais. O motivo é que, no parâmetro do acerto ao acaso, quando o perfilhamento na verossimilhança foi estimado para ser fixado, o valor resultou em zero. Logo, como 2PL, o c_i é igual a zero e o valor dos critérios de 2PLf também foi o mesmo. Contudo, os valores do acerto ao acaso para o " c_i " perfilhando dos 75 respondentes no 2PLf, é 0,1034. Por esse motivo, não foram observados valores iguais para o 2PL e 2PLf.

O padrão de resposta dos 94 respondentes é de 75, logo, conforme a metodologia

Tabela 17: Resultados das estatísticas de diagnósticos da qualidade dos 94 e 75 respondentes do ajuste dos modelos do conjunto de dados I

MLD1		94 respondentes			
		df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0$)		60	-1443,862	3007,72	3160,32
2PLf ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c$)		61 [”]	-1443,862	3007,72	3160,32
3PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c_i$)		90 ^{*”}	-1443,864	3067,73	3296,62
MLD2					
2PL ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0$)		90 [*]	-1414,82	3007,64	3233,99
2PLf ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c$)		91 ^{*”}	-1414,82	3007,64	3233,99
3PL ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c_i$)		120 [*]	-1412,86	3063,72	3366,37
MLD1		75 respondentes			
		df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0$)		60	-1338,75	2797,49	2936,54
2PLf ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0,103$)		61	-1336,05	2792,11	2931,16
3PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c_i$)		90 [*]	-1323,52	2827,04	3035,62
MLD2					
2PL ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0$)		90 [*]	-1311,21	2800,42	3006,67
2PLf ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0,103$)		91 [*]	-1302,99	2783,99	2999,25
3PL ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c_i$)		120 [*]	-1287,31	2812,61	3088,39

Nota:

- modelo logístico de um traço latente (MLD1);
- modelo logístico de dois traços latentes (MLD2);
- dois parâmetros logísticos (2PL) ;
- dois parâmetros logísticos fixando o c_i (2PLf);
- três parâmetros logísticos (3PL); e
- quantidade de parâmetros (df).

Os valores com (”) possuem a quantidade de parâmetros onde os valores fixo de c e c_i são estimados iguais a zero.

Os valores com (*) representam que a quantidade de parâmetros é maior do que o total de respondentes.

na subseção 3.2.7, é indicado verificar a quantidade de parâmetros (df) obtidos e aconselha-se a não ultrapassar da quantidade de respondentes tanto no total como nos padrões de respostas. No MLD2 todos os df ultrapassam o padrão de resposta, mesmo que os modelos ajustem-se pelos resultados da $logLik$, assim, não é aconselhado a escolha dos modelos da MLD2.

A indicação é para os modelos do MLD1 com 2PL para os 94 respondentes, pois no MLD1 do 3PL, o acerto ao acaso foi estimado, para os 94 avaliados, igual a zero e a quantidade de df não é aconselhável por extrapolar o valor de padrão de resposta. Como a $logLik$, AIC e BIC não apresentaram diferenças, os critérios estatísticos basearam-se pela quantidade de parâmetros.

Para os 75 respondentes no MLD1 a $logLik$, AIC e BIC indicam o modelo 3PL, mas a quantidade de números de parâmetros não é aconselhada. Assim, optou-se pelo modelo 2PLf pela diferença de parâmetros estimados ser de uma unidade e satisfazer também os critérios diagnósticos da estatística.

4.1.3.1 Estimação dos parâmetros nos modelos da TRIM no conjunto de dados I

Retomam-se as análises na TRIM e seguem os valores estimados para cada item dos 94 e 75 respondentes, conforme os modelos indicados na Tabela 18.

Tabela 18: Resultados das estimativas dos parâmetros dos MLD1 e MLD2 para os 94 e MLD1 para 75 respondentes

Q	94 respondentes							75 respondentes					
	MLD1			MLD2				MLD1					
	2PL/2PLf/3PL			2PL/2PLf				2PL			2PLf		
	α_{i1}	d_i	$c c_i$	α_{i1}	α_{i2}	d_i	c	α_{i1}	d_i	c	α_{i1}	d_i	c
Q1	2,33	-0,76	0	2,06	0,48	-0,72	0	1,24	0,25	0	1,1	0	0,1
Q2	1,24	-1,75	0	1,19	0,16	-1,75	0	0,64	-1,17	0	1,2	-2,1	0,1
Q3	2,31	-2,56	0	2	1,1	-2,72	0	1,16	-1,53	0	3,2	-3,9	0,1
Q4	2,54	-2,82	0	3,32	-0,35	-3,41	0	1,23	-1,66	0	13,8	-17,4	0,1
Q5	1,61	-1,32	0	2,16	-0,55	-1,54	0	0,74	-0,58	0	1,2	-1,1	0,1
Q6	0,52	-1,59	0	0,48	0,11	-1,59	0	-0,36	-1,26	0	-0,5	-1,9	0,1
Q7	1,06	-0,58	0	1,26	-0,27	-0,61	0	0,08	-0,03	0	0,2	-0,3	0,1
Q8	1,43	-1,56	0	1,87	-0,48	-1,77	0	0,53	-0,87	0	0,6	-1,3	0,1
Q9	2,3	-0,41	0	2,01	0,95	-0,42	0	1,04	0,57	0	1,2	0,4	0,1
Q10	1,66	-0,78	0	1,5	0,48	-0,78	0	0,7	-0,03	0	0,8	-0,3	0,1
Q11	2,16	-0,51	0	1,97	0,53	-0,5	0	1,43	0,5	0	2,9	0,3	0,1
Q12	0,77	-2,44	0	0,83	-0,14	-2,47	0	-0,08	-2	0	9,8	-23,6	0,1
Q13	0,59	-1,77	0	0,64	-0,07	-1,78	0	-0,19	-1,4	0	-0,3	-2,2	0,1
Q14	1	-1,23	0	0,9	0,26	-1,23	0	0,2	-0,7	0	0,2	-1,1	0,1
Q15	0,81	-2,13	0	0,58	0,45	-2,14	0	0,09	-1,66	0	0,3	-2,7	0,1
Q16	0,57	-0,93	0	0,34	0,55	-0,96	0	-0,43	-0,54	0	-0,5	-0,9	0,1
Q17	1,21	-1,32	0	0,84	1,54	-1,68	0	0,33	-0,71	0	0,4	-1,1	0,1
Q18	1,04	-0,68	0	0,89	0,35	-0,68	0	-0,09	-0,13	0	-0,2	-0,4	0,1
Q19	1,86	-0,79	0	1,62	0,61	-0,79	0	0,86	0,04	0	0,9	-0,2	0,1
Q20	1,1	-1,4	0	0,69	4,59	-3,78	0	0,25	-0,83	0	0,4	-1,3	0,1
Q21	0,61	-1,86	0	1,56	-1,56	-2,71	0	-0,22	-1,49	0	-0,5	-2,4	0,1
Q22	0,8	-1,34	0	0,92	-0,23	-1,38	0	-0,06	-0,88	0	-0,3	-1,3	0,1
Q23	1,34	-0,67	0	1,13	0,53	-0,67	0	0,32	-0,03	0	0,4	-0,3	0,1
Q24	0,93	-0,98	0	0,92	0,04	-0,98	0	0,1	-0,46	0	0,3	-0,8	0,1
Q25	1,74	-2,47	0	2,28	-0,55	-2,86	0	1,01	-1,75	0	2,4	-3,9	0,1
Q26	1,66	-1,54	0	1,45	0,54	-1,54	0	1,14	-0,86	0	1,3	-1,4	0,1
Q27	1,81	-0,96	0	1,66	0,33	-0,94	0	0,55	-0,14	0	0,3	-0,4	0,1
Q28	0,87	-1,24	0	1,11	-0,41	-1,33	0	0,05	-0,75	0	0	-1,1	0,1
Q29	1,4	-0,92	0	1,27	0,43	-0,93	0	0,51	-0,26	0	0,6	-0,6	0,1
Q30	1,34	-0,73	0	1,06	0,97	-0,8	0	0,37	-0,08	0	0,4	-0,3	0,1

Nota:

- modelo logístico de um traço latente (MLD1);
- modelo logístico de dois traços latentes (MLD2);
- dois parâmetros logísticos (2PL) ;
- dois parâmetros logísticos fixando o c_i (2PLf); e
- três parâmetros logísticos (3PL).

Para os resultados da Tabela 18, utilizou-se o pacote *mirt*, cujo procedimento iterativo de estimação do parâmetro aborda a máxima verossimilhança marginal. A integração pelo método de Gauss - Hermite utilizou a estimativa EM tradicional e os modelos exploratórios e de confirmação são estimados por algoritmo estocástico, detalhes em Cai

(2010).

Na Tabela 18, somente nos 75 respondentes no modelo MLD1 em 2PLf, o acerto ao acaso teve o valor estimado em 0,1. Outra observação importante é que, ao olhar a discriminância para todas as questões, a $Q4$ sobressai, principalmente para os 75 respondentes no MLD1/2PLf, com um valor de 13,8, mostrando que é o item que mais discrimina na prova quando o parâmetro é estimado pela TRIM.

4.1.3.2 Cargas fatoriais e CCI dos itens na TRIM no conjunto de dados I

Ao aplicar a TRIM no pacote *mirt*, a indicação para os 94 respondentes é o modelo 2PL $(\alpha_i, d_i, 0)$ e para os 75, o 2PLf (α_i, d_i, c) , ambos no MLD1 .

No pacote *ltm*, somente é possível detalhar os modelos unidimensionais e, como estão aninhados, a escolha apontou para o $M_2 (\alpha, \beta_i, 0)$, cuja diferença está na discriminância, onde no *ltm* foi estimada e fixada em um único valor para os 94 respondentes.

Para os 75 respondentes, o modelo indicado é $M_4 (\alpha, d_i, c)$. Conforme Nojosa (2002), Reckase (2009) e Chalmers (2012), a técnica das cargas fatoriais possibilitam informar e também se destinam à verificação de quais itens se relacionam a cada fator, pois quanto maior a carga fatorial, melhor é a relação entre item e o fator.

Os resultados, após a rotação oblíqua, das cargas fatoriais dos modelos indicados na TRIM, estão apresentados na Tabela 19.

Na Tabela 19 é possível verificar que para os 94 respondentes, quando são analisadas as cargas fatoriais do MLD2, somente as $Q20$ e $Q21$, explicam-se melhor na segunda dimensão, ou seja, para essa prova, a grande maioria das questões foi explicada no $F1$ com exceção de duas questões. Portanto, é conveniente o modelo multidimensional com a indicação de um traço latente, porque em 28 questões o aluno necessitou de apenas uma habilidade para responder à prova e somente duas questões carregaram mais no segundo fator.

Com relação às cargas fatoriais, Reckase (1988), considera que há unidimensionalidade quando o primeiro fator da conta de 20% da variância total. Para que um item se ajuste, é necessário que sua comunalidade seja maior que 0,3. Essa observação é relevante, e a comunalidade utilizada será maior que 0,15 pois, ao retirar o item $Q6$ do modelo, conforme as considerações das carga fatorial e comunalidade atendendo o pré-requisito, o modelo não sofreu alteração e nem houve indicação para outro modelo.

Tabela 19: Cargas fatoriais e comunalidades dos MLD1 e MLD2 para o CD1

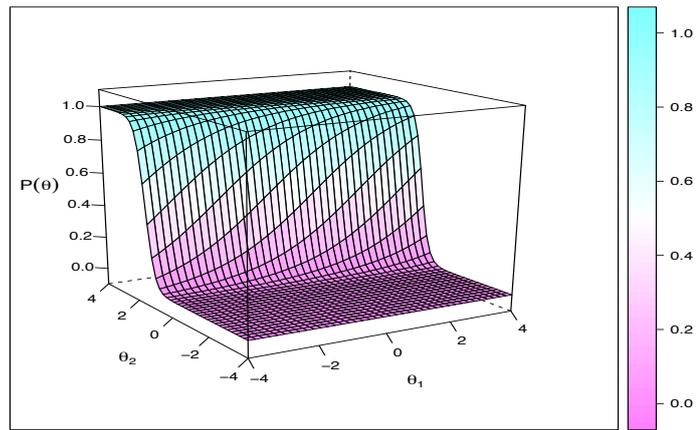
Q	MLD1 - 94		MLD2 - 94			MLD1 - 75			
	2PL/2PLf/3PL		2PL/2PLf			2PL		2PLf	
	F1	h2	F1	F2	h2	F1	h2	F1	h2
Q1	0,81	0,65	0,79	0,01	0,63	0,59	0,34	0,54	0,29
Q2	0,59	0,35	0,59	-0,04	0,35	0,35	0,12	0,58	0,34
Q3	0,81	0,65	0,8	0,2	0,69	0,56	0,32	0,88	0,78
Q4	0,83	0,69	0,85	-0,26	0,78	0,59	0,34	0,99	0,99
Q5	0,69	0,47	0,7	-0,34	0,6	0,4	0,16	0,56	0,32
Q6	0,29	0,08	0,29	0	0,09	-0,21	0,04	-0,28	0,08
Q7	0,53	0,28	0,53	-0,24	0,34	0,05	0	0,13	0,02
Q8	0,64	0,42	0,65	-0,32	0,53	0,29	0,09	0,34	0,11
Q9	0,8	0,65	0,8	0,16	0,67	0,52	0,27	0,59	0,34
Q10	0,7	0,49	0,7	0,06	0,49	0,38	0,15	0,41	0,17
Q11	0,79	0,62	0,79	0,03	0,62	0,64	0,41	0,86	0,74
Q12	0,41	0,17	0,4	-0,15	0,18	-0,05	0	0,99	0,97
Q13	0,33	0,11	0,33	-0,11	0,12	-0,11	0,01	-0,2	0,04
Q14	0,51	0,26	0,51	0,03	0,26	0,12	0,01	0,11	0,01
Q15	0,43	0,18	0,41	0,16	0,19	0,05	0	0,17	0,03
Q16	0,32	0,1	0,31	0,23	0,15	-0,25	0,06	-0,3	0,09
Q17	0,58	0,34	0,58	0,48	0,56	0,19	0,04	0,24	0,06
Q18	0,52	0,27	0,51	0,07	0,27	-0,06	0	-0,1	0,01
Q19	0,74	0,54	0,73	0,09	0,55	0,45	0,2	0,46	0,21
Q20	0,54	0,29	0,52	0,78	0,89	0,14	0,02	0,20	0,04
Q21	0,34	0,11	0,33	-0,67	0,56	-0,13	0,02	-0,3	0,09
Q22	0,43	0,18	0,42	-0,2	0,21	-0,04	0	-0,18	0,03
Q23	0,62	0,38	0,61	0,12	0,39	0,18	0,03	0,24	0,06
Q24	0,48	0,23	0,47	-0,08	0,23	0,06	0	0,15	0,02
Q25	0,71	0,51	0,72	-0,34	0,63	0,51	0,26	0,81	0,66
Q26	0,7	0,49	0,69	0,08	0,49	0,56	0,31	0,62	0,38
Q27	0,73	0,53	0,72	-0,01	0,52	0,31	0,1	0,19	0,04
Q28	0,45	0,21	0,45	-0,29	0,29	0,03	0	0,01	0
Q29	0,64	0,4	0,64	0,06	0,42	0,29	0,08	0,35	0,12
Q30	0,62	0,38	0,63	0,28	0,47	0,21	0,05	0,25	0,06

Nota: F1 carga de um fator; F2 carga de dois fatores; h2 comunalidade.

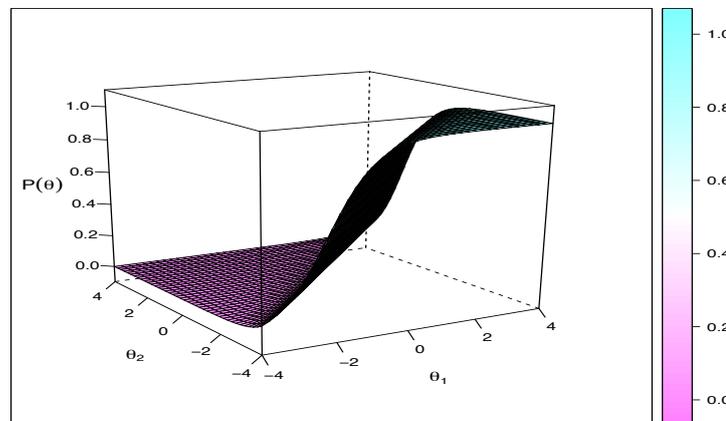
Como a intenção não é calibrar os itens e sim verificar as perspectivas e os desafios no exame multidisciplinar, optou-se em não retirar o item para as próximas análises.

O modelo logístico da TRI é uma função que apresenta na CCI a forma de um “S”alongado. Conforme a Figura 13, seguem as CCIs das questões 20, 21 e 4, que representam o comportamento das cargas fatoriais com duas dimensões, observado na Tabela 19, nas quais as $Q20$ e $Q21$ tiveram suas cargas no segundo fator, ou seja, em duas dimensões. Enquanto a $Q4$ teve o valor de 0,85 na primeira carga fatorial e, conforme o comportamento do gráfico, é possível verificar que a CCI é explicada na primeira dimensão.

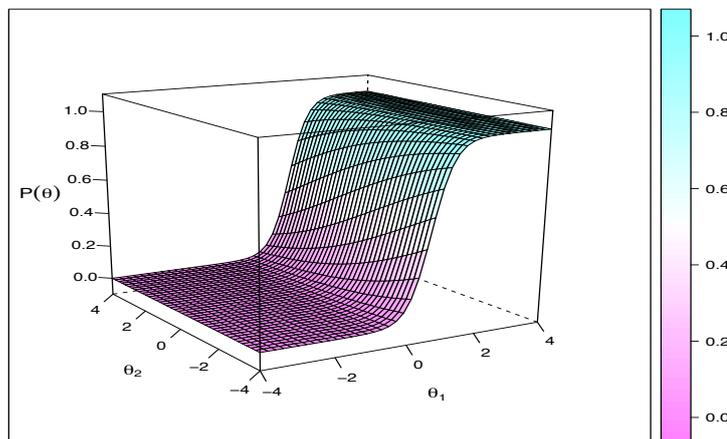
Na Figura 13(c) da $Q4$, é possível dizer, para este item, que os alunos necessitaram de apenas uma habilidade para respondê-lo e quase todos os outros itens também apresentaram esse comportamento, fora as $Q20$ e $Q21$ que necessitam de duas habilidades conforme a Figura 13(a) e 13(b) respectivamente.



(a)



(b)



(c)

Figura 13: CCI da Q20 (a), Q21 (b) e Q4 (c) do MLD2- 2PL com 94 respondentes

Como ainda é recente, na bibliografia, a técnica de como abordar a TRI em uma prova com itens unidimensional/bidimensional, devido o pacote mirt estar em desenvolvimento, seguem-se as análises a fim de verificar o escore dos respondentes, conforme a indicação unidimensional dos modelos.

4.1.3.3 Escores das habilidades da TRIM no conjunto de dados I

Para abordar os escores da TRIM para cada respondente, a habilidade avaliada foi representada pelo escore padronizado, conforme a metodologia proposta na subsecção 3.2.6, que possibilita analisar os resultados de cada respondente e como ele respondeu a prova. Caso o exame multidisciplinar avaliado aplicasse a TRIM, seguem as comparações com as notas lançadas pela TCT, que foi o método utilizado para a divulgação e tomada de decisões pedagógicas.

A Tabela 20 apresenta os resultados de alguns escores dos respondentes do CD1 e utiliza-se a escala padronizada conforme os modelos indicados pelos critérios AIC, BIC e $\log Lik$, com a possibilidade de comparar a pontuação da TCT, que foi o modelo utilizado na prova, mais detalhes no Apêndice D da escala padronizada e Apêndice E com todos os resultados.

Tabela 20: Comparação dos escores dos 20 primeiros alunos dos 94 e 75 respondentes dos modelos indicados para TRIM

Al	Total	TCT	MLD1 (94) - 2PL		MLD1 (75) - 2PLf	
	Acertos	Notas	Escore	EP	Escore	EP
1	12	4	0,501	0,45	-0,367	0,30
2	0	0	-1,639	0	-	-
3	11	3,7	0,519	0,45	0,525	0,50
4	12	4	0,657	0,48	0,512	0,50
5	12	4	0,515	0,45	0,292	0,45
6	14	4,7	0,831	0,51	0,736	0,55
7	13	4,3	0,888	0,53	1,333	0,68
8	13	4,3	0,79	0,51	0,745	0,55
9	15	5	1,22	0,60	1,83	0,79
10	0	0	-1,639	0	-	-
11	0	0	-1,639	0	-	-
12	16	5,3	0,75	0,50	-0,26	0,33
13	15	5	0,904	0,53	0,761	0,55
14	16	5,3	1,004	0,55	0,837	0,57
15	0	0	-1,639	0	-	-
16	14	4,7	0,887	0,53	0,664	0,53
17	14	4,7	0,836	0,52	0,591	0,52
18	8	2,7	0,258	0,39	-0,147	0,35
19	14	4,7	0,757	0,50	-0,22	0,33
20	8	2,7	0,049	0,35	-0,9	0,18

Nota: Alunos (Al); Habilidades (H); Escala padronizada (EP)

A Tabela 20 apresenta somente os 20 primeiros respondentes dos 94 em pesquisa. Os resultados dos escores fornecidos pelas teorias e a classificação com a utilização da escala padronizada tornam inevitável a comparação e a classificação das notas maiores para as menores, como os indivíduos 6, 16, 17 e 19. Todos possuem um total de acertos de 14 questões em que, na TCT, a pontuação corresponde a 4, 7 e ocupa a terceira posição

como classificação.

Para a TRIM MLD1 de 2PL, de acordo com a classificação das notas e dos resultados, a posição é diferente, pois os 4 indivíduos citados no parágrafo acima não possuem a mesma classificação seus escores são diferentes devido à teoria da resposta ao item, a classificação dos indivíduos 6, 16, 17 e 19 que acertaram 14 itens, a classificação da maior nota é do indivíduo 17, oposto a do respondente 16, que ficou na décima primeira posição.

Quando foram retirados os que zeraram na TRIM, os 75 respondentes no MLD1 em 2PL e 2PLf, o respondente 17 ficou na sétima posição. Entre os quatro, o melhor escore é do indivíduo 6, ocupando a quinta posição e o indivíduo 19 teve o menor escore.

Segue a Figura 14 com detalhes sobre as notas na escala padronizada dos 94 e 75 respondentes, conforme os modelos MLD1 2PL e 2PLf respectivamente.

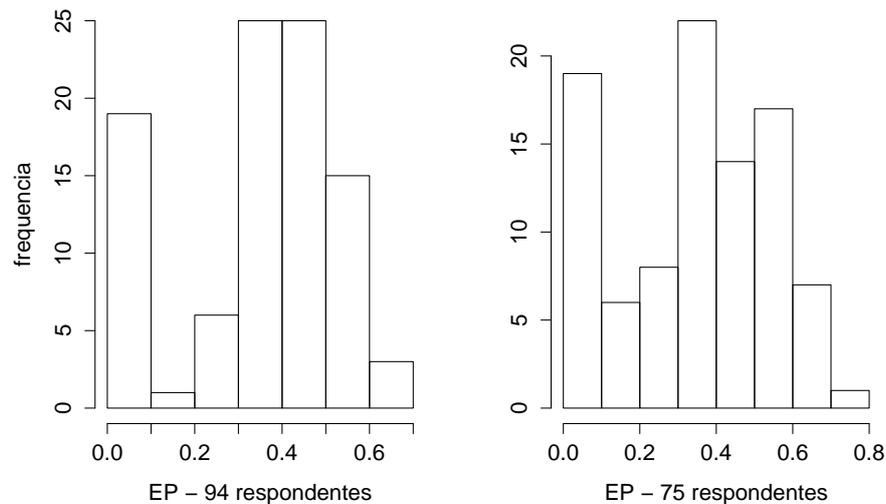


Figura 14: Escores dos 94 e 75 respondentes na escala padronizada

Na Figura 14, as notas na escala padronizada, conforme os modelos indicados nas análises, mostram que a avaliação multidisciplinar pela TRI estão abaixo da média considerada pela instituição - pela escala 0,7. As notas ficaram em torno de 0,4 mesmo com a exclusão dos que zeraram e ao se considerar o acerto ao acaso para os 75 respondentes, foram apresentados poucos respondentes acima da média. Todavia, o mesmo não foi observado para o modelo indicado para os 94 respondentes.

Observa-se que, dependendo da escolha da teoria com o modelo estipulado e as estimativas dos parâmetros, a classificação dos respondentes difere muito. Assim, surge

um questionamento ao se comparar a Tabela 16 que utiliza o pacote *ltm* com a Tabela 20 pelo *mirt*, ambas estão detalhadas no Apêndice F. Os valores dos escores não são os mesmos devido aos números de interações e os métodos de convergências utilizados para estimação dos parâmetros.

Contudo, esse questionamento foi verificado pela Figura 15(a), que relaciona os resultados dos escores do MLD1- 2PL do pacote *mirt*, no eixo das abcissas, com o modelo equivalente, o M_5 do *ltm*, representado pelo eixo das ordenadas, em que ambos possuem os parâmetros em $(\alpha, \beta_i, 0)$. Seguem os resultados a fim de verificar a correlação entre os escores dos dois *softwares*, pela Figura 15 (a) e complementado pelas Figuras 15 (b e c), de correlações da escala padronizada entre o MLD1-2PL e a TCT.

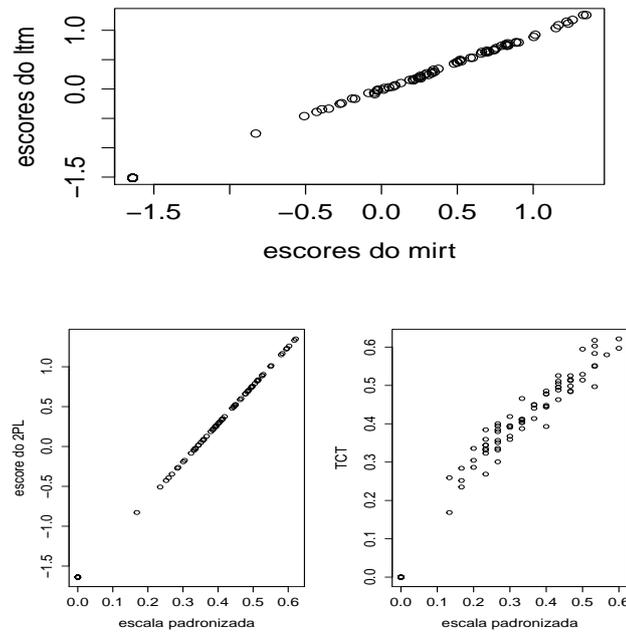


Figura 15: Correlação entre: os escores do pacote *mirt*/ MLD1-2PL & *ltm*/ M_5 e a escala padronizada com a MLD1-2PL e TCT

Na Figura 15, mesmo que o modelo M_5 do *ltm* não tenha sido indicado, ele é o modelo equivalente do MLD1-2PL, assim, os escores estão correlacionados. As Figuras da escala padronizada entre o escore 2PL também apresentaram correlações e para TCT existe aproximação nos valores, conforme mostra o gráfico, mesmo que tenham sido abordados pacotes com proposta computacional diferentes, foram comparáveis.

Como o cenário é de observações e automaticamente de comparações e reforçando as correlações dos resultados obtidos, analisa-se o CD1 na TRIU com o pacote *ltm* enquanto na TRIM utiliza-se o *mirt*. Logo, observou-se que os resultados dos critérios

estatísticos se aproximam em AIC, BIC e na $\log Lik$, conforme a Tabela 21.

Para a TRIU, o modelo conhecido na literatura com dois parâmetros é o ML2 e que na metodologia dessa tese, conforme a hierarquia, é o $M_5(\alpha, \beta_i, 0)$, foi comparado na metodologia da TRIM, onde este modelo está correlacionado ao MLD1 com 2PL ($\alpha, d_i, 0$). As informações estão apresentadas na Tabela 21.

Tabela 21: Valores observados no pacote *ltm* (M_5) e *mirt* (2PL) no MLD1

Observações	Pacote ltm	Pacote mirt
logLik	-1443,22	-1443,86
AIC	3006,445	3007,72
BIC	3159,043	3160,32

Essas diferenças ocorrem devido aos métodos utilizados para convergência dos modelos e também aos métodos para estimar os parâmetros. Portanto, ao serem utilizados os pacotes para a TRI, o método utilizado da estimação é de fundamental importância até para que não sejam feitas escolhas precipitadas de como ou qual delas utilizar.

4.2 MODELO LOGÍSTICO DA TRI NO CONJUNTO DE DADOS II

O conjunto de dados II é composto por 189 alunos da graduação, os quais realizaram o exame multidisciplinar com 40 itens. Tal avaliação multidisciplinar interna da PUCPR é baseada nos modelos da TRI e com a perspectiva de complementar, validar e apoiar os resultados já detalhados no CD1.

As análises descritivas estão delineadas na subseção 2.1.2 para que se avalie o aprendizado e a aquisição de habilidades e as competências, de acordo com o perfil dos estudantes. São abordadas as análises nos modelos exploratórios pela TRIM, que oferece subsídios para avaliação. Os modelos contribuíram com o trabalho de melhoria metodológica/ pedagógica do ensino e, juntamente com as análises, auxiliaram no processo de planejamento para a validação do modelo na TRI.

A primeira etapa das análises aborda a dimensionalidade do exame multidisciplinar na subseção 4.2.1.

4.2.1 Dimensionamento do conjunto de dados II

A princípio, a dimensionalidade da TRI no CD2 é estudada com base na análise paralela da análise fatorial (FA) e dos componentes principais (PC). A metodologia, na subseção 3.2.6 e 3.2.7, é aplicada no exame multidisciplinar com 40 questões para verificar como é tratada a dimensão na avaliação toda.

A avaliação, para os 189 respondentes no CD2, de acordo com o gráfico da Figura 16 (a), possui um autovalor superior aos demais e apresenta um indicativo de que o instrumento é unidimensional, ou que, pelo menos, há um fator dominante que se destacou tanto na FA como no PC.

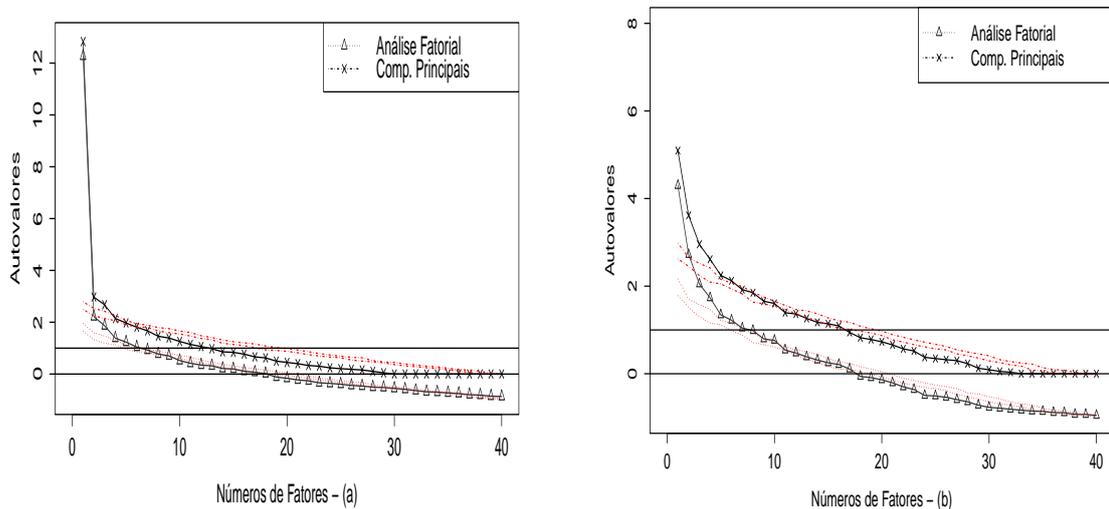


Figura 16: Diagnóstico da dimensão dos 189 (a) e 154 (b) respondentes

O estudo da dimensionalidade é obtido pelo pacote *psych* juntamente com as análises paralelas dos componentes principais sobre a matriz de correlação tetracórica e das análises dos fatores. O instrumento verifica as dimensões interpretáveis dos 189 respondentes e, bem como de quando se retira os 35 alunos que zelaram na prova, totalizando 154 respondentes. As informações dos gráficos da matriz tetracórica possibilitam a indicação da dimensionalidade, conforme as relações estruturais dos itens na prova, como sintetizada pela Figura 16.

Na Figura 16(a), para os 189 respondentes, existem 4 fatores dos componentes principais (PC) acima da linha de corte, ou seja, da linha tracejada, que são valores simulados dos PC. Os fatores acima dessa linha representam fatores explicativos, mas o ganho, o quanto cada fator explica, aponta a existência de apenas um fator bem definido.

Assim, há necessidade de investigar até o segundo fator porque, a partir desse, as distâncias entre os fatores são as mesmas e não há ganhos de fatores explicativos. A mesma verificação acontece para análise fatorial (FA), onde há nove fatores acima da linha de corte, mas somente dois distanciam-se entre si. Portanto, a pesquisa é feita em até duas dimensões, visto que, para os 154 respondentes, existem oito fatores acima dos componentes principais e dez fatores na análise fatorial, conforme a Figura 16(b). Ao serem retirados os alunos que zeraram na prova, os fatores das análises paralelas se aproximaram e é possível indicar que há, não somente uma dimensão dominante, mas ganhos menores de dimensões. Por esse motivo, há necessidade de que se explorem os modelos da TRIM até em duas dimensões.

A prova é multidisciplinar e aborda as áreas específicas de administração e de exatas, mas há um fator dominante como é observado na Figura 16, tanto nos 189 como nos 154 respondentes. Vale ressaltar que há indicação do segundo fator para os respondentes.

Na subseção 4.2.2 são feitas as análises sobre os resultados obtidos com o pacote *mirt*, que aborda a teoria de resposta ao item multidimensional. Assim, pode-se dizer que a prova é primeiramente unidimensional, mas é necessário investigar a segunda dimensão para a validade da aplicação da teoria de resposta ao item e explorar também a abordagem da TRIM.

4.2.2 TRIM no conjunto de dados II

A metodologia descrita na subseção 3.2.2 é empregada no modelo logístico da TRIU e TRIM. A diferença entre ambas é que, na análise TRIM, ao aumentar a dimensão no modelo, essa também se adapta às dimensões propostas. O questionamento é: quando vale a pena abordar mais que uma dimensão? Leva-se em conta que já houve a indicação dominante na subseção 4.2.1 do modelo ser unidimensional, mas também verifica-se a segunda dimensão pelo fato do exame ser multidisciplinar. Com a indicação, tornam-se necessárias outras análises para diagnosticar qual modelo deve ser escolhido, dada a dimensionalidade do conjunto. Portanto, são obtidos e analisados os resultados apresentados ao utilizar o pacote *mirt* do *R*, aplicando os critérios de AIC, BIC e a verossimilhança conforme os resultados adquiridos da Tabela 22.

Na Tabela 22, os valores estimados do c_i foram iguais a zero tanto no MLD1 como MLD2 nos 2PLf ao utilizarem o método do perfilamento. Logo, os resultados da *logLik*, dos critérios AIC e BIC são iguais para 2PL e 2PLf. Ademais, tais critérios apontam o menor valor do modelo como melhor e, neste caso, em ambas dimensões é o

Tabela 22: Resultados das estatísticas de diagnósticos da qualidade dos 189 e 154 respondentes do ajuste dos modelos do conjunto de dados II

		189 respondentes			
MLD1		df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0$)		80	-3623,06	7406,13	7665,47
2PLf ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c$)		81"	-3623,06	7406,13	7665,47
3PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c_i$)		120	-3623,09	7486,19	7875,20
MLD2		df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0$)		120	-3560,55	7359,10	7744,87
2PLf ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c$)		121"	-3560,55	7359,10	7744,87
3PL ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c_i$)		160*	-3558,36	7434,72	7950,16
		154 respondentes			
MLD1		df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0$)		80	-3375,95	6911,90	7154,86
2PLf ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0,256$)		81	-3476,79	7113,57	7356,52
3PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c_i$)		120	-3361,84	6963,67	7328,11
MLD2		df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0$)		120	-3313,96	6865,93	7227,33
2PLf ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0,05$)		121	-3311,60	6861,2	7222,60
3PL ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c_i$)		160*	-3276,54	6871,08	7353,96

Nota:

- modelo logístico de um traço latente (MLD1);
- modelo logístico de dois traços latentes (MLD2);
- dois parâmetros logísticos (2PL) ;
- dois parâmetros logísticos fixando o c_i (2PLf);
- três parâmetros logísticos (3PL); e
- quantidade de parâmetros (df).

Os valores com (") possuem a quantidade de parâmetros onde os valores fixo de c e c_i são estimados iguais a zero.

Os valores com (*) representam que a quantidade de parâmetros é maior do que o total de respondentes.

2PL. Portanto, a indicação é para um modelo mais simples.

Sabendo que existem 155 padrões de resposta dos 189 indivíduos que responderam à multidisciplinar, não é conveniente abordar o modelo MLD3. Conforme Reckase (2009), a quantidade dos padrões de resposta e o total dos respondentes devem ser maiores do que a quantidade de parâmetros. Os resultados da Tabela 22 destacados com (*) são valores em que a quantidade de parâmetros ultrapassa a quantidade de respondentes, o qual restringe na escolha do modelo.

Como os modelos são acumulativos, o teste da razão da verossimilhança é utilizado para verificar se compensa um modelo com mais parâmetros, dada a *log-verossimilhança* aproximada. No MLD2 dos 154 respondentes, em ambos os modelos, 2PL e 2PLf, há pouca diferença na *logLik*, dada a quantidade de parâmetros de 120 a 121, respectivamente. Ao se aplicar o teste da razão da verossimilhança com 95% de confiança, escolhe-se o modelo com 2PLf com mais parâmetros. Ao se comparar o 2PLf com 3PL, o teste também

mostrou que compensa a escolha do modelo 3PL. Vale ressaltar que a quantidade de parâmetros é maior que os padrões de respondentes. Portanto, para o conjunto com 154 respondentes, são viáveis o modelo 2PLf com MLD2 e os critérios de AIC e BIC, mas é necessário, ainda, que se abordem outras análises subsequentes, antes da tomada de decisão.

Para os 189 respondentes, os critérios AIC e BIC indicam o modelo 2PL ou 2PLf por possuírem o mesmo resultado. Apesar da *logLik* indicar 3PL, o teste da razão da verossimilhança entre 2PLf e 3PL indica que não compensa o modelo com mais parâmetros. Outro fator importante, mesmo com as indicações feitas pela Tabela 22, é a necessidade de que se verifiquem as estimativas dos parâmetros apresentadas na subseção 4.2.2.1.

4.2.2.1 Estimação dos parâmetros nos modelos da TRIM no conjunto de dados II

No MLD1 para os 189 respondentes, a estimação dos parâmetros do 3PL, visto nos critérios de AIC e BIC, aproximou-se do 2PL e 2PLf porque o c_i , estimado para o 3PL somente no item da Q33, que pertence a área específica de administração, é de $c_i = 0,001$ e para as outras questões todos os c_i foram estimados iguais a zero, desta maneira, os valores das *logLik* também são muito próximos. Por esse motivo, os valores das estimativas do MLD1 de 2PL e 2PLf também são iguais para as questões do 3PL, com exceção da Q33, mas a aproximação é quase imperceptível. Assume-se, então, que são iguais devido a aproximação ser centesimal, conforme mostra a Tabela 23.

Para o MLD2 no 3PL a mesma situação se repete para a questão Q33 cujo $c_i = 0,051$. Como no 2PLf o acerto ao acaso também foi perfilhado em zero, assume-se que o valor para o 3PL aproxima-se de zero.

Na Tabela 23, ao serem retirados os alunos que zeraram, os 154 respondentes, quanto aos valores do acerto ao acaso, foram diferentes de zero quando estimados. No modelo MLD1 em 2PLf, o acerto ao acaso teve o valor fixo perfilhado e estimado em 0,26. Para o 3PL do MLD1, é possível também verificar que existem valores para o acerto ao acaso em que os alunos não “chutaram”. A Q40 em que $c_i=0$ e nas Q5 e Q10, o valor estimado do $c_i=0,69$, ou seja, esses dois itens indicam, conforme o modelo, os valores estimados com maiores acertos ao acaso.

No modelo MLD2, o valor do acerto ao acaso para o 2PLf, fixado e que utiliza o perfilhamento do modelo estimado, é de 0,05 menor do que foi proposto pela TRIU. Não se recomenda utilizar o 3PL na TRIM, pois conforme a Tabela 22, o número de

Tabela 23: Resultados das estimativas dos parâmetros dos MLD1 e MLD2 para os 189 e 154 respondentes do CD2

Q	189 respondentes												154 respondentes															
	MLD1			MLD2			MLD1						MLD2															
	2PL/2PLf/3PL	2PL/2PLf/3PL	2PL/2PLf/3PL	2PL	2PLf	3PL	2PL	2PLf	3PL	2PL	2PLf	3PL	2PL	2PLf	3PL													
α_i	d_i	c_i	α_{i1}	α_{i2}	d_i	c_i	α_{i1}	d_i	c_i	α_{i1}	d_i	c_i	α_{i1}	α_{i2}	d_i	c_i	α_{i1}	α_{i2}	d_i	c_i								
Q1	1.59	-0.15	0	0.79	-1.02	-0.17	0	0.36	0.55	0	0.46	0	0.26	0.96	-1.25	0.5	-0.05	-0.46	0.56	0	-0.1	-0.52	0.48	0.05	26.1	-3.53	-30.27	0.57
Q2	0.76	-0.7	0	0	-0.93	-0.74	0	-0.4	-0.27	0	-0.84	-1.28	0.26	-0.52	-0.28	0	0.61	-0.07	-0.28	0	0.68	-0.13	-0.43	0.05	36.72	-9.99	-61.84	0.39
Q3	1.37	-1.43	0	0.94	-0.54	-1.43	0	0.39	-0.79	0	0.52	-2.4	0.26	0.38	-0.79	0	-0.31	-0.21	-0.79	0	-0.29	-0.28	-0.97	0.05	-0.09	0.45	-0.8	0
Q4	2.19	-0.29	0	1.67	-0.74	-0.31	0	0.68	0.62	0	0.93	0.04	0.26	0.83	0.3	0.16	-0.59	-0.33	0.62	0	-0.64	-0.27	0.53	0.05	0.12	0.7	0.62	0
Q5	2.98	0.62	0	1.95	-1.37	0.58	0	1.09	1.83	0	1.13	1.39	0.26	2.56	-0.18	0.65	-0.7	-0.82	1.84	0	-0.87	-0.67	1.77	0.05	3.17	3.76	0.07	0.62
Q6	2.99	-0.1	0	2.11	-1.09	-0.09	0	1.05	1.11	0	0.83	0.56	0.26	0.99	1.09	0	-0.79	-0.56	1.1	0	-0.69	-0.49	0.98	0.05	0.25	0.96	1.09	0
Q7	1.75	0.02	0	1.44	-0.45	0.02	0	0.42	0.76	0	0.28	0.26	0.26	0.41	0.76	0	-0.46	-0.07	0.77	0	-0.46	-0.07	0.69	0.05	0.13	0.44	0.77	0
Q8	1.32	-1.67	0	1.32	-0.07	-1.71	0	0.55	-1.08	0	17.5	-29.5	0.26	0.65	-1.11	0	-0.63	-0.08	-1.1	0	-0.67	-0.02	-1.33	0.05	0.21	0.56	-1.1	0
Q9	2.52	0.64	0	2.8	-0.26	0.72	0	0.83	1.65	0	0.81	1.22	0.26	0.9	1.68	0	-1.17	-0.07	1.83	0	-1.17	0.03	1.75	0.05	-0.22	7.48	-3.02	0.7
Q10	1.87	0.28	0	1.46	-0.61	0.27	0	0.32	1.04	0	0.41	0.6	0.26	2.16	-1.9	0.65	-0.36	-0.09	1.05	0	-0.34	-0.09	0.97	0.05	-0.02	0.96	-0.62	0.58
Q11	1.39	-0.66	0	0.7	-0.9	-0.69	0	0.29	-0.03	0	0.4	-0.77	0.26	0.28	-0.03	0	-0.05	-0.36	-0.03	0	-0.09	-0.38	-0.14	0.05	0.45	-0.02	-0.04	0
Q12	1.69	-1.09	0	1.45	-0.38	-1.1	0	0.6	-0.34	0	0.85	-1.42	0.26	1.12	-1.24	0.21	-0.63	-0.17	-0.35	0	-0.83	-0.17	-0.52	0.05	0	1.39	-1.12	0.17
Q13	0.91	-0.44	0	0.14	-0.95	-0.46	0	-0.34	0.05	0	-0.39	-0.64	0.26	-0.36	0.05	0	0.55	-0.1	0.06	0	0.53	-0.16	-0.05	0.05	0.09	-0.42	0.05	0
Q14	1.71	-1.21	0	1.31	-0.57	-1.23	0	0.58	-0.45	0	1.66	-2.18	0.26	3.35	-4.59	0.32	-0.5	-0.27	-0.45	0	-0.59	-0.23	-0.61	0.05	-0.07	2.12	-2.85	0.29
Q15	1.9	-2.12	0	1.9	-0.1	-2.2	0	0.86	-1.32	0	4.02	-7.12	0.26	3.66	-5.12	0.16	-1	-0.11	-1.38	0	-1.85	0.12	-2.2	0.05	2.82	28.8	-37.68	0.16
Q16	1.21	-0.34	0	0.12	-1.45	-0.38	0	0.02	0.24	0	0.25	-0.39	0.26	-0.01	-0.08	0.15	0.49	-0.61	0.27	0	0.36	-0.59	0.16	0.05	1	-0.8	-0.92	0.34
Q17	2.74	-0.6	0	2.13	-0.83	-0.6	0	1.16	0.57	0	1.48	-0.16	0.26	1.5	0.15	0.17	-0.97	-0.56	0.57	0	-1	-0.49	0.46	0.05	0.26	0.91	0.53	0
Q18	2.95	-0.32	0	1.5	-2.28	-0.43	0	1.08	0.9	0	1.45	0.38	0.26	1.12	0.9	0	-0.36	-1.42	1.03	0	-0.56	-1.25	0.9	0.05	0.95	0.84	0.92	0
Q19	2.74	0.29	0	1.65	-1.37	0.27	0	1	1.42	0	0.98	0.94	0.26	0.88	1.37	0	-0.53	-0.83	1.42	0	-0.68	-0.79	1.36	0.05	1.06	0.77	1.52	0
Q20	1.33	-0.24	0	1.52	0.06	-0.25	0	0.11	0.37	0	0.02	-0.2	0.26	0.1	0.32	0.03	-0.48	0.38	0.39	0	-0.37	0.34	0.29	0.05	-0.3	0.35	0.36	0.02
Q21	1.71	-1.09	0	2.02	0.11	-1.19	0	0.77	-0.35	0	1.52	-1.73	0.26	0.93	-0.39	0	-1.07	0	-0.39	0	-1.42	0.23	-0.61	0.05	0.22	2.69	-2.34	0.24
Q22	2.27	-0.09	0	1.21	-1.37	-0.11	0	0.86	0.88	0	0.82	0.35	0.26	0.64	0.83	0	-0.36	-0.86	0.91	0	-0.46	-0.85	0.82	0.05	1.01	0.58	0.38	0.24
Q23	1.56	-2.13	0	1.9	0.24	-2.31	0	0.73	-1.48	0	21.8	-42.2	0.26	1.68	-2.57	0.08	-1	0.02	-1.59	0	-1.29	-0.01	-2.06	0.05	0.24	1.48	-2.51	0.08
Q24	1.75	0.14	0	3.16	0.77	0.24	0	0.48	0.9	0	0.55	0.42	0.26	5.06	-3.37	0.59	-1.5	0.67	1.21	0	-1.4	0.72	1.1	0.05	-0.89	1.22	1.24	0
Q25	1.99	-3.53	0	0.62	-1.99	-3.99	0	0.81	-2.64	0	0.31	-13.1	0.26	0.59	-2.52	0	-0.06	-1.32	-3.05	0	-0.37	-2.05	-2.68	0.05	3.86	1.71	-7.77	0.04
Q26	0.45	-1.95	0	0.7	0.27	-2	0	-0.56	-1.74	0	-9.28	-21.6	0.26	-1.24	-3.01	0.09	0.03	0.86	-1.87	0	0.42	1.19	-2.54	0.05	-0.82	-0.3	-1.85	0
Q27	1.49	-0.7	0	0.28	-1.68	-0.8	0	0.32	-0.03	0	0.62	-0.84	0.26	0.94	-1.84	0.39	0.28	-0.88	-0.02	0	0.18	-1.09	-0.16	0.05	4.42	1.06	-4.85	0.39
Q28	2.01	-0.45	0	0.78	-1.79	-0.53	0	0.61	0.4	0	1.04	-0.25	0.26	0.7	-0.01	0.19	-0.01	-1.03	0.46	0	-0.18	-1.15	0.36	0.05	37.06	4.55	-21.95	0.43
Q29	1.94	-1.7	0	1.03	-1.24	-1.78	0	0.83	-0.87	0	19.3	-25.1	0.26	14.11	-17.97	0.25	-0.35	-0.86	-0.9	0	-0.68	-1	-1.22	0.05	18.23	15.2	-24.85	0.2
Q30	1.1	-2.12	0	-0.71	-2.49	-2.88	0	0.19	-1.56	0	0.21	-12.3	0.26	0.04	-1.66	0.02	0.83	-1.26	-2.08	0	3.54	-11.5	-14.9	0.05	37.6	-8.82	-61.03	0.11
Q31	2.48	-0.04	0	0.83	-2.9	-0.11	0	0.82	0.97	0	1.46	0.56	0.26	0.89	0.99	0	0.18	-1.9	1.37	0	-0.15	-2.19	1.38	0.05	40.22	0.85	-0.72	0.4
Q32	2.05	-0.82	0	1.2	-1.08	-0.84	0	0.78	0.06	0	1.15	-0.75	0.26	0.89	0.05	0	-0.4	-0.69	0.06	0	-0.59	-0.67	-0.05	0.05	0.66	0.63	0.04	0
Q33	0.09	-3.42	0	-0.5	-0.63	-3.54	0	-1.15	-3.78	0	-0.14	-13.3	0.26	-0.98	-3.86	0.01	1.31	0.37	-4	0	5.89	1.65	-14.6	0.05	3.53	-22.5	-44.68	0.01
Q34	1.94	0.75	0	1.23	-0.92	0.73	0	0.05	1.51	0	0.23	1.14	0.26	0.17	1.44	0.06	0	-0.08	1.51	0	0.01	-0.11	14.5	0.05	3.41	-3.38	-8.86	0.81
Q35	0.6	-2.2	0	0.5	-0.16	-2.21	0	-0.31	-1.88	0	-0.21	-12.7	0.26	-1.24	-3.78	0.1	0.1	0.29	-1.88	0	0.35	0.42	-2.44	0.05	-12.55	0.33	-18.8	0.08
Q36	0.93	-2.34	0	0.74	-0.27	-2.35	0	0.07	-1.85	0	-0.03	-12.7	0.26	0.12	-1.88	0	-0.17	0.06	-1.86	0	-0.34	0.27	-2.38	0.05	-10.48	18	-38.5	0.1
Q37	1.3	-2.71	0	1.19	-0.15	-2.73	0	0.57	-2.14	0	14.6	-33.9	0.26	3.39	-7.36	0.1	-0.59	-0.11	-2.16	0	-1.31	2.41	-22.3	0.05	0.11	22.3	-35.61	0.07
Q38	1.39	-1.92	0	1.58	0.14	-2.02	0	0.58	-1.31	0	20.9	-36.3	0.26	1.98	-3.4	0.15	-0.8	0.04	-1.38	0	-0.88	0	-1.7	0.05	0.89	1.86	-3.57	0.15
Q39	0.88	-2.85	0	0.57	-0.39	-2.86	0	-0.05	-2.39	0	-0.02	-13.1	0.26	0.02	-3.81	0.06	-0.01	0.09	-2.39	0	-0.11	0.21	-3.29	0.05	-1.6	0.64	-3.35	0
Q40	1.37	-1.77	0	1.61	0.18	-1.86	0	0.42	-1.12	0	0.83	-4.19	0.26	0.38	-1.13	0	-0.76	0.23	-1.22	0	-0.76	0.42	-1.49	0.05	-25.44	22.1	-41.37	0.17

parâmetros estimados é maior do que os padrões de resposta. O motivo de se manter o 3PL do MLD2 na Tabela 23 é para que se verifique o comportamento dos parâmetros estimados. E, mesmo que não seja recomendado, verifica-se que, de fato, os valores dos parâmetros foram superestimados, justificando o motivo da não utilização desse modelo.

Outra observação importante é que, ao se analisar d_i como um parâmetro relacionado à dificuldade do item, medido na mesma escala da habilidade, ele é o escalar que representa $d_i = -\alpha_i \cdot \beta_i$. O valor mais alto de d_i em relação aos outros itens da Tabela 23 está no 3PL do MDL1 e corresponde ao da Q29 para os 154 respondentes. É uma questão da área de exatas que ignora o 3PL do MDL2. Para compreender melhor os itens, aborda-se a CCI na subseção 4.2.2.2.

4.2.2.2 CCI e cargas fatoriais dos itens na TRIM no conjunto de dados II

A curva característica do item é a base da construção da TRI e todas as outras construções dependem dela. Para dar sequência à discussão levantada do item Q29, observa-se no MLD1/3PL com 154 respondentes, o comportamento dos parâmetros pela CCI. O valor de $d_i = -17,97$, cuja dificuldade $\beta_i = 1,27$ é apresentada na Tabela 23. Portanto, o sinal negativo dos parâmetros, inclusive o do exemplo da Q29, representa na

CCI que a questão é difícil, conforme a Figura 17(a).

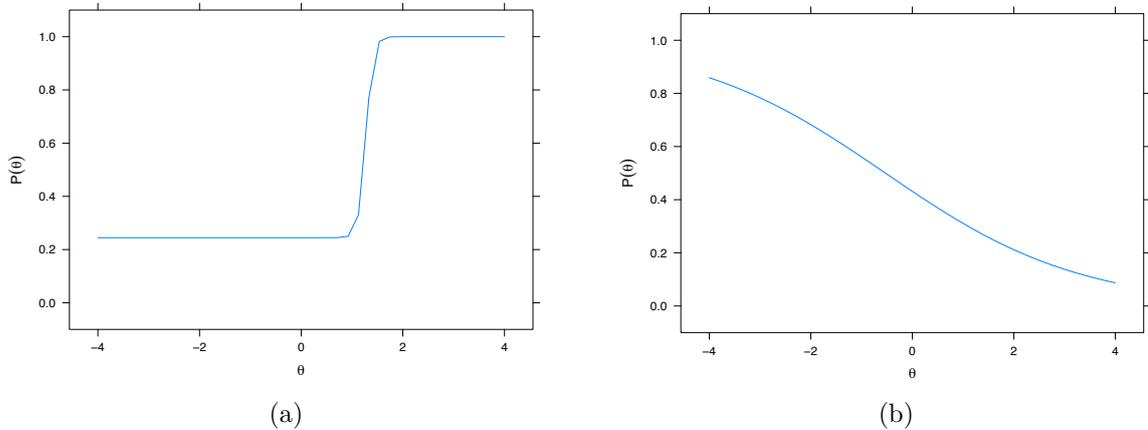


Figura 17: CCI dos 154 respondentes da Q29 e Q2 do MLD1/3PL

Diferente da questão Q2 do 3PL do MLD1, conforme a Figura 17(b), quando ambos os parâmetros possuem o mesmo sinal negativo, a CCI com $a < 0$ mostra que o item é inconsistente, com a possibilidade de acerto ao acaso, ou seja, indica necessidade de análise mais detalhada desse item devido aos sinais estimados nos parâmetros pelo pacote *mirt*. Nessas análises, há a necessidade de especialistas da área para rever o motivo do item apresentar esse decrescimento, para que seja excluído, como mostra a Figura 17 e também os valores negativos dos parâmetros da Tabela 23.

Para observar o comportamento de todos os itens nas CCI e os três parâmetros estimados do modelo unidimensional, seguem os gráficos da Figura 18 para os 189 e 154 respondentes do exame multidisciplinar de administração para o MLD1 com 3PL.

Para que se detalhem as informações necessárias do comportamento dos 40 itens respondidos e se tenha uma visualização mais clara, os itens foram divididos em 2 grupos de 1 a 20 e de 21 a 40. Leva-se em consideração que as questões da área de exatas são 9, entre as Q24 a Q32, os demais itens são da área específica de administração.

Na Figura 18, o comportamento das CCI dos 189 respondentes são representados pelos 20 primeiros itens, na Figura (a), que apresentaram comportamento íngreme no formato “S” e pelos itens entre 21 a 40, na Figura (b), com dificuldade nas habilidades e quase nada de discriminação, no item Q33. Para todos os avaliados, o acerto ao acaso não apresentou valores significativos.

A CCI dos 154 respondentes na Figura 18 (c), representado pelos itens 1 a 20 e a Figura (d) dos itens 21 a 40, apresentaram o acerto ao acaso para vários itens com a probabilidade acima de 0.5 como as questões Q5 e Q24. Nas Figuras (c) como a (d),

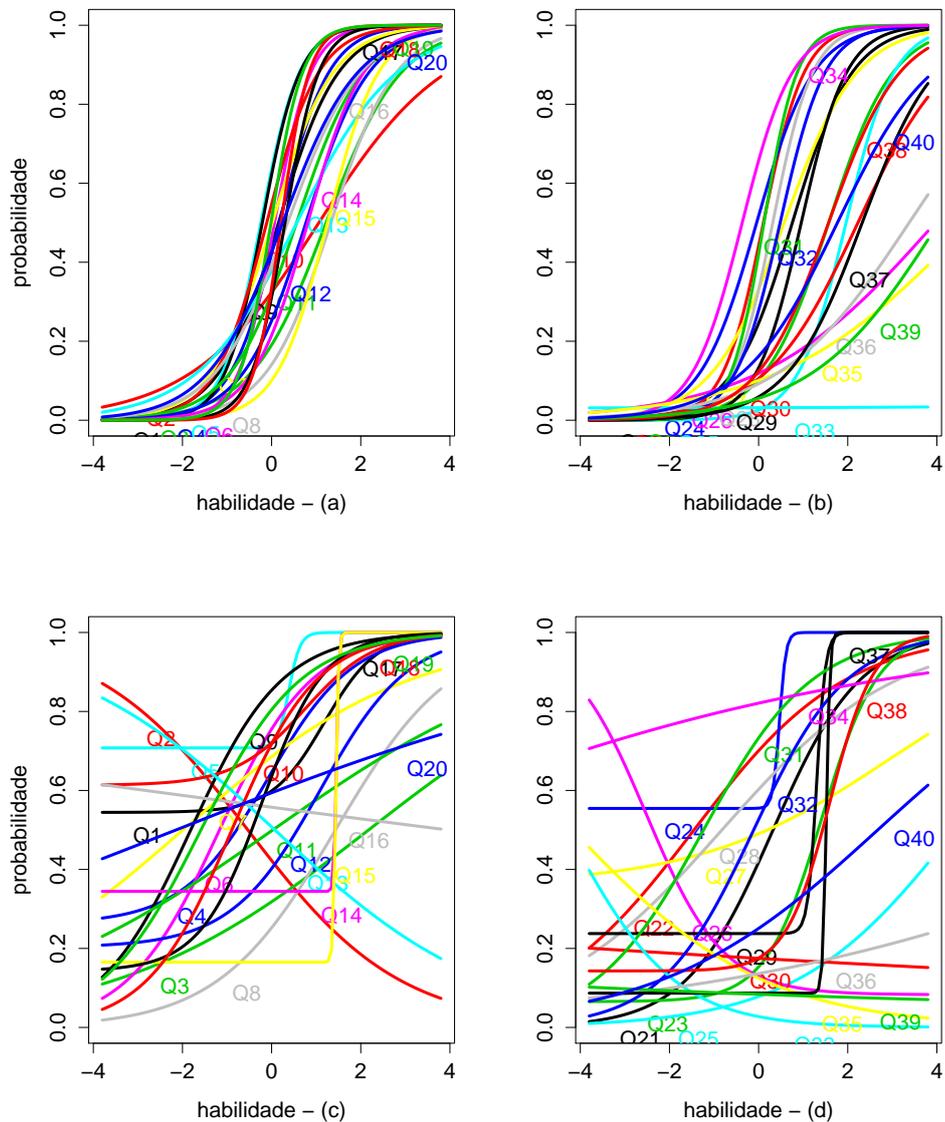


Figura 18: CCI dos 189 e 154 respondentes do MLD1- 3PL
 Nota: Os 189 e 154 respondentes quanto aos itens de 1 a 20 (a e c) e do 21 a 40 (b e d), respectivamente.

é possível detectar itens sem discriminação, decrescentes e com a CCI bem íngreme. Logo, mostra-se que, ao serem retirados os alunos que zeraram para o mesmo modelo, o MLD1 com 3PL, as diferenças foram visualmente observadas no comportamento dos itens, indicando a influência dos zeros inflacionados no modelo para análise.

Segue a Figura 19 que mostra a influência dos zeros no modelo ajustado.

Como as CCI para os 154 respondentes apresentaram, em vários itens, um comportamento diferenciado dos 189 respondentes, verificou-se o quanto o modelo é capaz de receber as influências dos 35 zeros.

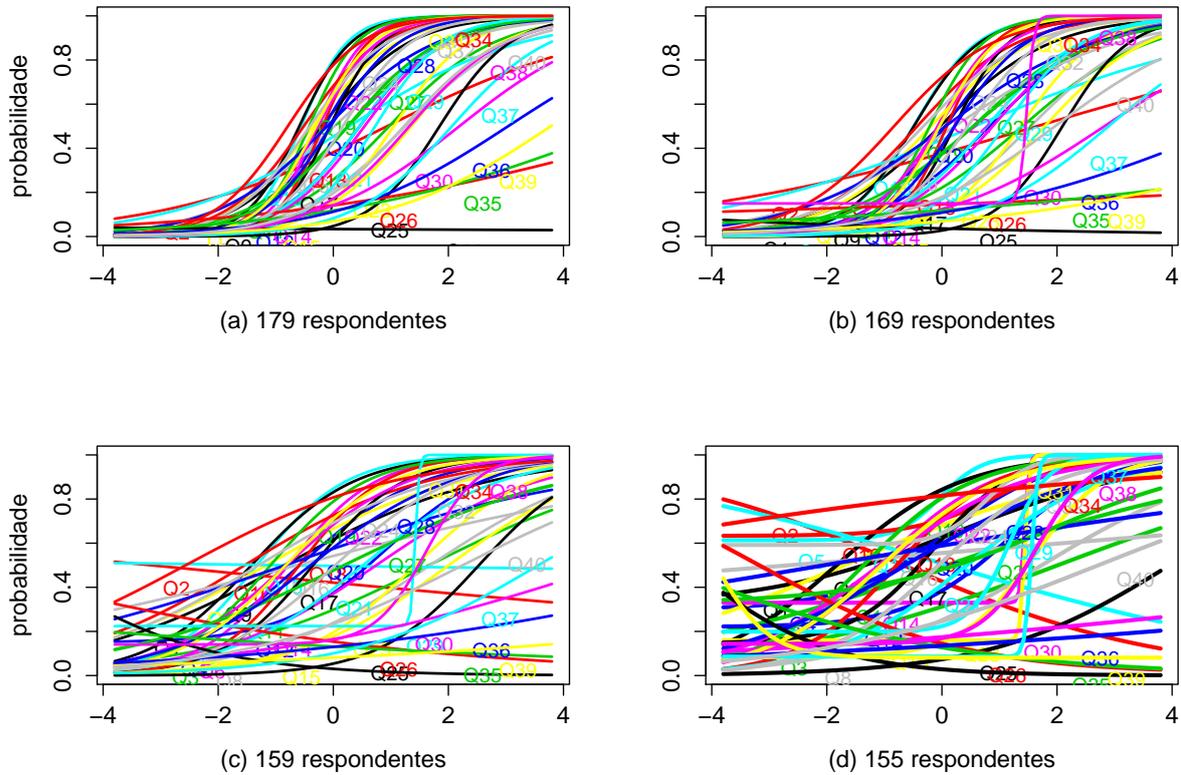


Figura 19: Influências dos zeros do MLD1- 3PL na CCI

Primeiramente, na Figura 19 (a), retirou-se somente 10 zeros, escolhidos ao acaso, ou seja, avalia-se 179 respondentes. Na Figura (b) excluiu-se 20 zeros resultando em 169 respondentes. Para a Figura (c) retirou-se 30 zeros e na Figura (d), eliminou-se 34 zeros e obteve-se 155 respondentes.

Na Figura 19 (a) e (b), observa-se que quando os alunos que zeraram são gradativamente retirados da prova, as CCIs não influenciam consideravelmente o MLD1-3PL. Em contrapartida, ao retirar 30 alunos dos 35 que são o total que zeraram no exame, ou como mostra a Figura (d) excluídos 34 alunos, as influências dos zeros já apresentaram diferenças nas CCI. Essas análises necessitam de investigações mais detalhadas nas estimações dos parâmetros, conforme os sinais e as influências dos zeros, nos modelos em relação aos respondentes.

Ao se aplicar a TRIM, é necessário analisar e verificar os resultados antes da tomada de decisão de qual modelo adotar. Por esse motivo, os carregamentos dos fatores de explicação fazem parte das análises. No pacote *mirt*, conforme Nojosa (2002), Reckase (2009) e Chalmers (2012), as cargas fatoriais possibilitam informar e também verificar

quais itens se relacionam a cada fator. Assim, quanto maior a carga fatorial melhor é a relação entre item e o fator. Os resultados, após a rotação oblíqua, das cargas fatoriais dos modelos indicados na TRIM, estão apresentados na Tabela 24.

Tabela 24: Cargas fatoriais e comunalidades dos MLD1 e MLD2 do CD2

Q	189 respondentes						154 respondentes														
	MLD1			MLD2			MLD1						MLD2								
	2PL/2PLf/3PL			2PL/2PLf/3PL			2PL		2PLf		3PL		2PL			2PLf			3PL		
	F1	h2		F1	F2	h2	F1	h2	F1	h2	F1	h2	F1	F2	h2	F1	F2	h2	F1	F2	h2
Q1	0,68	0,47	0,68	-0,12	0,48	0,21	0,04	0,26	0,07	0,49	0,24	-0,2	-0,17	0,07	-0,19	-0,24	0,09	0,88	0,48	1	
Q2	0,41	0,17	0,41	-0,25	0,23	-0,23	0,05	-0,44	0,2	-0,29	0,09	0,24	-0,24	0,11	0,3	-0,22	0,14	0,93	0,36	1	
Q3	0,63	0,39	0,62	0,01	0,39	0,23	0,05	0,29	0,08	0,22	0,05	-0,22	0,02	0,05	-0,22	-0,07	0,05	-0,2	0,17	0,07	
Q4	0,79	0,62	0,79	0,06	0,63	0,37	0,14	0,48	0,23	0,44	0,19	-0,38	0,06	0,14	-0,38	0,02	0,14	-0,18	0,34	0,15	
Q5	0,87	0,75	0,87	-0,02	0,76	0,54	0,29	0,55	0,31	0,83	0,69	-0,54	-0,08	0,3	-0,54	-0,11	0,3	0,04	0,94	0,89	
Q6	0,87	0,76	0,86	0,04	0,75	0,52	0,28	0,44	0,19	0,5	0,25	-0,51	0,04	0,26	-0,45	-0,08	0,2	-0,2	0,46	0,25	
Q7	0,72	0,51	0,71	0,1	0,52	0,24	0,06	0,16	0,03	0,24	0,06	-0,23	0,13	0,07	-0,25	0,08	0,07	-0,1	0,24	0,07	
Q8	0,61	0,37	0,6	0,18	0,39	0,3	0,09	1	0,99	0,36	0,13	-0,3	0,18	0,12	-0,34	0,14	0,14	-0,1	0,31	0,11	
Q9	0,83	0,69	0,84	0,23	0,75	0,44	0,19	0,43	0,18	0,47	0,22	-0,47	0,32	0,32	-0,51	0,25	0,32	-0,63	0,75	0,95	
Q10	0,74	0,55	0,74	0,06	0,55	0,19	0,03	0,23	0,05	0,79	0,62	-0,2	0,09	0,05	-0,2	0,03	0,04	-0,31	0,38	0,24	
Q11	0,63	0,4	0,64	-0,11	0,42	0,17	0,03	0,23	0,05	0,16	0,03	-0,16	-0,14	0,04	-0,15	-0,17	0,05	0,21	0,14	0,07	
Q12	0,7	0,5	0,7	0,12	0,5	0,33	0,11	0,45	0,2	0,55	0,3	-0,34	0,14	0,13	-0,43	0,11	0,2	-0,39	0,5	0,4	
Q13	0,47	0,22	0,47	-0,23	0,27	-0,19	0,04	-0,22	0,05	-0,2	0,04	0,21	-0,23	0,09	0,23	-0,2	0,09	0,19	-0,16	0,06	
Q14	0,71	0,5	0,71	0,05	0,51	0,32	0,1	0,7	0,49	0,89	0,79	-0,32	0,06	0,11	-0,35	0,02	0,13	-0,51	0,6	0,61	
Q15	0,74	0,55	0,72	0,22	0,57	0,45	0,2	0,92	0,85	0,91	0,82	-0,44	0,27	0,26	-0,64	0,36	0,54	-0,54	0,84	1	
Q16	0,58	0,34	0,59	-0,32	0,45	0,01	0	0,14	0,02	-0,01	0	-0,02	-0,4	0,16	0,03	-0,37	0,14	0,61	-0,02	0,37	
Q17	0,85	0,72	0,85	0,08	0,72	0,56	0,32	0,66	0,43	0,66	0,44	-0,56	0,09	0,32	-0,55	-0,01	0,3	-0,18	0,44	0,23	
Q18	0,87	0,75	0,88	-0,18	0,8	0,54	0,29	0,65	0,42	0,55	0,3	-0,55	-0,36	0,44	-0,49	-0,4	0,4	0,12	0,58	0,35	
Q19	0,85	0,72	0,85	-0,06	0,72	0,51	0,26	0,5	0,25	0,46	0,21	-0,49	-0,14	0,26	-0,49	-0,2	0,28	0,18	0,58	0,36	
Q20	0,62	0,38	0,61	0,24	0,43	0,07	0	0,01	0	0,06	0	-0,07	0,32	0,11	-0,1	0,26	0,08	-0,26	0,05	0,07	
Q21	0,71	0,5	0,7	0,28	0,57	0,41	0,17	0,67	0,44	0,48	0,23	-0,42	0,33	0,28	-0,53	0,36	0,41	-0,47	0,7	0,71	
Q22	0,8	0,64	0,8	-0,11	0,65	0,45	0,2	0,43	0,19	0,35	0,12	-0,44	-0,21	0,24	-0,41	-0,28	0,25	0,22	0,51	0,31	
Q23	0,67	0,45	0,65	0,31	0,51	0,39	0,16	1	0,99	0,7	0,49	-0,39	0,32	0,25	-0,55	0,25	0,36	-0,32	0,58	0,44	
Q24	0,72	0,51	0,72	0,45	0,72	0,27	0,07	0,31	0,09	0,95	0,9	-0,31	0,61	0,46	-0,41	0,54	0,45	-0,64	0,19	0,45	
Q25	0,76	0,58	0,77	-0,28	0,67	0,43	0,18	0,18	0,03	0,33	0,11	-0,44	-0,43	0,38	-0,47	-0,62	0,6	0,46	0,8	0,86	
Q26	0,26	0,07	0,24	0,21	0,1	-0,31	0,1	-0,98	0,97	-0,59	0,35	0,32	0,32	0,21	0,43	0,41	0,36	-0,25	-0,37	0,2	
Q27	0,66	0,43	0,67	-0,31	0,55	0,19	0,03	0,34	0,12	0,48	0,23	-0,19	-0,42	0,22	-0,17	-0,51	0,29	0,6	0,71	0,87	
Q28	0,76	0,58	0,78	-0,23	0,66	0,34	0,11	0,52	0,27	0,38	0,14	-0,36	-0,38	0,27	-0,33	-0,46	0,32	0,73	0,68	1	
Q29	0,75	0,56	0,76	-0,12	0,59	0,44	0,19	1	0,99	0,99	0,99	-0,44	-0,21	0,24	-0,51	-0,28	0,34	0,23	0,97	0,99	
Q30	0,54	0,3	0,53	-0,57	0,61	0,11	0,01	0,12	0,01	0,02	0	-0,09	-0,64	0,42	-0,17	-0,97	0,98	0,92	0,39	1	
Q31	0,82	0,68	0,84	-0,32	0,81	0,43	0,19	0,65	0,42	0,46	0,22	-0,45	-0,59	0,55	-0,41	-0,68	0,63	0,79	0,61	1	
Q32	0,77	0,59	0,77	-0,06	0,6	0,42	0,17	0,56	0,31	0,46	0,21	-0,41	-0,14	0,19	-0,44	-0,18	0,22	0,08	0,46	0,22	
Q33	0,05	0	0,04	-0,28	0,08	-0,56	0,31	-0,08	0,01	-0,5	0,25	0,59	-0,24	0,4	0,95	-0,16	0,93	0,73	-0,68	0,99	
Q34	0,75	0,56	0,75	-0,03	0,57	0,03	0	0,13	0,02	0,1	0,01	-0,03	-0,03	0	-0,02	-0,06	0	0,94	-0,12	0,89	
Q35	0,33	0,11	0,33	0,04	0,11	-0,18	0,03	-0,12	0,02	-0,59	0,35	0,16	0,08	0,03	0,28	0,13	0,1	-0,82	-0,56	0,98	
Q36	0,48	0,23	0,48	0,05	0,23	0,04	0	-0,02	0	0,07	0	-0,05	0,09	0,01	-0,11	0,22	0,06	-0,92	0,37	0,99	
Q37	0,61	0,37	0,59	0,15	0,36	0,32	0,1	0,99	0,99	0,89	0,8	-0,3	0,15	0,11	-0,81	0,58	0,98	-0,62	0,78	0,99	
Q38	0,63	0,4	0,61	0,26	0,44	0,32	0,1	1	0,99	0,76	0,58	-0,32	0,28	0,18	-0,42	0,19	0,21	-0,17	0,75	0,59	
Q39	0,46	0,21	0,45	-0,01	0,21	-0,03	0	-0,01	0	0,01	0	0,03	0,04	0	0	0,14	0,02	-0,69	-0,18	0,51	
Q40	0,63	0,39	0,6	0,27	0,44	0,24	0,06	0,44	0,19	0,22	0,05	-0,24	0,34	0,17	-0,26	0,36	0,2	-1	0,07	1	

Nota: Questões (Q); F1 carga de um fator; F2 carga de dois fator; h2 comunalidade; modelo logístico de um traço latente (MLD1); modelo logístico de dois traços latentes (MLD2); dois parâmetros logísticos (2PL) e fixando o ci (2PLf); três parâmetros logísticos (3PL).

Na Tabela 24, é possível verificar que, para os 189 respondentes, quando analisadas as cargas fatoriais do MLD2, somente a Q30 carregou um pouco melhor na segunda dimensão e faz parte de uma das questões da área de exatas, ou seja, para essa prova as questões foram explicadas no F1. Assim, a indicação foi comprovada desde a Tabela 22 e, pelo teste da razão da verossimilhança, os modelos escolhido são os MLD1 com 2PL/2PLf/3PL, pois apresentaram a mesma estimação para o acerto ao acaso. Logo, devido à quantidade de df, escolheu-se o modelo com o menor parâmetro, neste caso o 2PL.

Os itens Q26 e Q33 têm carga fatorial menor do que 0,30 e comunalidade menor que 0,15 no MLD1 - 2PL com relação às cargas fatoriais e à comunalidade, seguindo

os pré-requisitos informados por (PASQUALI, 2003, p 166). Para o modelo sugerido, ao serem retirados os itens indicados, não houve indicação de outro modelo. Como o procedimento verifica as perspectivas e os desafios no exame, não foram retirados os itens para as próximas análises. O MLD2 dos 189 respondentes não foi indicado, porque quase todos os itens explicaram e carregaram no F1, contudo, não se justifica a escolha de um modelo com mais dimensões, sendo que os itens indicam um fator.

Ao serem retirados os alunos que zeraram a prova multidisciplinar, os 154 respondentes no MLD1 apresentaram as cargas que se justificam melhor no F1 do modelo 3PL. Logo, no MLD2, tanto no 2PL como em 2PLf, mais de 10 questões não tiveram explicação no F1 e nem no F2. Todavia, em 3PL, o valor dos parâmetros estimados é maior do que os dos respondentes, assim, é inviável a indicação para o 3PL e também pelos critérios do AIC e BIC da Tabela 22.

Conforme a Tabela 24, a Q30 foi a única do MLD2 dos 189 respondentes que carregou no F2 como nos 154 respondentes. As questões 24, 25, 27, 28, 30 e 31 tiveram suas cargas explicadas no F2 e este fator representa a área de exatas, a Q16 carregou também no F2 mas é um item da área de administração. Assim, para representar o comportamento das cargas fatoriais em duas dimensões e comparar os modelos, seguem na Figura 20, as CCI da Q30 no MDL2 dos 189 no 2PLf e dos 154 respondentes nos 2PLf.

Conforme a Figura 20, a Q30 é um item em que os alunos necessitam de duas habilidades para responder devido ao comportamento da CCI. Na Figura (a) foram necessárias as duas habilidades, enquanto na (b) a segunda habilidade explicou melhor o item. Em ambas as figuras a inclinação indica que a Q30 é difícil de ser respondida, pertence à área de exatas e teve comportamentos distintos quando tratou os 189 respondentes e ao excluir os 35 alunos que zeraram.

Como a prova do curso de administração é multidisciplinar, ao rever o item, há indícios do carregamento fatorial para a segunda habilidade por ser uma questão que exige conhecimento matemático e observou-se todos os itens que carregaram no F2 dos 154 e realmente são questões que necessitam da habilidade de matemática. Para definir a habilidade dessa questão nos 189 respondentes, há indicativos de que o modelo pode não ser tão representativo por ter sido mascarado com os respondentes que zeraram na prova. Assim, aconselha-se uma análise detalhada dos itens por professores das áreas teóricas de administração e de matemática, visto que a prova abordava essas duas habilidades.

Vale ressaltar que é necessário basear-se em um dos critério, AIC, BIC e $\log Lik$, que seriam penalizados ao usar apenas quando o número de parâmetros é o mesmo. Para

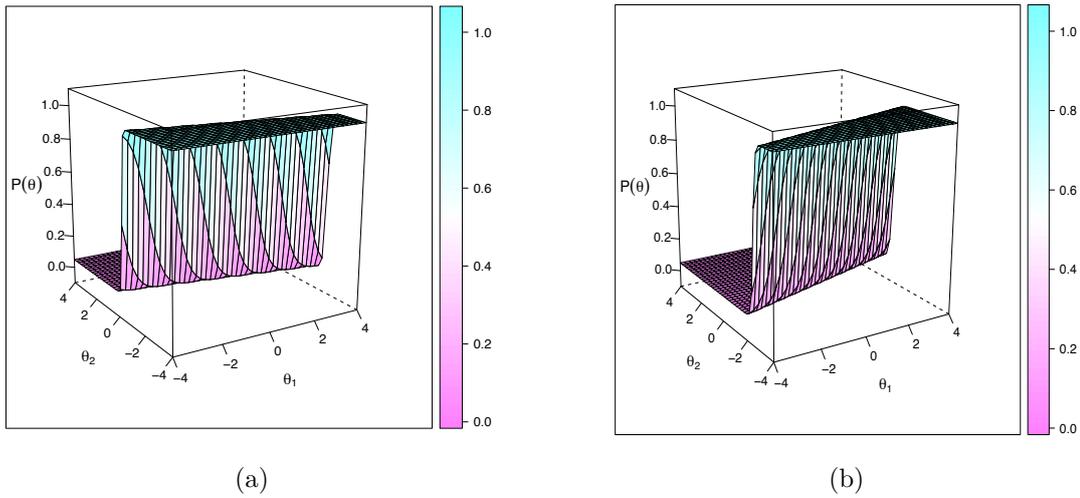


Figura 20: CCI da Q30 dos 189 (a) e 154 (b) respondentes MLD2-2PLf

indicar o modelo, no CD2 com 189 respondentes, seria o MLD2/2PL, mas ao abordar as cargas fatoriais explicadas no primeiro fator, verificou-se somente o item da Q30 que explicou-se um pouco mais no fator 2. Posto isto, os resultados não justificam o modelo multidimensional com dois traços latentes, visto que somente um item carrega na segunda habilidade. Por esse motivo, há a necessidade de se analisar os escores da TRIM utilizando o modelo unidimensional 2PL para os 189 respondentes e ao retirar os alunos que zeraram, o modelo com dois traços latentes com 2PLf, visto que quase todos os itens da área de matemática explicaram no F2.

Pelos critérios AIC, BIC, de acordo com a *logLik* e a quantidades de parâmetros estimados, foi indicado o 2PL para 189 e 2PLf para os 154 respondentes. Com a indicação dos modelos para ambos os respondentes, segue a subseção 4.2.2.3 com as discussões e os resultados dos escores do exame multidisciplinar do segundo período de administração.

4.2.2.3 Escores das habilidades da TRIM no conjunto de dados II

A fim de comparar a pontuação da TCT, que foi o modelo do exame multidisciplinar, utilizou-se a escala padronizada proposta pela metodologia na subseção 3.2.8. Nos 189 respondentes para o modelo indicado de um traço latente, bastou aplicar a equação 3.7 do MLD1 e assim padronizar, para cada respondente, o escore estimado pelo modelo. Os 154 avaliados tiveram seus escores individuais em duas habilidades, pois o modelo sugerido é de duas dimensões. Assim, para cada respondente, também há a escala padronizada dada a Equação 3.8 e a dimensão da habilidade avaliada.

Outra forma de abordagem é de como tratar os traços latentes e obter uma única

nota, que represente e seja viável à interpretação dos escores e na escala do conhecimento. Neste caso, o CD2 e as duas habilidades dos 154 avaliados sugerem a aplicação da escala padronizada multidimensional da equação 3.21 para representar a nota ponderada por habilidades.

Tabela 25: Escores e escala padronizada dos 189 e 154 respondentes

Al	Total Acertos	Nota TCT	189 respondentes		154 respondentes				
			Escore MLD1-2PL	Escala Padronizada	Escore F1	Escore F2	Escala Padonizada	EPM	
1	27	6,8	1,34	0,64	-1,875	-1,026	0,83	0,71	0,78
2	8	2	-0,33	0,31	1,41	0,93	0,16	0,22	0,18
3	20	5	0,66	0,5	-1,344	-0,046	0,72	0,47	0,62
4	16	4	0,28	0,43	0,394	0,165	0,37	0,41	0,39
5	22	5,5	0,84	0,54	-0,85	-0,647	0,62	0,62	0,62
6	23	5,8	0,77	0,52	-0,203	-1,223	0,49	0,76	0,6
7	21	5,3	0,66	0,5	-0,564	-0,4	0,56	0,56	0,56
8	22	5,5	0,79	0,53	-1,639	-0,494	0,78	0,58	0,7
9	14	3,5	0,13	0,4	0,518	0,135	0,34	0,42	0,37
10	13	3,3	0,05	0,38	0,694	0,218	0,31	0,4	0,34
11	21	5,3	0,61	0,49	-0,078	-0,306	0,46	0,53	0,49
12	22	5,5	0,7	0,51	-1,039	0,688	0,66	0,28	0,51
13	0	-	-1,87	0	-	-	-	-	-
14	0	-	-1,87	0	-	-	-	-	-
15	17	4,3	0,35	0,44	-0,428	0,495	0,54	0,33	0,45
16	12	3	-0,16	0,34	1,528	1,12	0,14	0,17	0,15
17	22	5,5	0,63	0,5	-0,541	0,656	0,56	0,29	0,45
18	23	5,8	0,97	0,57	-1,64	-1,02	0,78	0,71	0,75
19	20	5	0,54	0,48	-0,605	0,059	0,57	0,44	0,52
20	21	5,3	0,71	0,51	-0,993	0,286	0,65	0,38	0,54

Nota: Itens explicados pelas cargas fatoriais (F1) e (F2); Alunos (Al); Escala Padronizada Multidimensional (EPM)

A Tabela 25 aborda os vinte primeiros alunos, com a proposta dos resultados da escala padronizada e a possibilidade da utilização de duas formas: uma empregando as escalas padronizadas individuais para cada habilidade e a outra seria calcular uma média ponderada e obter a escala padronizada multidimensional que represente as duas habilidades. Com esse argumento, é possível a comparação com os resultados pela TCT e seguem os resultados completos dos avaliados do CD2 no Apêndice G. Os comandos para efetuar as análises constam no Apêndice D, com linguagem *R*.

Primeiramente, como são diversas as abordagens para as discussões da Tabela 25 e que impossibilitam comentar todos os avaliados, delimitam-se os resultados exemplificando os alunos 5, 8, 12 e 17 com o mesmo total de acertos - 22 itens - e a mesma nota pela TCT - 5,5. Caso a nota fosse baseada pela TRI e a indicação do modelo seguisse as análises abordadas, os escores difeririam entre si. Com a escala padronizada no MLD1/2PL para 189 respondentes, a comparação é diretamente conveniente 0,54; 0,53; 0,51; 0,50 respectivamente aos alunos como exemplo, ou seja, o aluno 5 desempenhou

melhor do que seus concorrentes perante ao mesmo acerto.

Ao excluir os alunos que zeraram, consideram-se duas habilidades na prova: as específicas de administração com as cargas fatoriais explicadas no (F1) e a da área de exatas no fator (F2). Percebe-se, assim, que o aluno 5 possui a mesma habilidade para ambas, conforme os itens que acertou, o aluno 8, com a habilidade 0,78 para F1 e 0,58 para exatas, indicou desenvolvimento para as questões específicas em administração. O mesmo caso pode ser observado para os alunos 12 e 17, em que seus escores foram ainda menores na habilidade de exatas em comparação com a habilidade específica.

Para verificar a escala padronizada multidimensional, leva-se em conta que o exame multidisciplinar possuía mais questões na área específica de administração do que de exatas. As notas ponderadas apontavam para esta conclusão, visto que as cargas explicativas pelas análises fatoriais foram 07 itens na habilidade de exatas. A prova possuía 10 itens, entre eles existem indicativos de que os itens não foram carregados somente na habilidade de exatas, porque necessitavam de conhecimento específico de administração.

Na escala padronizada multidimensional, dos alunos exemplificados, o conhecimento de cada aluno é ponderando a partir de duas habilidades avaliadas.

Assim, os seguintes alunos apresentam as seguintes notas: aluno 5 com a nota de 0,62; aluno 8 com a nota 0,7; aluno 12 com a nota 0,51 e o aluno 17 com a nota 0,45. A possibilidade que a TRIM proporciona é a de que o aluno 8 tenha seu reconhecimento pela nota proporcional e sua habilidade é desenvolvida conforme seu desempenho de responder, dado o conhecimento apreendido.

A Figura 21 mostra as escalas padronizadas pelos modelos numa visão geral para os 189 e 154 respondentes, conforme a sugestão dos modelos da TRI.

Na Figura 21, para a escala padronizada dos 189 respondentes, caso o exame multidisciplinar de administração utilize a TRI no MD1/2PL, o conhecimento estaria com uma frequência acima de 80 alunos com as notas entre 0,4 a 0,5 conforme os 40 itens respondidos. Retirando-se os alunos que zeraram no exame, a escala padronizada indicou que a frequência é de 50 alunos que obtiveram os valores entre 0,4 a 0,5 e o histograma apresentou simetria. Para os dois histogramas, é relevante observar que, para os 189 respondentes, a escala máxima é até 0,7. Enquanto que para os 154 respondentes, na escala padronizada multidimensional, a máxima é acima de 0,8.

No histograma com a escala padronizada multidimensional para 154 respondentes, utilizou-se a equação 3.21, que representa a habilidade multidimensional, ou seja,

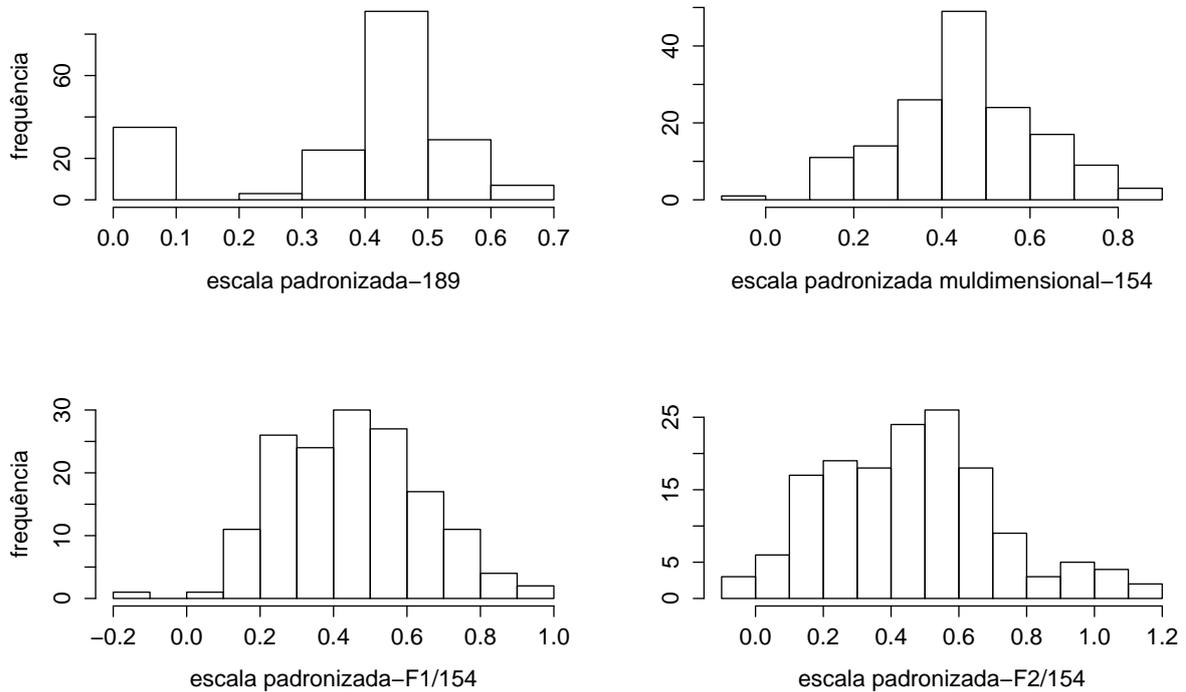


Figura 21: Escalas padronizadas dos 189 e 154 respondentes

ponderou-se as duas habilidades avaliadas e obteve-se uma habilidade que represente os dois conhecimentos. Observou-se um valor atípico, pois um aluno extrapolou a escala entre 0 a 1, como é possível observar no histograma o valor $-0,09$. Ao olhar o Apêndice G, o avaliado teve um escore no F1 abaixo dos alunos que zeraram a prova.

Para detalhar a escala padronizada multidimensional dos dois fatores dos 154 respondentes com a indicação do MLD2/2PLf, dividiu-se dois histogramas: o primeiro representou o F1/154 e o segundo o F2/154.

Na Figura F1/154, encontra-se a escala padronizada dos itens que explicam a habilidade em que estavam relacionadas as áreas específicas de administração e mostra que as notas padronizadas para esse conhecimento concentraram entre 0,4 a 0,6. Também observou-se o valor atípico no F1, que é o mesmo aluno com a escala padronizada multidimensional fora da padronização.

Na F2/154 da Figura 21, as escalas padronizadas desse fator concentraram-se entre 0,4 a 0,6 aproximadamente em 20 a 25 avaliados. Nesse fator, que representa a área de exatas, foi possível observar, pelo escore e pela escala no Apêndice G, 4 alunos com a carga fatorial acima de 1, ou seja, valores acima da estimativa do escore do aluno que acertou todas as questões e o mesmo aconteceu para 3 alunos que tiveram sua escala

padronizada abaixo de zero, levando-se assim, a repensar nos motivos desses valores.

Os escores dos 154 respondentes não devem ser indicados por dois motivos, primeiramente, por mostrar, na Figura 19, que ao retirar os 35 alunos que zeraram, as CCI apresentaram valores decrescentes, em segundo, os itens que não explicaram devem ser retirados. Isso indica um estudo detalhado para esse exame multidisciplinar quando é tratado um modelo global e retira-se as influências tanto dos respondentes como dos itens. Uma sugestão seria abordar o modelo com a priori no parâmetro do acerto ao acaso, detalhes dessa abordagem consta no conjunto de dados III.

4.3 MODELO LOGÍSTICO DA TRI NO CONJUNTO DE DADOS III

O objetivo da presente seção é discutir os resultados do exame multidisciplinar e abordar o conjunto de dados III do Enade/2009 com 436 graduados de estatística em que foram excluídos os exames de 7 alunos que zeraram - com os 32 itens e conforme a determinação da comissão do INEP, 30 itens - em que os itens G1 e Q23 foram excluídos. Em algumas discussões e resultados relevantes, a retirada da questão G2 também é explorada juntamente com os outros dois itens excluídos. Isso ocorre para confirmar o seu comportamento mantendo-a nas análises, visto nos Apêndices A e B, explorado e detalhado na seção 2.2

A finalidade deste estudo é verificar quantas habilidades o aluno utilizou para responder a cada item e indicar um modelo que seja viável para representar o conhecimento do avaliado. Trata-se a prova como geral, ou seja, sem dividir o exame em formação geral ou específica, para verificar a TRI, bem como relatar e captar essas observações.

Para que se desenvolvam estudos e se avance nas perspectivas de avaliação dos resultados no Enade, segue uma visão geral das análises dos modelos exploratórios da TRI. Primeiro, é preciso abordar, identificar, validar e comparar os modelos. Assim, a TRI se ajusta conforme os critérios, respeitando o número de parâmetros do modelo e finaliza com os escores, conforme a metodologia proposta no Capítulo 3. A partir da importância dos assuntos dos resultados da TRI, surge a necessidade de subdividi-los em subseções, para dar continuidade e ordenar os resultados dos estudos.

4.3.1 Dimensionamento do conjunto de dados III

O pacote *psych* (Revelle, 2013), baseado na análise paralela sobre a matriz de correlação tetracórica, auxilia na identificação da dimensionalidade dos dados em estudo e adota a junção de duas análises (componentes principais e análise fatorial). Espera-se encontrar pelo menos dois traços latentes, visto que a prova objetiva possui questões gerais e questões específicas. A seguir, na Figura 22(a), pode-se observar o resultado da técnica aplicada aos dados deste trabalho, considerando todos os 32 itens e também ao excluir as duas questões (G1 e Q23), determinadas pela comissão do INEP.

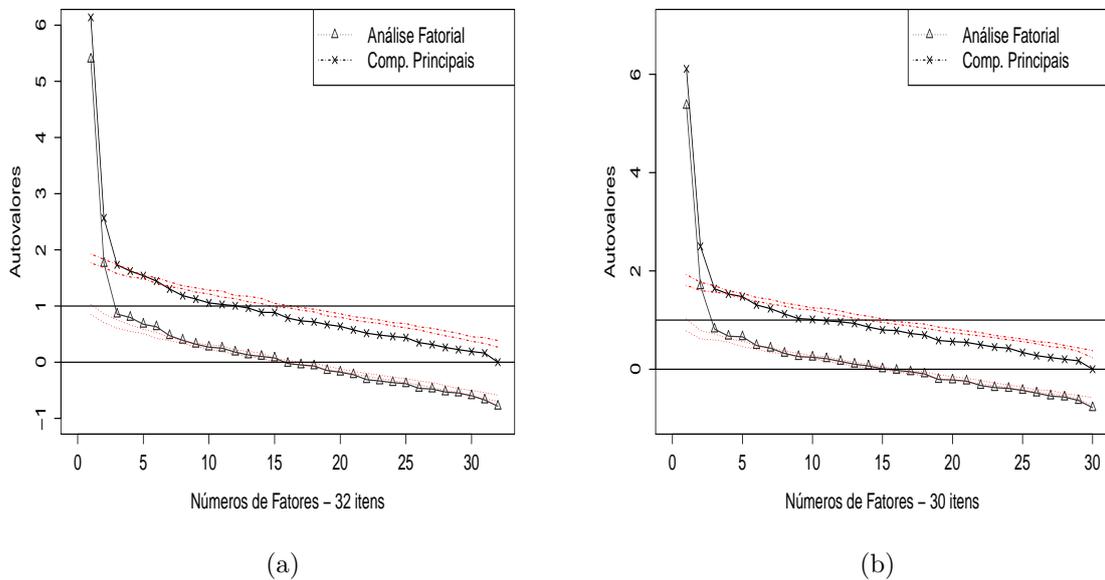


Figura 22: Diagnóstico da dimensão dos 436 respondentes com: 32 questões; 30 questões excluindo a G1 e Q23

O instrumento verifica as dimensões interpretáveis e observa, pelos autovalores, que existe um fator bastante expressivo. Na sequência, outro fator bem menor, mas ainda relevante em relação aos demais, um indicativo de dois traços latentes, configura a bidimensionalidade. Os demais fatores associados aos dados têm contribuições ínfimas. Todavia, se fossem assumidos mais traços latentes na modelagem, isso não traria melhorias significativas ao ajuste do modelo nem ao entendimento do comportamento dos dados.

Visualmente, no gráfico da matriz tetracórica, quando retiraram-se os itens G1, G2 e Q23, não houve diferença do gráfico apresentado na Figura 22(b) com 30 itens, excluindo G1 e Q23, ou seja, o impacto foi pequeno para diagnosticar diferenças na dimensão, logo, optou-se por não apresentá-lo por ser semelhante.

Quanto às competências do Enade, pode-se supor que o fator mais relevante poderia ser a interpretação, bastante útil nas duas habilidades avaliadas e o segundo fator, a competência estatística. Essa constatação pode indicar que os dois testes avaliam basicamente as mesmas coisas. Diante disso, optou-se por verificar o ajuste a modelos uni/multidimensionais de dois fatores, de resposta dicotômica, conforme a subseção 4.3.2.

4.3.2 TRIM no conjunto de dados III

Emprega-se, no modelo logístico da TRIM, o modelo logístico com 1 traço latente (MLD1) e com 2 traços latentes (MLD2), visto que ambos contemplam a TRIU e a TRIM, cujos parâmetros logísticos dos itens são abordados em: estimação da discriminância e o escalar que representa a dificuldade (2PL).

Considera-se o acerto ao acaso fixo como um único valor (2PLf); e os três parâmetros dos itens são estimados (3PL). A diferença é que, conforme aumenta-se a dimensão, o modelo também adapta-se às dimensões propostas, questionando sempre o que vale a pena indicar dada a dimensão. Com esse questionamento, apresenta-se a Tabela 26 e são consideradas as 32 questões para essas análises. Também são excluídas as questões indicadas pela comissão do INEP e as indicadas pelas estatísticas exploratórias - G1, G2 e Q23.

Para os resultados da Tabela 26, utilizou-se o pacote *mirt*, com o procedimento iterativo de estimação dos parâmetros por máxima verossimilhança marginal. Sabe-se que a integração é pelo método de Gauss - Hermite utilizando a estimativa de máxima verossimilhança tradicional (EM). Os modelos exploratórios são estimados por um algoritmo estocástico. Mais detalhes podem ser vistos em Cai (2010).

Na Tabela 26, com os modelos aninhados para a indicação do modelo, os resultados são comparados a partir do valor da verossimilhança maximizada ($\log Lik$) e pelos critérios de AIC/BIC. Observa-se que todos os modelos se ajustaram e o MLD2 com 3PL foi indicado com os 32 itens pelo $\log Lik$, pelo critério de AIC; enquanto o menor é o MLD2/2PLf e o BIC o MLD1/2PLf. É importante definir que o valor considerado fixo do acerto ao acaso do 2PLf foi determinado estimando-se e perfilhando-se os modelos pela máxima verossimilhança, conforme seguem os valores do c_i que, neste caso, foram quase equivalentes entre os modelos dimensionais.

As possibilidades de retirar os itens G1 e Q23 e depois G1, G2 e Q23 mostraram que os resultados das estatísticas de diagnóstico de qualidade foram os mesmos para

Tabela 26: Resultados das estatísticas de diagnósticos da qualidade do ajuste dos modelos

MLD1	436 respondentes			
	df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0$)	64	-7727,53	15583,07	15844,04
2PLf ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0,097$)	65	-7699,55	15527,11	15789,1
3PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c_i$)	96	-7674,83	15541,65	15934,64
MLD2				
2PL ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0$)	96	-7665,22	15520,44	15909,33
2PLf ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0,097$)	97	-7649,13	15488,27	15877,16
3PL ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c_i$)	128	-7624,53	15503,07	16022,95
436 respondentes sem as questões G1 e Q23				
MLD1	df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0$)	60	-7335,14	14790,28	15034,94
2PLf ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0,103$)	61	-7306,36	14732,74	14977,39
3PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c_i$)	90	-7287,16	14754,33	15121,32
MLD2				
2PL ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0$)	90	-7275,09	14728,18	15091,09
2PLf ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0,103$)	91	-7259,39	14696,79	15059,7
3PL ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c_i$) ($c_i = c_i$)	120	-7239,05	14716,10	15201,34
436 respondentes sem as questões G1, G2 e Q23				
MLD1	df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0$)	58	-7041,62	14199,26	14435,76
2PLf ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0,103$)	59	-7012,66	14141,33	14377,83
3PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c_i$)	87	-6993,77	14161,55	15416,31
MLD2				
2PL ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0$)	87	-6982,58	14137,17	14487,85
2PLf ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0,103$)	88	-6968,77	14109,54	14460,22
3PL ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c_i$) ($c_i = c_i$)	116	-6945,37	14120,75	14589,68

Nota:

- questões gerais (G);
- questões específicas (Q);
- diferença de parâmetros (df);
- modelo logístico de um traço latente (MLD1);
- modelo logístico de dois traços latentes (MLD2);
- dois parâmetros logísticos (2PL) ;
- dois parâmetros logísticos estimado e o acerto ao acaso fixado (2PLf); e
- três parâmetros logísticos (3PL).

ambos, ou seja, o modelo pela $\log Lik$ é o MLD2/3PL, para o AIC, é o MLD2/2PLf; para o BIC, é o MLD1/2PLf.

Como os resultados não foram unânimes, compara-se a quantidade estimada de parâmetros (df) com os padrões de resposta. Neste caso, para todos os modelos propostos, existem mais padrões de resposta do que parâmetros estimados. Para que se esgotem e se finalizem as análises desta tabela, testa-se a razão da verosimilhança entre os modelos aninhados, para verificar se compensa ir de um modelo mais simples para um mais complexo. Testaram-se o MLD1/2PLf com MLD2/2PLf e o MLD1/3PL com MLD2/3PL a 95% de confiança e $df = 32$ para os 436 respondentes. E também quando retiram-se as questões, o teste compensa o modelo com mais parâmetros estimados e com duas dimensões.

Assim, antes de tomar a decisão e indicar o modelo da TRI para o Enade, segue a subsubseção 4.3.2.1 e averiguam-se as estimativas dos parâmetros nos modelos analisados.

4.3.2.1 Estimação dos parâmetros nos modelos da TRIM no conjunto de dados III

A análise torna-se necessária para que se entenda tanto o comportamento dos modelos como as diferenças entre os parâmetros estimados e também as vantagens da escolha de um modelo para o outro. Na Tabela 27, seguem alguns itens dos modelos logísticos de uma, duas e três dimensões e juntamente os parâmetros estimados seguindo a hierarquia de cada modelo em 2PL, 2PLf e 3PL.

Tabela 27: Estimativas dos parâmetros dos modelos do conjunto de dados III

Q	436 respondentes																				
	MLD1									MLD2											
	2PL			2PLf			3PL			2PL				2PLf				3PL			
	α_{i1}	d_i	c_i	α_{i1}	d_i	c_i	α_{i1}	d_i	c_i	α_{i1}	α_{i2}	d_i	c_i	α_{i1}	α_{i2}	d_i	c_i	α_{i1}	α_{i2}	d_i	c_i
G1	0,11	-2,17	0	0,4	-4,85	0,1	0,15	-2,25	0,01	-0,09	-0,28	-2,19	0	-0,28	-2,07	-6,3	0,1	-0,12	-0,33	-2,24	0
G2	0,18	0,34	0	0,21	0,16	0,1	0,2	0,19	0,08	-0,01	-0,27	0,34	0	-0,12	-0,43	0,16	0,1	-0,1	-0,38	0,26	0,04
G3	0,59	0,2	0	0,66	-0,02	0,1	0,61	0,18	0	0,1	-0,73	0,2	0	0,18	-0,74	-0,02	0,1	0,01	-0,75	0,2	0
G4	0,41	-0,65	0	0,49	-1	0,1	1,06	-2,26	0,25	0,48	-0,02	-0,65	0	0,61	0	-1,02	0,1	0,46	-0,02	-0,66	0
G5	0,59	-0,44	0	0,7	-0,77	0,1	0,92	-1,19	0,18	0,31	-0,46	-0,44	0	0,35	-0,63	-0,79	0,1	0,41	-1,09	-1,66	0,23
G6	0,76	0,16	0	0,86	-0,07	0,1	0,83	0,06	0,04	0,24	-0,84	0,17	0	0,28	-1,1	-0,09	0,1	0,12	-1,36	-0,41	0,19
G7	1,2	2,14	0	1,38	2,13	0,1	1,31	2,23	0	-0,32	-2,45	3,06	0	-0,36	-3,94	4,08	0,1	-0,91	-3,19	3,49	0,02
G8	1,42	2,59	0	1,5	2,54	0,1	1,57	2,75	0	0,07	-1,93	3,06	0	0,18	-1,94	2,93	0,1	0,07	-1,87	3,01	0
Q11	0,65	0,09	0	0,73	-0,15	0,1	0,69	0,04	0,02	0,41	-0,41	0,08	0	0,54	-0,34	-0,15	0,1	0,61	-0,25	0,08	0
Q12	0,56	0,28	0	0,65	0,08	0,1	0,65	0,27	0	-0,1	-0,98	0,3	0	0,05	-0,94	0,08	0,1	-0,19	-0,96	0,28	0
Q13	0,97	-1,84	0	2,37	-3,92	0,1	2,47	-3,84	0,08	1,61	0,4	-2,19	0	4,31	1,24	-5,97	0,1	2,15	0,51	-2,69	0,02
Q15	0,59	-0,5	0	0,79	-0,88	0,1	2,24	-3,06	0,28	0,66	-0,09	-0,51	0	0,76	-0,18	-0,88	0,1	1,76	-1,99	-4,1	0,29
Q16	0,87	-0,62	0	1,17	-1,09	0,1	1,58	-1,7	0,17	0,67	-0,42	-0,62	0	0,84	-0,58	-1,07	0,1	0,96	-0,73	-1,55	0,16
Q17	0,69	-1,1	0	1,05	-1,79	0,1	2,82	-4,47	0,2	0,74	-0,11	-1,12	0	0,96	-0,31	-1,79	0,1	2,57	-2,03	-5,98	0,21
Q18	0,62	-1,83	0	4,71	-8,29	0,1	6,86	-11,4	0,11	0,97	0,3	-1,97	0	8,91	0,46	-14,9	0,1	14,7	2,38	-22,5	0,11
Q19	0,85	-1,13	0	1,68	-2,16	0,1	2,55	-3,33	0,15	0,84	-0,19	-1,14	0	1,48	-0,32	-2,07	0,1	2,21	-0,55	-3,3	0,15
Q20	0,5	-1,72	0	1,97	-4,09	0,1	2,39	-4,77	0,12	0,79	0,21	-1,82	0	1,57	-0,24	-3,61	0,1	1,34	-0,64	-3,78	0,11
Q21	0,97	-1,04	0	1,47	-1,76	0,1	1,5	-1,76	0,09	0,78	-0,41	-1,04	0	1,35	-0,39	-1,78	0,1	1,58	-0,38	-2,1	0,12
Q23	0,24	-1,02	0	0,41	-1,52	0,1	4	-7,52	0,24	0,38	0,12	-1,04	0	0,67	0,2	-1,6	0,1	4,72	-0,9	-9,73	0,24
Q24	0,48	-0,86	0	0,66	-1,31	0,1	0,68	-1,27	0,09	0,28	-0,32	-0,86	0	0,44	-0,34	-1,3	0,1	1,98	-0,67	-4,34	0,25
Q25	1	-1,76	0	2,02	-3,42	0,1	1,66	-2,74	0,07	0,82	-0,42	-1,76	0	1,65	-1,46	-3,98	0,1	0,87	-0,41	-1,82	0
Q26	1,08	-1,46	0	1,67	-2,45	0,1	1,37	-1,9	0,05	0,8	-0,58	-1,46	0	1,25	-0,88	-2,48	0,1	0,88	-0,88	-1,99	0,05
Q27	1,25	-0,07	0	1,46	-0,42	0,1	1,33	-0,19	0,02	0,74	-0,86	-0,08	0	1,02	-0,89	-0,41	0,1	0,86	-0,74	-0,12	0
Q28	0,7	-0,84	0	0,99	-1,37	0,1	1,86	-2,66	0,2	0,69	-0,15	-0,85	0	1	-0,07	-1,37	0,1	4,63	0,45	-6,29	0,24
Q29	0,36	-0,86	0	0,59	-1,33	0,1	1,77	-3,41	0,24	0,46	0,03	-0,87	0	0,74	0,1	-1,36	0,1	2,19	0,49	-3,52	0,23
Q30	0,46	-0,01	0	0,5	-0,25	0,1	0,65	-0,71	0,23	0,43	-0,13	-0,01	0	0,53	-0,06	-0,25	0,1	0,44	-0,11	-0,02	0
Q31	1,19	-0,42	0	1,35	-0,8	0,1	1,25	-0,47	0	0,69	-0,88	-0,43	0	0,99	-0,7	-0,79	0,1	0,76	-0,76	-0,46	0
Q32	0,43	0,07	0	0,48	-0,16	0,1	0,43	0,05	0,01	0,22	-0,34	0,07	0	0,28	-0,34	-0,16	0,1	4,9	-9,2	-17,5	0,48
Q33	1,18	-0,92	0	1,83	-1,72	0,1	2,43	-2,46	0,15	0,98	-0,5	-0,92	0	1,44	-0,76	-1,69	0,1	2,01	-1,26	-2,96	0,17
Q34	0,85	-1,65	0	2,41	-3,93	0,1	3,46	-5,39	0,12	1,28	0,26	-1,83	0	2,56	0,34	-3,8	0,1	1,95	0,3	-2,75	0,06
Q36	0,74	-1,48	0	1,75	-2,93	0,1	2,22	-3,61	0,12	0,97	0,07	-1,56	0	1,66	-0,09	-2,79	0,1	1,96	-0,01	-3,19	0,11
Q37	0,98	-0,25	0	1,34	-0,65	0,1	2,36	-1,89	0,25	0,83	-0,4	-0,25	0	1,33	-0,31	-0,65	0,1	1,9	-1,15	-2,05	0,25

Na Tabela 27 e com mais detalhes no Apêndice H, observam-se as estimativas dos parâmetros, primeiramente quando os itens G1 e Q23 foram retirados da análise e depois com a subtrações dos itens G1, G2 e Q23. Para cada modelo dimensional, com os 32 itens, excluem-se G1 e Q23. Diferente de quando se comparam as estimativas dos parâmetros entre os 32 itens e excluem-se G1, G2 e Q23. A mesma observação é feita entre G1; G2; Q23 e G1; Q23, as diferenças são numéricas e em sinais, ou seja, há a indicação que a G2 influencia nas estimativas dos parâmetros.

Na Tabela 27 a discriminação aumenta ao considerar o acerto ao acaso e o parâmetro d_i , o vetor que relaciona a dificuldade reduz, quanto maior for o “chute”. No modelo ML1, o item com o maior discriminância é o G8 para o 2PL, ao se considerar o acerto ao acaso em 2PLf e 3PL, pois a discriminação e a dificuldade são elevadas para Q18. Quando se observa o item Q18 para o MLD2, sobrecarrega-se na primeira discriminação (α_{i1}) e em consequência o d_i , quando o valor do parâmetro aumenta.

Outra observação importante feita é que, ao olhar d_i , medido na mesma escala da habilidade, que é $d_i = -\alpha_i \cdot \beta_i$, que representa a dificuldade do item no MLD1/3PL, ele não é visto diretamente na tabela, uma vez que é necessário determinar β_i para saber que o item considerado mais difícil é o G1. Isso é resolvido quando se encontra o valor da dificuldade para o G1 no MDL1/3PL, em que $\beta_i = -2,25 / -0,15$ com $\beta_i = 15$. Na Tabela 7 do Relatório Síntese do Enade (2009), o valor encontrado para a G1 mostra, nas análises exploratórias em relação aos outros itens, que G1 é muito difícil e tem fraca discriminação.

A Tabela 27 mostra que a estimação do parâmetro acerto ao acaso para o MLD1 está em torno de $0 \leq c_i \leq 0.28$, cujo item Q15 é o de maior “chute” e quando passa para MLD2, a distância aumenta de $0 \leq c_i \leq 0.48$ e Q32 é o que apresenta o maior acerto ao acaso.

Nos resultados da estimação dos parâmetros apresentados na Tabela 27, como no Apêndice H, mostra-se o cuidado antes de se indicar o modelo para representar o exame do Enade. Como as diferenças de resultados entre os MLD1 e MLD2 existem, é necessário observar o carregamento dos fatores para os itens com a análise fatorial para determinar a dimensão e decidir qual modelo compensa levar em conta e que, ao se excluir a G2 na estimação, os valores alteram consideravelmente.

Para dar continuidade e entender a relação dos valores considerados para os parâmetros dos itens e o comportamento dos parâmetros estimados, segue a subsubseção 4.3.2.2, a Curva Característica do Item (CCI), resumidos graficamente.

4.3.2.2 CCI e cargas fatoriais dos itens na TRIM no conjunto de dados III

A CCI é observada e verifica-se que, quanto maior for a habilidade do respondente, maior é a probabilidade de responder corretamente ao item. O modelo proposto baseia-se no fato de que o indivíduo com maior habilidade, possua a maior probabilidade de acertar o item e que esta relação não é linear.

Como o modelo não está definido, observa-se a CCI do MLD1 com os três parâmetros estimados - 3PL para os 32 itens; excluindo G1 e Q23 - 30 itens e retirando G1, G2 e Q23 - 29 itens.

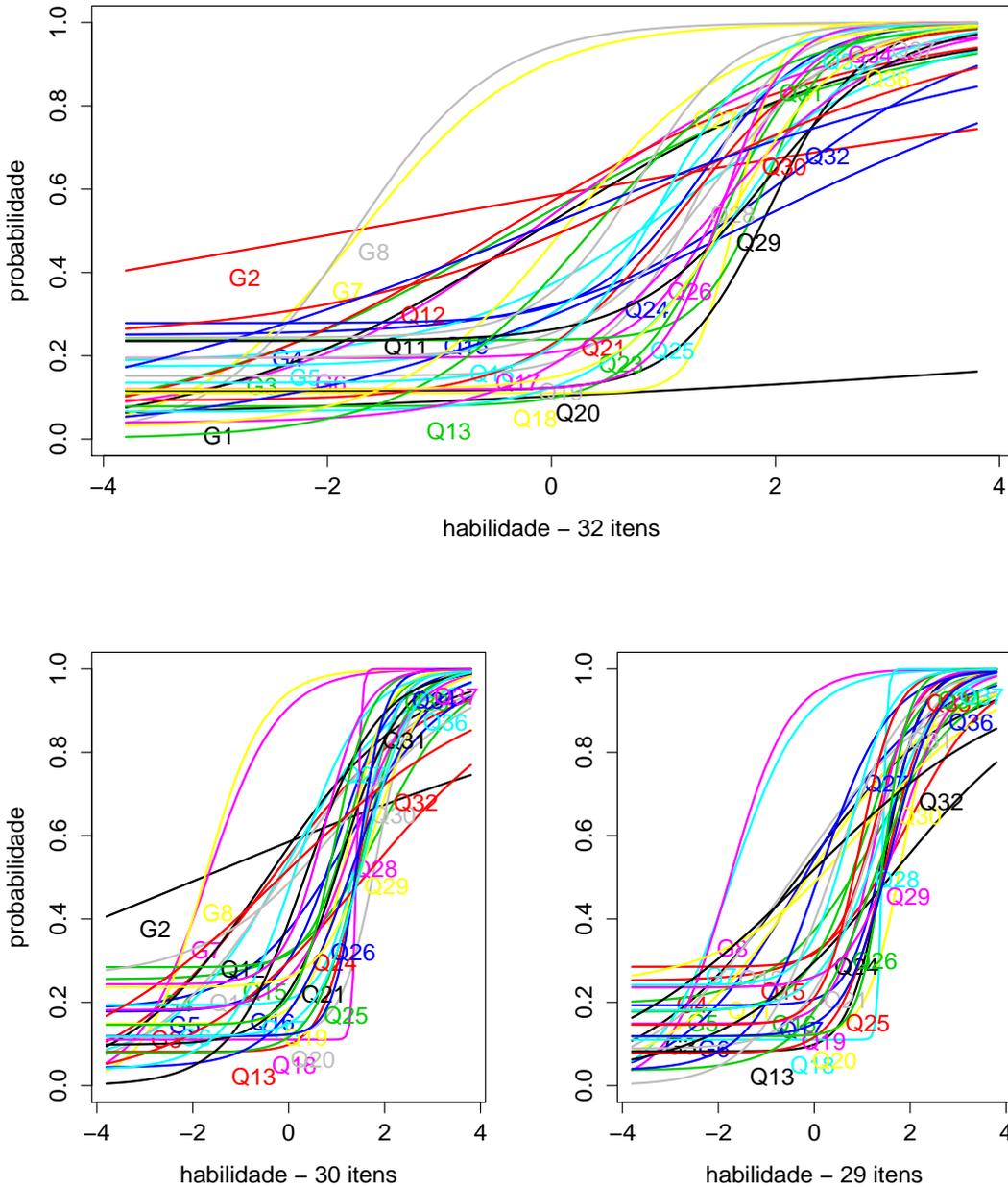


Figura 23: CCI do MLD1-3PL para 436 respondentes com 32, 30 e 29 itens

Na Figura 23 é possível identificar que nas três CCI existem duas curvas que visualmente permanecem na mesma posição com a inclinação diferente dos itens agrupados, os quais representam a discriminação na habilidade próxima de 1.3 e a dificuldade abaixo de zero, ou seja, uma questão fácil. Eles, são os itens G8 e G7 que podem ser vistos na

Tabela 27, enquanto o Apêndice H, para a quantidade de itens 32, 30 e 29 a G8 e G7, apresenta os valores dos três parâmetros equivalentes.

Outros três itens que necessitam observação nos MLD1/3PL são G1, G2 e Q23, mas o G2 é o item que menos discrimina e possui o maior valor no acerto ao acaso. Quando retira-se G1 e Q23 ficando com 30 itens, conforme o relatório do Enade (2009), o item G2 continua com probabilidade do acerto ao acaso próximo de 0.4, com baixa discriminação com 0.2 e dificuldade equivalente para ambas as figuras, detalhadas na Tabela 27.

A G1 é a questão muito difícil, com a probabilidade de acerto ao acaso abaixo da G2. Está difícil localizar a Q23 na tabela, porque ela se encontra no aglomerado dos itens e acompanha o formato da CCI em “S”.

A Figura 24, para não dificultar as visualizações, mostra apenas os itens em discussão, G1, G2 e Q23 no MLD1/3PL, a partir das Curvas Características dos Itens (CCI) que apresentam todas as informações relevantes, provenientes dos 32 itens.

Na Figura 24 é possível identificar a questão Q23, que apresenta o íngreme em “S”, segundo Andrade et al.(2000) com discriminação no valor de 4, dificuldade 1,88 e acerto ao acaso de 0,24.

A G2 teve elevada probabilidade de chute e pouca discriminância entre as questões. Em comparação com G1, notou-se também que a curva linearizou-se e está mais próxima do que a G2 do eixo das habilidades. Isso mostra que é o item mais difícil, sem discriminação e com a probabilidade de acerto ao acaso pequena. Mais detalhes podem ser encontrados nos relatórios do INEP (2009). Recomenda-se olhar as questões para entender os motivos que levaram a esse comportamento da área geral analisada, disponíveis no portal do INEP.

Continuando nos modelos exploratórios, na Tabela 28, apresenta os resultados dos carregamentos dos fatores de explicação, e conforme Nojosa (2002), Reckase (2009) e Chalmers (2012), a técnica das cargas fatoriais possibilita informar e também averiguar quais itens se relacionam a cada fator. Quanto maior a carga fatorial, melhor é a relação entre item e fator. Até porque, mesmo com os testes já realizados e sendo o modelo indicado por ser um dos mais complexos, vale ressaltar que computacionalmente e metodologicamente a relação entre custo benefício tem que valer a pena.

Segue a Tabela 28 com alguns resultados relevantes e, no Apêndice I, a tabela completa. Os resultados indicados são propostos, conforme diagnósticos estatísticos das análises dos modelos, após a rotação oblíqua das cargas fatoriais:

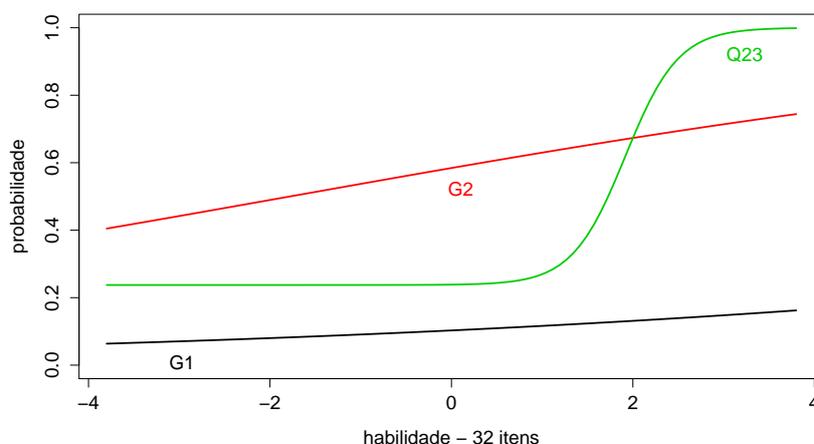


Figura 24: Curva característica dos itens G1, G2 e Q23 no MLD1-3PL

Na Tabela 28 não consta as análises da retirada da G1, G2 e Q23, mas para indicar os impactos das questões nos modelos, é necessário, nas discussões, as observações do Apêndice I.

Na Tabela 28, para os 436 respondentes com as 32 questões, conforme os modelos indicados, é possível analisar as cargas fatoriais dos itens nos modelos de uma e duas dimensões. O item G1 somente explica sua inclusão quando considera o acerto ao acaso fixo, no MLD2/2PLf, enquanto a Q23 explica MLD1 e MLD2 no 3PL. Observa-se, também, que incluindo ou não a G1, G2 e a Q23, poucas questões carregaram, explicando o F2. Para essa prova, a grande maioria das questões foi explicada no F1.

Observa-se, no modelo dimensional MLD1, que 22 itens explicam melhor no 3PL, ao serem considerados todos os parâmetros estimados, do que fixando o acerto ao acaso no 2PLf. Outro fator importante é que, devido ao resumo, não foi incluído o MLD1 das 30 questões na tabela, porque, como é visto até no MLD2, há poucas diferenças numéricas entre os carregamentos, como mostra o Apêndice I.

No Apêndice I, ao comparar o MLD1/3PL com as 32 questões e MLD1/3PL retirando G1, G2 e Q23, as cargas fatoriais explicadas nos itens não alteram consideravelmente, como é o exemplo da Q36 com comunalidade de 0,61 ao excluir os itens e mantendo 0,63. Outro exemplo é o item G3, cuja comunalidade de 0,11 é a mesma para ambas as análises. As diferenças seriam consideráveis se fosse escolhido um modelo com mais de um traço latente.

Outro fator importante, e com relação às cargas fatoriais no MLD1 e à comunalidade, é que o item tem que explicar 30% da prova com a comunalidade acima de 0,15,

Tabela 28: Cargas fatoriais dos modelos para as 32 questões e retirando a G1 e Q23

Q	436 respondentes e 32 questões										436 respondentes e 30 questões					
	MLD1				MLD2						MLD2					
	2PLf		3PL		2PLf			3PL			2PLf			3PL		
	F1	h2	F1	h2	F1	F2	h2	F1	F2	h2	F1	F2	h2	F1	F2	h2
G1	0,23	0,05	0,09	0,01	0,37	-0,67	0,59	0,07	-0,16	0,03	-	-	-	-	-	-
G2	0,12	0,01	0,12	0,01	0,08	-0,22	0,06	0,1	-0,18	0,04	0,08	-0,26	0,07	0,16	-0,26	0,09
G3	0,36	0,13	0,34	0,11	0,32	-0,28	0,18	0,29	-0,29	0,17	0,32	-0,28	0,18	0,29	-0,27	0,16
G4	0,28	0,08	0,53	0,28	0,31	0,12	0,11	0,25	0,08	0,07	0,32	0,12	0,12	0,25	0,09	0,07
G5	0,38	0,15	0,48	0,23	0,37	-0,2	0,18	0,54	-0,29	0,37	0,36	-0,24	0,19	0,52	-0,27	0,35
G6	0,45	0,2	0,44	0,19	0,44	-0,37	0,33	0,48	-0,42	0,41	0,43	-0,38	0,34	0,49	-0,41	0,4
G7	0,63	0,4	0,61	0,37	0,47	-0,78	0,84	0,41	-0,76	0,75	0,48	-0,81	0,88	0,46	-0,76	0,8
G8	0,66	0,44	0,68	0,46	0,5	-0,57	0,58	0,54	-0,51	0,55	0,51	-0,56	0,58	0,56	-0,49	0,56
Q11	0,39	0,15	0,38	0,14	0,38	-0,04	0,15	0,4	0,02	0,16	0,39	-0,02	0,15	0,4	0,05	0,16
Q12	0,35	0,13	0,36	0,13	0,31	-0,38	0,24	0,26	-0,39	0,22	0,32	-0,35	0,22	0,27	-0,38	0,22
Q13	0,81	0,66	0,82	0,68	0,73	0,57	0,86	0,63	0,43	0,58	0,76	0,51	0,84	0,61	0,45	0,57
Q15	0,42	0,18	0,8	0,63	0,43	0,07	0,19	0,85	-0,22	0,77	0,44	0,05	0,19	0,85	-0,16	0,75
Q16	0,57	0,32	0,68	0,46	0,55	-0,08	0,31	0,63	-0,08	0,41	0,55	-0,08	0,31	0,64	-0,05	0,41
Q17	0,52	0,27	0,86	0,73	0,53	0,05	0,29	0,91	-0,13	0,84	0,54	0,06	0,3	0,86	-0,05	0,75
Q18	0,94	0,88	0,97	0,94	0,89	0,41	0,96	0,86	0,49	0,99	0,9	0,39	0,96	0,85	0,5	0,98
Q19	0,7	0,49	0,83	0,69	0,67	0,13	0,47	0,82	0,13	0,68	0,69	0,16	0,51	0,81	0,17	0,68
Q20	0,76	0,57	0,81	0,66	0,68	0,16	0,48	0,71	0,01	0,5	0,77	-0,01	0,6	0,77	-0,06	0,6
Q21	0,65	0,43	0,66	0,44	0,65	0,08	0,44	0,71	0,12	0,52	0,66	0,13	0,46	0,68	0,16	0,49
Q23	0,23	0,05	0,92	0,85	0,28	0,22	0,13	0,93	0,19	0,9	-	-	-	-	-	-
Q24	0,36	0,13	0,37	0,14	0,34	-0,06	0,12	0,8	0,08	0,65	0,35	-0,02	0,12	0,72	0,12	0,53
Q25	0,76	0,58	0,7	0,49	0,8	-0,19	0,68	0,54	0,01	0,29	0,82	-0,21	0,72	0,71	-0,07	0,51
Q26	0,7	0,49	0,63	0,39	0,7	-0,1	0,5	0,64	-0,14	0,43	0,72	-0,12	0,53	0,66	-0,12	0,45
Q27	0,65	0,42	0,62	0,38	0,65	-0,15	0,44	0,61	-0,1	0,39	0,66	-0,14	0,45	0,62	-0,09	0,39
Q28	0,5	0,25	0,74	0,54	0,49	0,16	0,26	0,84	0,41	0,87	0,5	0,16	0,28	0,82	0,44	0,86
Q29	0,33	0,11	0,72	0,52	0,34	0,19	0,15	0,64	0,42	0,59	0,36	0,22	0,18	0,61	0,43	0,55
Q30	0,28	0,08	0,36	0,13	0,29	0,08	0,09	0,28	0,04	0,08	0,3	0,07	0,1	0,27	0,03	0,07
Q31	0,62	0,39	0,59	0,35	0,62	-0,09	0,39	0,59	-0,13	0,36	0,62	-0,09	0,39	0,58	-0,11	0,35
Q32	0,27	0,07	0,24	0,06	0,26	-0,09	0,08	0,91	-0,4	0,98	0,27	-0,09	0,08	0,92	-0,35	0,98
Q33	0,73	0,54	0,82	0,67	0,72	-0,03	0,53	0,85	-0,06	0,73	0,74	-0,03	0,55	0,85	-0,04	0,73
Q34	0,82	0,67	0,9	0,81	0,72	0,41	0,68	0,64	0,36	0,54	0,75	0,37	0,71	0,62	0,36	0,51
Q36	0,72	0,51	0,79	0,63	0,67	0,23	0,49	0,71	0,26	0,57	0,69	0,19	0,51	0,7	0,27	0,56
Q37	0,62	0,38	0,81	0,66	0,64	0,11	0,42	0,83	-0,05	0,7	0,64	0,09	0,42	0,84	-0,04	0,71

Nota:

- questões gerais (G);
- questões específicas (Q);
- modelo logístico de um traço latente (MLD1);
- modelo logístico de dois traços latentes (MLD2);
- dois parâmetros logísticos estimados e o acerto ao acaso fixado (2PLf);
- três parâmetros logísticos (3PL);
- carga fatorial com uma (F1) e duas (F2) habilidades e
- comunalidade (h2)

assim, os itens G1, G2 e Q32 não atendem esse pré-requisito. O procedimento é realizado com a retirada dos itens G1 e G2 conforme as outras análises, mas, a Q32 fica como indicação para análise na CCI. Como o CD3 não adota a TRI, a intenção é do modelo relatar o exame na perspectiva de como foi aplicado.

Na Tabela 28, existe a indicação de se excluir a G2, já que não carrega em nenhum modelo e foi indicada também na análise descritiva pelo ponto bisserial na prova. Como o Enade envolve análise de itens, seria necessário um especialista em questões para verificar os itens G1, G2 e Q23 antes de excluí-los. Assim, seria conveniente um modelo

com dois traços latentes, visto que é um exame multidisciplinar, mas com benefício, e como são poucas questões que carregaram na segunda dimensão, é vantajoso indicar o modelo MLD1/3PL. Portanto, para continuar as discussões, são abordados os escores na subsubseção 4.3.2.3, em que se consideram os 32 itens e excluem-se G1, G2 e Q23.

4.3.2.3 Escores das habilidades da TRIM no conjunto de dados III

Como no Enade ainda não se tem a utilização da TRI e as notas pela TCT não representam as pontuações do Enade, por conter somente as questões objetivas, prosseguem-se as análises a fim de se verificar o escore dos respondentes, conforme a indicação unidimensional do MDL1/3PL e transformando o escore resultante para escala padronizada entre 0 e 1. Segue a Tabela 29 para discussão dos escores do exame do Enade.

Na Tabela 29, são inevitáveis as comparações e, não fica tão fácil observar os resultados do escore de cada respondente obtido pelo MLD1/3PL mas, ao transformar esses resultados em uma escala padronizada, pela equação 3.20, fica evidente que a TRI diferencia o escore da habilidade do conhecimento da TCT.

Toma-se como exemplo o aluno 19 que a nota pela TCT é acima da média - 0,75 - considerada na grande maioria das IES. Caso o escore fosse pela TRI, considerando todos os 32 itens, sua nota seria igual, mas ao serem retirados os itens G1, G2 e Q23, sua habilidade abaixa para 0,65.

Outro exemplo são os alunos 1 e 15, ambos possuem a mesma nota (0,69) pela TCT e acertaram 22 itens. Na TRI no modelo MLD1/32 itens, o aluno 1 teve uma nota de 0,71 e se aproximou mais da TCT do que o aluno 15 com a nota de 0,73, cuja diferença também é proporcional quando retiram-se os itens indicados.

Pela TRI e TCT, esses alunos se diferenciam nas notas pela composição de itens que acertaram. Segue a Figura 25 para compreender os resultados no banco de questão dos itens corretos, dos alunos em estudo, pela CCI.

Como exemplos os alunos 5 e 12 que acertaram somente 10 itens. A nota pela TCT 0,31 e para os 32 itens seria 0,32; 0,46 e retirando-se os itens 0,27; 0,39 respectivamente, ou seja, no último caso, o aluno 12 teve uma menor habilidade. Para essa compreensão, segue a Figura 25 dos itens que os dois alunos acertaram.

O aluno 5, para os 32 itens, acertou mais questões com acerto ao acaso entre 0,2 a 0,4, além da pouca discriminação com itens fáceis e difíceis. Já o aluno 12 mostra uma habilidade um pouco melhor, por acertar o item G1, considerado o item mais difícil. As

Tabela 29: Escores e escala padronizada do Enade no MLD1-3PL para 436 graduados de estatística - 2009

Al	T. A	TCT	MLD1 - 32 itens		MLD1 - sem G1,G2 e Q23	
			3PL	Escala P.	3PL	Escala P.
1	22	0,69	1,45	0,71	1,38	0,61
2	16	0,5	0,4	0,5	0,3	0,42
3	12	0,38	0,24	0,47	0,21	0,41
4	18	0,56	1	0,62	0,96	0,54
5	10	0,31	-0,57	0,32	-0,58	0,27
6	19	0,59	1,77	0,77	1,67	0,66
7	13	0,41	0,31	0,49	0,35	0,43
8	21	0,66	1,52	0,72	1,6	0,65
9	12	0,38	0,34	0,49	0,34	0,43
10	24	0,75	2,03	0,82	1,96	0,71
11	18	0,56	1,53	0,72	1,37	0,61
12	10	0,31	0,18	0,46	0,1	0,39
13	18	0,56	1,14	0,65	1,17	0,57
14	13	0,41	0,4	0,5	0,39	0,44
15	22	0,69	1,58	0,73	1,64	0,65
16	14	0,44	0,66	0,56	0,69	0,49
17	17	0,53	0,74	0,57	0,67	0,49
18	15	0,47	0,26	0,48	0,26	0,42
19	24	0,75	1,58	0,73	1,64	0,65
20	17	0,53	1,11	0,64	1,16	0,57

Nota: Alunos (Al); total de acertos (T.A);teoria clássica dos testes (TCT) modelo linear unidimensional (MLD1) e 3 parâmetros estimados (3PL)

questões corretas discriminam melhor pelo íngreme da CCI e por acertarem menos itens com a probabilidade de acerto ao acaso, próximo de 0,2 a 0,4.

Na Figura 25, ao serem retirados os itens G1, G2 e Q23, o aluno 5 foi penalizado pela G2, ficando com 9 itens para serem avaliados no MDL1-3PL. O aluno 12 foi punido pela G1 e G2 e seu escore é baseado em 8 questões corretas. Para o aluno 12, a CCI continua com o íngreme no formato de “S” melhor do que os itens do aluno 5. Por esse fator, o escore do aluno 12 sobressaiu-se em decorrência dos itens seguirem um equilíbrio de acertos.

Na verdade, o que chama atenção é a conformidade dos fatos em acertar o item que seja realmente relevante para TRI, em que o desempenho de sua habilidade seja favorável em acertar itens fáceis, médios e difíceis. Não vale a pena acertar itens difíceis sem saber responder aos fáceis e médios. Assim, justifica-se que o indivíduo acertou ao acaso o item difícil já que não soube responder ao item fácil ou o médio.

Para uma visualização detalhada de todo escore do Enade pela TRI do MLD1/3PL com os 32 itens, segue a Figura 26, que representa em (a) a distribuição dos responden-

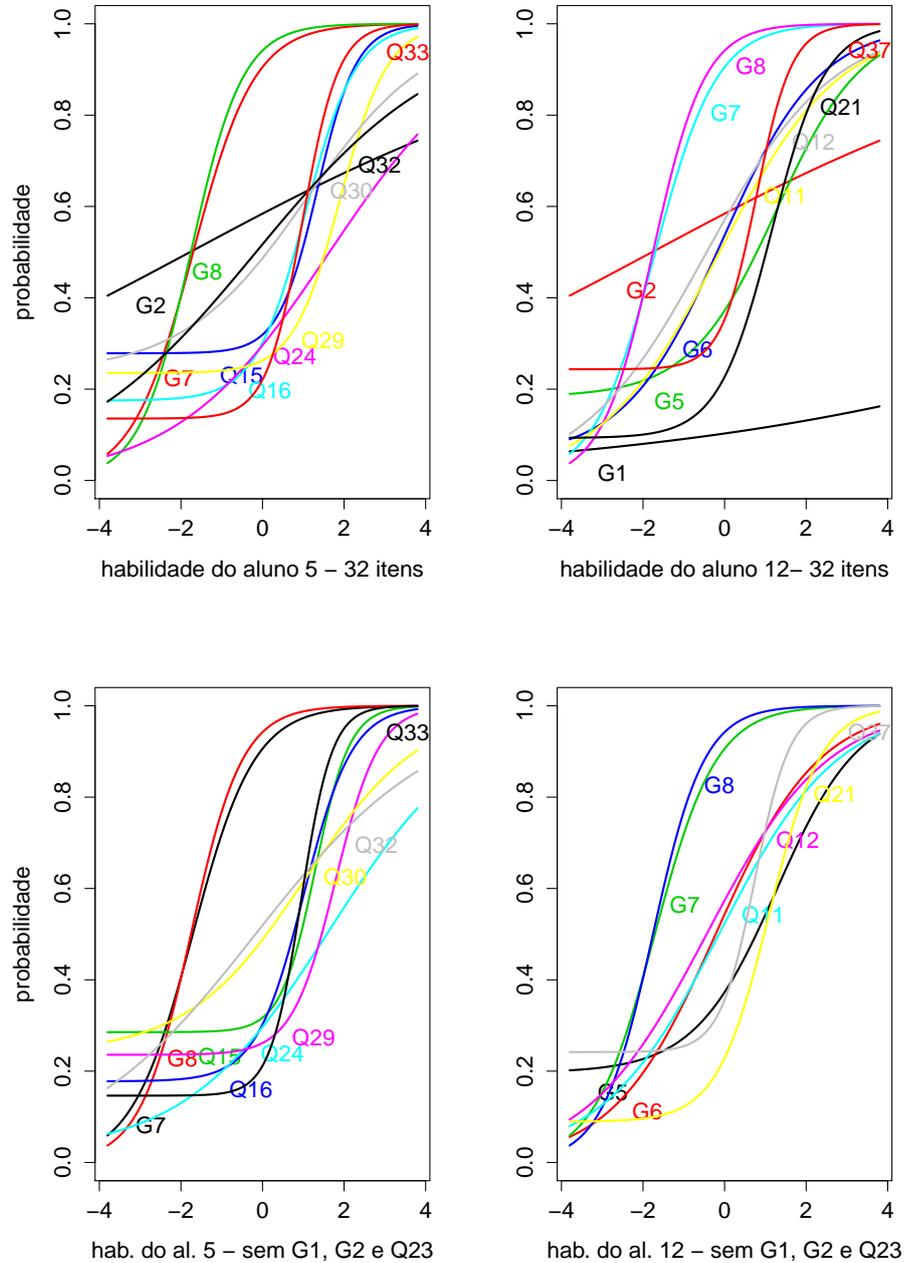


Figura 25: CCI do MLD1-3PL dos itens corretos para os alunos 5 e 12 com 32 itens e excluindo G1, G2 e Q23

tes e as suas respectivas notas pela TRI e na Figura (b) o histograma que caracteriza a frequência das notas pelos respondentes.

Ambas as Figuras (a) e (b) da Figura 26 estão relacionadas e proporcionam uma visualização geral de como os alunos sucederam na avaliação objetiva do Enade/2009 com 32 questões, ou seja, poucos alunos tiveram nota acima 0,8 e a grande maioria ficou entre 0,4 a 0,6.

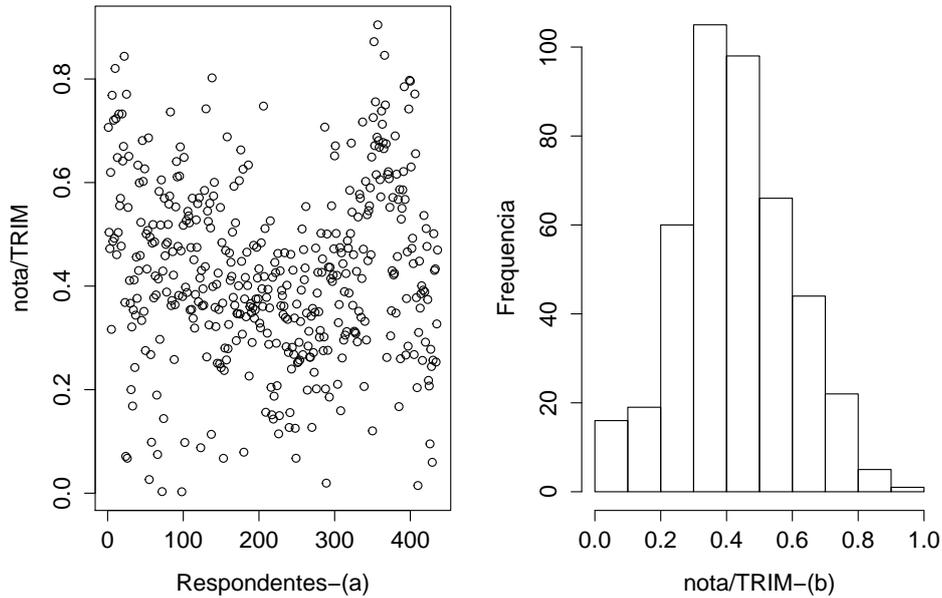


Figura 26: Escore dos respondentes pela TRIU: (a) dispersão entre as Notas/TRIU dos respondentes e (b) histograma das Notas/TRIU

A Figura 26 (a) proporciona uma visualização geral de como os alunos sucederam na avaliação objetiva do Enade/2009 com 32 questões, ou seja, poucos alunos tiveram nota acima de 0,8. A grande maioria ficou entre 0,4 a 0,6.

Na Figura 26 (b) é possível visualizar a assimetria positiva com a concentração de notas entre 0,3 a 0,5. Logo, os alunos de estatística que responderam o Enade/2009 tiveram, na prova objetiva, notas abaixo da média (7,0), considerada pelas instituições.

Com as análises descritas, é possível detalhar os itens do Enade/2009 e ter o apanhado geral da prova e o desempenho dos concluintes do curso de Estatística. Como a informação não tem a pretensão de ser conclusiva e finita na abordagem exploratória do tema, segue, na subseção 4.3.3 o método com “a priori” para estimar o parâmetro do acerto ao acaso.

4.3.3 Estimação Bayesiana para o parâmetro do acerto ao acaso

A perspectiva do processo avaliativo na abordagem do método bayesiano, haja vista a metodologia 3.2.3, é a aplicação do parâmetro de acerto ao acaso no modelo da TRIM que restringe a estimação “a priori”, conforme o Apêndice C.

Os modelos com três parâmetros estimados foram definidos como 3PL. Nesta

abordagem, são estimados os três parâmetros, com a restrição do parâmetro “chute” seguir “a priori”, definido com 3PLp. Como a recomendação já abordada por Klein (2003) é usar, inicialmente, um modelo unidimensional com “priors”, neste caso, adota-se essa análise especialmente para os parâmetros c e verifica-se o comportamento nos modelos com dois traços latentes para os dados do Enade.

Primeiro, na Tabela 30 verificam-se os resultados diagnósticos da qualidade do ajuste do modelo.

Tabela 30: Resultados das estatísticas de diagnóstico da qualidade do ajuste dos modelos

MLD1	436 respondentes		
	logLik	AIC	BIC
3PLp ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = pc_i$)	-7683,80	15495,60	15756,57
MLD2			
3PLp ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_i, d_i = d_i, c_i = pc_i$)	-7639,0	15468,01	15855,39

Nota:

- modelo logístico de um traço latente (MLD1);
- modelo logístico de dois traços latentes (MLD2); e
- modelo de três parâmetros logísticos com o acerto ao acaso “priors” (3PLp).

Na Tabela 30, os critérios de AIC e BIC indicaram o modelo MLD1 e MLD2, respectivamente; e a verossimilhança maximizada indicou o MLD2. Seguindo a mesma metodologia desenvolvida para os outros conjuntos de dados, verifica-se o teste da razão da verossimilhança com 95%, assim compensa ir para o MLD2.

Para indicar o modelo, segue a estimação do MLD1 e MLD2 para os três parâmetros em que o acerto ao acaso foi fixado com “a priori” para ambos os modelos, conforme mostra a Tabela 31.

Na Tabela 31, o acerto ao acaso é fixo com “a priori” para MLD1 e MLD2, tendo a variação entre 0,091 a 0,245 em comparação com a Tabela 27 do MLD1 e MLD2 em 3PL. O c_i , em muitas questões, foi estimado em zero, com a metodologia adotada na subseção 3.2.5, os valores do c_i seguem o conhecimento pré estabelecido da distribuição da log-normal.

O aumento do parâmetro “ c ” diminui a informação do item, pois aumenta a chance do acerto casual, com isso, aborda-se “a priori” e restringe-se a estimação “chute”.

Na Tabela 31, o item que apresenta maior discriminância e dificuldade é o Q18 em MLD1 e MLD2. Assim, para indicar o modelo no Enade do acerto ao acaso com “a priori”, baseado na metodologia abordada, segue a Tabela 32, que representa as cargas

Tabela 31: Estimação dos parâmetros com o acerto ao acaso a *priori* no MLD1 e MLD2

Q	MLD1			MLD2			
	α_i	d_i	c_i	α_{i1}	α_{i2}	d_i	c_i
G1	0,604	-7,506	0,109	2,758	4,513	-17,484	0,109
G2	0,249	-0,089	0,201	-0,19	0,57	-0,096	0,201
G3	0,745	-0,225	0,172	0,188	0,801	-0,233	0,172
G4	0,736	-1,562	0,188	0,764	0,041	-1,554	0,188
G5	0,913	-1,162	0,175	0,345	0,917	-1,23	0,175
G6	0,987	-0,281	0,165	0,261	1,275	-0,327	0,165
G7	1,452	2,047	0,175	-0,857	5,746	5,156	0,175
G8	1,609	2,517	0,173	0,228	1,943	2,835	0,173
Q11	0,857	-0,403	0,177	0,807	0,158	-0,404	0,177
Q12	0,721	-0,072	0,157	0,061	0,972	-0,085	0,157
Q13	2,635	-4,134	0,091	4,315	-1,289	-5,651	0,091
Q15	1,805	-2,344	0,245	1,334	0,948	-2,454	0,245
Q16	1,531	-1,615	0,164	1,055	0,675	-1,569	0,164
Q17	2,619	-4,1	0,186	2,039	1,091	-4,199	0,186
Q18	6,745	-11,294	0,115	14,516	-1,596	-23,078	0,115
Q19	2,526	-3,283	0,15	2,267	0,302	-3,185	0,15
Q20	2,54	-5,032	0,121	1,9	0,578	-4,593	0,121
Q21	1,757	-2,138	0,125	1,768	0,144	-2,17	0,125
Q23	3,598	-6,716	0,225	3,548	0,311	-6,984	0,225
Q24	0,922	-1,79	0,158	0,746	0,282	-1,787	0,158
Q25	2,082	-3,43	0,097	1,594	1,276	-3,843	0,097
Q26	1,847	-2,653	0,103	1,393	0,851	-2,692	0,103
Q27	1,618	-0,597	0,133	1,2	0,753	-0,582	0,133
Q28	1,529	-2,161	0,17	1,805	-0,175	-2,361	0,17
Q29	1,403	-2,665	0,205	1,599	-0,326	-2,689	0,205
Q30	0,596	-0,54	0,19	0,601	0,065	-0,539	0,19
Q31	1,455	-0,89	0,109	1,133	0,534	-0,868	0,109
Q32	0,531	-0,415	0,187	0,253	0,449	-0,423	0,187
Q33	2,445	-2,474	0,146	1,905	0,919	-2,47	0,146
Q34	3,399	-5,317	0,121	3,245	-0,2	-4,919	0,121
Q36	2,317	-3,767	0,126	2,203	0,042	-3,632	0,126
Q37	2,094	-1,54	0,213	1,857	0,484	-1,491	0,213

fatoriais, de como os itens carregam e explicam na habilidade.

Na Tabela 32, os carregamentos indicam o F1 com mais itens sendo explicados, tanto no MLD1 como MLD2, e com aplicação da “*priori*” no acerto ao acaso, as cargas carregaram mais com a comunalidade expressa mais próxima de 1, ou seja, estão melhores explicadas para a habilidade avaliada.

Observa-se, na Tabela 32, que as cargas fatoriais mais significantes estão no fator 1, portanto, é difícil avaliar a contribuição do F2. Para melhorar o *insight* sobre o real significado destes dois fatores, pondera-se e leva-se em consideração que o exame é multidisciplinar e que, as 8 primeiras questões pertencem aos conhecimentos gerais e as 24 questões restantes, aos conteúdos dos graduados de estatística.

Outro fator importante é a questão G2, que não carregou em nenhum dos dois

Tabela 32: Cargas fatoriais dos MLD1 e MLD2 com o 3PLp no CD3

Q	MLD1 - 3PLp		MLD2 - 3PLp		
	F1	h2	F1	F2	h2
G1	0,32	0,1	0,9	0,39	0,97
G2	0,14	0,02	0,11	0,27	0,09
G3	0,4	0,16	0,37	0,28	0,22
G4	0,4	0,16	0,4	-0,11	0,17
G5	0,47	0,22	0,46	0,29	0,3
G6	0,5	0,25	0,5	0,4	0,41
G7	0,65	0,42	0,53	0,8	0,92
G8	0,69	0,47	0,57	0,52	0,59
Q11	0,45	0,2	0,46	-0,07	0,21
Q12	0,39	0,15	0,36	0,36	0,25
Q13	0,84	0,71	0,75	-0,53	0,85
Q15	0,73	0,53	0,74	0,12	0,56
Q16	0,67	0,45	0,65	0,08	0,43
Q17	0,84	0,7	0,84	0,06	0,71
Q18	0,97	0,94	0,91	-0,4	0,99
Q19	0,83	0,69	0,8	-0,16	0,67
Q20	0,83	0,69	0,79	-0,06	0,62
Q21	0,72	0,51	0,71	-0,17	0,54
Q23	0,9	0,82	0,88	-0,21	0,82
Q24	0,48	0,23	0,47	-0,01	0,22
Q25	0,77	0,6	0,8	0,16	0,67
Q26	0,73	0,54	0,74	0,08	0,56
Q27	0,69	0,47	0,69	0,08	0,49
Q28	0,67	0,45	0,66	-0,28	0,51
Q29	0,64	0,4	0,58	-0,32	0,44
Q30	0,33	0,11	0,34	-0,08	0,12
Q31	0,65	0,42	0,64	0,02	0,42
Q32	0,3	0,09	0,3	0,14	0,11
Q33	0,82	0,67	0,82	0,04	0,67
Q34	0,89	0,8	0,82	-0,32	0,78
Q36	0,81	0,65	0,76	-0,23	0,63
Q37	0,78	0,6	0,77	-0,08	0,6

fatores. Na comunalidade e nas cargas fatoriais, segundo Pasquali (2006), a indicação de exclusão do item é mais uma vez observada, pois explica menos 30% da prova com comunalidade abaixo de 0,15. A Q23, indicada a ser retirada, tem explicação no F1 com 0,9 e com a comunalidade de 0,82, próxima de 1.

A Tabela 32 indica que o melhor modelo é o MLD1 com 3PLp e o exame não reportou as duas habilidades que deveriam ser avaliadas, ou seja, para essa prova, há a suposição de que existe outro fator que explica e se sobressai nas duas habilidades, por exemplo, a interpretação. Para afirmar qual habilidade é essa, são necessários especialistas que verifiquem todas as questões e relatem a área de conhecimento fundamental para cada item.

A CCI do MDL1 com 3PLp, representada na Figura 26, implica toda a informação da estimação dos seguintes itens: dificuldade do item e seu poder de discriminação.

Na Figura 27 estão representadas 32 curvas características do item no mesmo

gráfico. Todas apresentam o acerto ao acaso “a priori”, ou seja, possuem “chute” pré-definido.

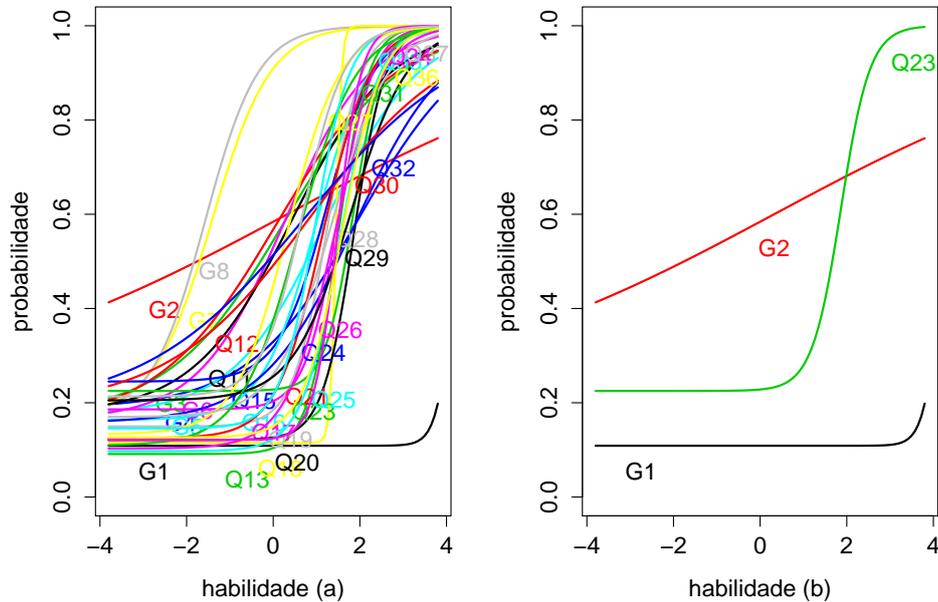


Figura 27: CCI do MLD1-3PLp : (a) com 32 itens e (b) itens Q1, Q2 e Q23 do Enade-2009 de estatística

Para o nível de discriminação, a curva com maior formato de S e que tem um alto nível de discriminação, é o item Q18, enquanto a G1 possui um poder de discriminação baixo que é a curva mais próxima da habilidade, indicando o item difícil.

Num contexto em que as respostas podem ser certas ou erradas, as duas curvas da esquerda, próximas de G7 e G8, representam um item fácil porque a probabilidade de resposta correta é alta para habilidade baixa.

Para abordar o escore dos 436 respondentes do Enade-2009 com os 32 itens, e o MLD1 - 3PLp, a Figura 28 representa a informação geral da habilidade dos avaliados.

A metodologia aplicado na subseção 3.2.8 é a escala padronizada entre 0 e 1 e representa os escores do MLD1-3PLp. Na Figura 28 do histograma, a habilidade com maior frequência está entre 0,3 a 0,4 e decresce quando a habilidade aumenta. A frequência de alunos entre 0,4 a 0,5 decresceu rapidamente em comparação à anterior, aproximadamente de 120 para 90, ou seja, para cada intervalo da escala padronizada, as diferenças apresentadas são em decorrência do acerto ao acaso.

Na Figura 28, o gráfico de correlação entre a nota pela TCT e a escala padronizada apresenta somente duas notas que se dispersaram do aglomerado de informações, mas, a

grande maioria dos avaliados possui a nota pela TCT e pela TRIM, indicada pelos MLD1 - 3PLp, que se aproximam.

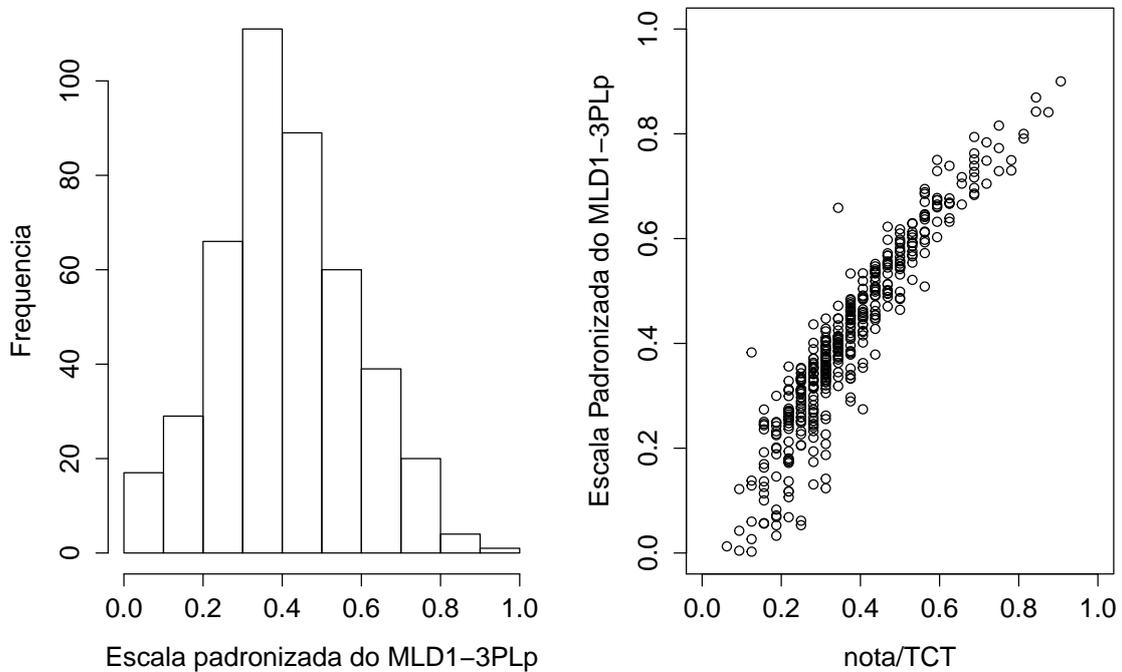


Figura 28: Escala padronizada e a correlação das notas da TCT com a escala padronizada do MLD1-3PLp

Esta pesquisa não tem a pretensão de ser conclusiva nem finita na abordagem exploratória do tema, visto que os resultados apresentados nesta subseção restringiram-se ao CD3, mas, a aplicabilidade da metodologia pode ser estendida aos outros conjuntos de dados.

Outra abordagem é analisar somente as 24 questões específicas da prova do Enade, visto que o estudo, quando tratou de todos os itens, 8 gerais e 24 específicos, o modelo escolhido foi de um traço latente, assim, a intenção é verificar a consistência do traço latente nos resultados, excluídos os itens gerais.

A Figura 29 auxilia a identificação da dimensionalidade em estudo e é possível verificar as dimensões interpretáveis, pelos autovalores, com o indicativo de um traço latente quando retirou-se as questões gerais.

Visualmente, comparou-se com a Figura 22, que indicava dois traços latentes e diante disso, optou-se por verificar o ajuste e também o diagnóstico da qualidade do ajuste dos modelos uni/multidimensionais.

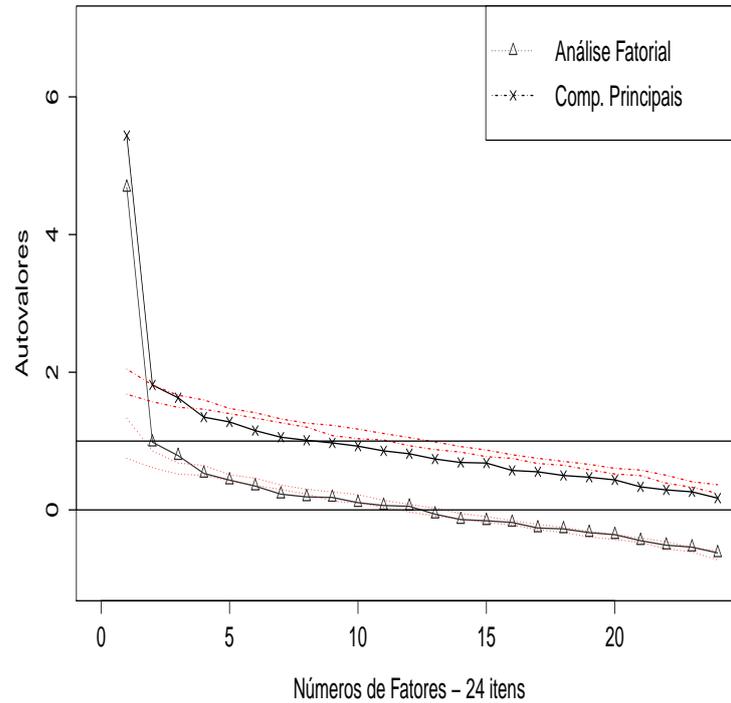


Figura 29: Diagnóstico da dimensão dos 436 respondentes com 24 questões específicas

Na Tabela 33, verificam-se os critérios de AIC e BIC que indicaram o modelo MDL1 e MLD2, respectivamente, e a verossimilhança maximizada indicou o MLD2. No entanto, houve indicação na dimensionalidade para um traço latente, assim, manteve-se o MDL2 para comparar com os resultados quando consideradas as 32 questões, o que indicou também a mesma análise.

Tabela 33: Resultados das estatísticas de diagnóstico da qualidade do ajuste dos modelos para as 24 questões específicas

MLD1	436 respondentes		
	logLik	AIC	BIC
3PLp ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = pc_i$)	-5812,20	11732,56	11916,14
MLD2			
3PLp ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_i, d_i = d_i, c_i = pc_i$)	-5783,68	11709,37	11998,89

Nota:

- modelo logístico de um traço latente (MLD1);
- modelo logístico de dois traços latentes (MLD2); e
- modelo de três parâmetros logísticos com o acerto ao acaso “priori” (3PLp).

A estimação dos parâmetros com o acerto ao acaso “a priori”, para as 32 questões na Tabela 31, mostrou a Q18 e a Q23 com maior discriminação e dificuldade em ambos, MLD1 e MLD2. Na Tabela 34, aplicou-se a mesma análise para verificar se ao excluir as

questões gerais esses mesmos itens teriam outros resultados dos parâmetros estimados.

Tabela 34: Estimação dos parâmetros com o acerto ao acaso *a priori* no MLD1 e MLD2

Q	MLD1			MLD2			
	α_i	d_i	c_i	α_{i1}	α_{i2}	d_i	c_i
Q11	0.87	-0.45	0.19	1.1	0.09	-0.48	0.19
Q12	0.61	-0.12	0.18	-0.25	-1.08	-0.15	0.18
Q13	2.81	-4.36	0.09	3.06	-0.05	-4.65	0.09
Q15	1.35	-1.77	0.22	0.19	-1.91	-2.25	0.22
Q16	1.45	-1.63	0.17	1.07	-0.5	-1.62	0.17
Q17	1.98	-3.18	0.17	1.23	-0.96	-3.19	0.17
Q18	9.09	-15.12	0.11	11.27	-3.68	-23.2	0.11
Q19	2.36	-3.07	0.14	1.64	-0.95	-3.05	0.14
Q20	2.35	-4.85	0.12	-0.69	-14.14	-23.88	0.12
Q21	1.86	-2.18	0.12	2.39	0.17	-2.48	0.12
Q23	3.46	-6.62	0.22	6.01	-0.62	-12.24	0.22
Q24	1.03	-2.02	0.18	1.1	0.01	-2.06	0.18
Q25	1.8	-3.09	0.09	0.64	-1.67	-3.43	0.09
Q26	1.66	-2.42	0.1	0.54	-1.51	-2.6	0.1
Q27	1.59	-0.59	0.14	1.09	-0.67	-0.59	0.14
Q28	1.69	-2.37	0.18	1.68	-0.31	-2.59	0.18
Q29	1.18	-2.31	0.19	2.48	1.01	-3.01	0.19
Q30	0.63	-0.52	0.19	-0.12	-0.9	-0.55	0.19
Q31	1.42	-0.92	0.12	0.77	-0.83	-0.92	0.12
Q32	0.55	-0.42	0.19	-0.09	-0.77	-0.44	0.19
Q33	2.22	-2.28	0.14	1	-1.99	-2.66	0.14
Q34	3.19	-5.01	0.12	2.21	-1.46	-5.19	0.12
Q36	1.91	-3.15	0.11	1.62	-0.45	-3.19	0.11
Q37	1.8	-1.15	0.18	0.58	-1.77	-1.31	0.18

A Tabela 34 mostra que a estimação dos parâmetros dos itens com “a priori” no acerto ao acaso, somente para os itens específicos Q18 e Q23, continuam indicando valores altos de dificuldade e discriminação em ambos os traços latentes. Há indicação, também, de que estas questões sejam retiradas da prova.

O estudo o modelo indicado é o MLD1 mesmo considerando as questões gerais ou somente abordando as questões específicas. Assim há indícios que as questões gerais não apresentam consistência para influenciar o traço latente. Como também é possível observar na Tabela 35 que as cargas fatoriais carregam no F1 tanto no MLD1 como no MLD2.

Ao comparar a Tabela 35 com a 32, em que os 32 itens foram analisados, verificou-se poucas diferenças nas cargas fatoriais, por exemplo, a questão Q18, que possuía a carga fatorial no MLD1 com o valor de 0,97 e ao excluir todas as questões gerais, a carga fatorial passou ter o valor de 0,98 e na Q23, em ambas as Tabelas, o $F1$ apresentou o mesmo valor. Contudo, o estudo das cargas fatoriais mostra que existem poucas diferenças ao

Tabela 35: Cargas fatoriais dos MLD1 e MLD2 com priori

Q	MLD1 - 3PLp		MLD2 - 3PLp		
	F1	h2	F1	F2	h2
Q11	0.45	0.21	0.47	0.24	0.28
Q12	0.34	0.11	0.34	-0.35	0.24
Q13	0.85	0.73	0.81	0.32	0.77
Q15	0.62	0.39	0.67	-0.37	0.58
Q16	0.65	0.42	0.64	0.06	0.41
Q17	0.76	0.57	0.75	-0.02	0.56
Q18	0.98	0.97	0.98	0.17	0.98
Q19	0.81	0.66	0.81	0.04	0.65
Q20	0.81	0.66	0.81	-0.58	0.98
Q21	0.74	0.54	0.72	0.35	0.65
Q23	0.9	0.8	0.92	0.3	0.93
Q24	0.52	0.27	0.5	0.21	0.29
Q25	0.73	0.53	0.74	-0.25	0.6
Q26	0.7	0.49	0.7	-0.25	0.55
Q27	0.68	0.47	0.68	0.02	0.46
Q28	0.7	0.5	0.71	0.19	0.55
Q29	0.57	0.33	0.54	0.56	0.6
Q30	0.35	0.12	0.34	-0.28	0.19
Q31	0.64	0.41	0.64	-0.07	0.41
Q32	0.31	0.09	0.3	-0.25	0.15
Q33	0.79	0.63	0.82	-0.22	0.71
Q34	0.88	0.78	0.89	0.01	0.79
Q36	0.75	0.56	0.73	0.14	0.56
Q37	0.73	0.53	0.74	-0.27	0.62

manter ou retirar as questões da prova do Enade.

Para comparação, segue a Figura 30 da escala padronizada, considerando os 32 itens, ou seja, as questões gerais e específicas em correlação com a escala padronizada dos 24 itens específicos.

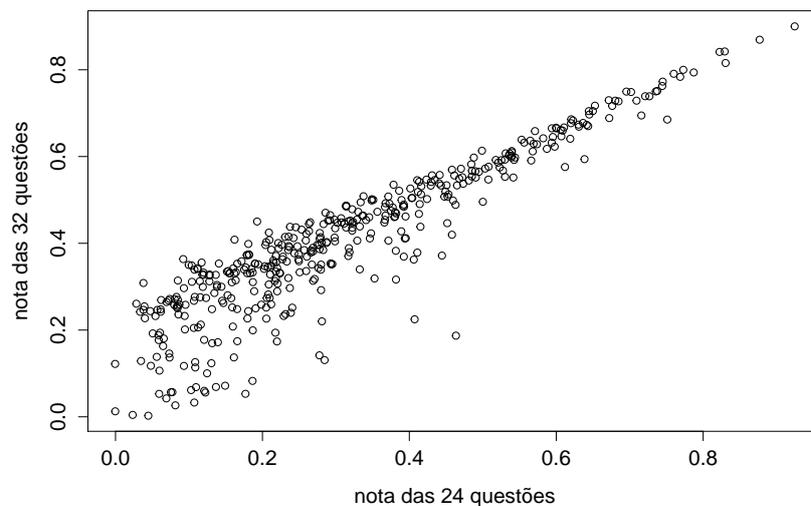


Figura 30: Correlação das notas com a escalas padronizadas das 32 e 24 questões

Assim, na Figura 30 pode-se observar que, para os 436 alunos do Curso de Esta-

tística que realizaram a prova objetiva do Enade, considerando as 32 questões (8 gerais e 24 específicas) ou apenas as 24 questões específicas, existe correlação entre as notas. Há a constatação, dentro do total de indivíduos, que há alguns alunos com notas diferenciadas da aglomeração. Ressalta que ambas as notas é de um traço latente e a consistência das questões gerais não indicam abordar mais do que uma dimensão.

Com os resultados, o Enade pode contar com mais um instrumento para a busca da melhoria da qualidade do ensino e desempenho do aluno. Dentre os métodos discutidos esta pesquisa não encerra as análises exploratórias baseadas em TRI, visto que os desafios e perspectivas em avaliação estão em constante aperfeiçoamento.

5 CONCLUSÕES E SUGESTÕES DE TRABALHOS FUTUROS

Esta tese apresentou o emprego dos modelos de TRI a um exame multidisciplinar, aplicado pela PUCPR, aos alunos de graduação e os modelos de TRIU e TRIM no Enade/2009 no curso de graduação de Estatística para 443 concluintes. Em ambas as investigações, as perspectivas são os desafios da modelagem e análise de avaliações multidisciplinares que não utilizam a TRI. Os resultados, com o uso dos modelos aplicados em amostras de pequena, média e grande escalas, mostram que, em geral, a prova apresenta boa coerência interna nos diagnósticos da qualidade dos ajustes dos modelos bem como boa aderência ao instrumento de medida proposto.

A TRI, aplicada sem uma análise de critérios estatísticos de diagnósticos, resulta em modelos que não condizem com a estrutura dos exames que são multidisciplinares, uma vez que medidas e qualidades de ajustes estatísticos auxiliam, porém, não substituem a capacidade de interpretação do pesquisador. Dizer que o modelo da TRI foi igual, aproximado ou se diferencia com os resultados do TCT, sem uma análise dos resultados na escolha do modelo, pode levar a conclusões precipitadas.

Desta maneira, um estudo mais metucioso, como o feito sobre o conjunto de dados I,II e III, reforça a necessidade processual para se avaliar adequadamente a aprendizagem. O fato de os itens terem sido explicados a partir de um traço latente em uma prova multidisciplinar, a comparação dos resultados por escores, implica na necessidade de especialistas que fundamentem cada uma das áreas analisadas.

Na perspectiva de analisar uma avaliação multidisciplinar, indica-se um modelo com mais de um traço latente, quando 30% dos itens estão associados a outro traço latente. O modelo proposto para os conjuntos de dados oferece a mais completa descrição possível dos respondentes, conforme as influências da quantidade de indivíduos, de itens em habilidades e de parâmetros, ou seja, a quantidade estimada dos parâmetros deve ser menor do que os padrões de respostas.

Assim, buscou-se identificar o modelo mais adequado para se avaliar o desempenho do educando, visando as habilidades necessárias para responder os itens. Foram também observados diversos fatores múltiplos, que representam diferentes misturas de habilidades necessárias para as questões. O que se evidencia é a necessidade de uma estrutura de avaliação que detecte e considere esta complexidade de múltiplas habilidades e que seja capaz de identificar qual a habilidade dominante que representa vários conhecimentos em um único modelo.

Na TRIM, o modelo exploratório apresenta eficiência e detecta as influências entre os testes avaliados e representa adequadamente a possibilidade de pesquisar a dimensão do conjunto de habilidades observadas com a indicação do modelo.

O modelo da TRIU supõe que todas as questões avaliam a dimensão latente. Contudo, para acompanhar o desenvolvimento das habilidades específicas, verificar o aprendizado, o desenvolvimento do uso destas habilidades e qual delas o candidato está usando, há necessidade de utilizar a TRIM com análise fatorial, dos componentes principais e das cargas fatoriais. Isso indica qual item explica a habilidade em estudo e o conjunto de itens, que compõem a prova, determinam a habilidade dominante.

Outro fator de resposta foi detectado quando constatou-se, no conjunto de dados I e II, na escolha do modelo e na estimação dos parâmetros, que o acerto ao acaso não representou de modo satisfatório o valor esperado da probabilidade de acerto casual, devido aos zeros inflacionados. O conjunto de dados II mostrou, que o modelo foi capaz de receber influências de acordo com a quantidade de zeros retirados do total de 35 alunos que zeraram a prova.

Ao detalhar o modelo pela TRI, no conjunto de dados III, foi verificado que os itens requerem considerações e não devem ser excluídos somente pela análise descritiva, antes de se verificar as CCIs e as cargas fatoriais. Essas indicações são relevantes para tomada de decisões precisas e baseadas na prova, não como um todo, mas em itens. E, nesse caso, quanto a dificuldade de cada item com diferentes níveis de discriminância e probabilidade de acerto casual, apresentaram resultados concisos.

Estas questões são relevantes para o processo de avaliação educacional e, devem ser consideradas na construção das provas que usam a TRI para serem adequadas e capazes de representar o exame.

Com os estudos, para indicar os modelos, foram propostos passos, que não apenas melhorem ou ajustem os conjuntos de dados, mas que contemplem aspectos da avaliação

multidisciplinar. Um procedimento foi não usar a escala métrica da TRI, já que os exames não a utilizam. A proposta foi de apresentar os escores dos modelos na escala padronizada entre 0 a 1. Esta convenção foi estendida para os modelos da TRIM e representa os escores de mais de uma habilidade com ponderação do conhecimento em uma habilidade composta.

A principal contribuição da TRI para os exames multidisciplinares é apresentar, de forma sistemática, a análise, não apenas dos alunos, mas também do instrumento, dando uma visão crítica de sua construção e capacidade em aferir o conhecimento dos alunos com a escala padronizada entre 0 a 1.

Deixa-se aqui, como proposta para futuras pesquisas, a avaliação deste exame considerando modelos da TRIM, que sejam capazes de contemplar mais de uma habilidade latente avaliada, como exemplo, o confirmatório. Da mesma forma, faz-se também necessário considerar que várias habilidades podem influenciar a resposta de cada item.

O pacote *mirt* que foi utilizado nos modelos da TRIM está em desenvolvimento até o presente momento e análises futuras dos escores ajustados pelas definições das estimativas dos modelos devem ser refeitas, até para futuras comparações dos resultados discutidos, juntamente com a abordagem de métodos bayesianos. Com a atualização deste pacote *mirt*, seria relevante rever as estimações dos parâmetros, que receberam sinais negativos e positivos neste estudo, necessitando a inversão dos sinais e demandando cuidado nas análises de cada item.

Outro fator importante de futuros estudos que auxiliarão na condução de futuros estudos é a utilização “a priori” no acerto ao acaso para o CD2 e CD1, como visto no CD3, que mostrou um desempenho favorável nas cargas, por explicar melhor os itens nos fatores como também indicações no uso da estimação bayesiana para os três parâmetros dos itens e a análise do modelo confirmatório.

Contudo, pode-se observar, com os resultados aqui discutidos que, na avaliação, a elaboração e composição dos itens tem um papel importante e deve ser uma aliada, tanto para o discente como para o docente, a serviço da aprendizagem.

Desta maneira, pode-se dizer que além dos resultados das avaliações multidisciplinares soma-se mais um instrumento - a TRI - para buscar melhorias na qualidade de ensino, desempenho do aluno e aporte, que pode auxiliar os professores a identificar as disciplinas/habilidades que geram mais dificuldades e que necessitam de mais atenção no desenvolvimento do conhecimento dos envolvidos na educação.

Referências

- ADAMS, R. J.; WILSON, M.; WANG, W. C. The multidimensional random coefficients multinomial logit model. **Applied Psychological Measurement**, v. 21(1), p. 1–23, 1997.
- ANDRADE, D. F.; BORTOLOLLI, S. Aplicação de um modelo de desdobramento graduado generalizado-ggum da teoria da resposta ao item. **Estudos Avaliação Educ**, v. 18(37), p. 157–87, 2007.
- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria da resposta ao item: conceitos e aplicações**. São Paulo: Associação Brasileira de Estatística - ABE, 2000.
- ANDRIOLA, W. B. Uso da teoria de resposta ao item (tri) para analisar a equidade do processo de avaliação do aprendizado discente. **Revista Iberoamericana de Avaliação Educacional**, v. 1, p. 171–189, 2008.
- ANDRIOLA, W. B. Psicometria moderna: características e tendências. **Estudos em Avaliação Educacional**, São Paulo, v. 20(43), mai-ago 2009.
- ARIAS, M. R. M.; LLOREDA, M. J. H.; LLOREDA, M. V. H. **Psicometria**. Madrid: Alianza, 2006.
- BAI, J.; NG, S. Determining the number of factors in approximate factor models. **Econometrica**, v. 70, p. 191–221, 2002.
- BAKER, F. B. **The Basics of Item Response Theory**. 2. ed. New York: CRC Press, 2001.
- BAKER, F. B.; KIM, S. **Item Response Theory: parameter estimation techniques**. 2. ed. New York: CRC Press, 2004.
- BARBETTA, A.; TREVISAN, M. V. L. A.; AL. et. Análise da prova de conhecimentos gerais do vestibular da unesp-2012 por modelos uni e multidimensionais da teoria da resposta ao item. In: III CONGRESSO BRASILEIRO DE TEORIA DE RESPOSTA AO ITEM - III CONBRATRI. Belém, PA: Associação Brasileira de Avaliação Educacional (ABAVE), 2013. p. 4 – 6.
- BARTHOLOMEW, D. J.; KNOTT, M. Latent variable models and factor analysis. In: LORD, F. M.; NOVICK, M. R. (Ed.). **Kendall's Library of Statistics**. 2. ed. Reading, MA: Addison-Wesley, 1999. p. 395–479.
- BATTAUZ, M. Laplace approximation in measurement error models. **Biometrical Journal**, v. 53(3), p. 411–425, 2011.
- BIRNBAUM, A. Some latent trait models and their use in inferring and examinee's ability. In: LOED F. M., L. M. R. (Ed.). **Statistical theories of mental test scores**. Reading, MA: Addison Wesley, 1968. p. 395–479.

- BOCK, R.; AITKIN, M. Marginal maximum likelihood estimation of item parameters: Application of an em algorithm. **Psychometrika**, v. 46, p. 443–459, 1981.
- BOCK, R. D.; GIBBONS, R.; MURAKI, E. Full information item factor analysis. **Applied Psychological Measurement**, v. 12, p. 261–280, 1988.
- BOCK, R. D.; LIEBERMAN, M. Fitting a response model to n dichotomously scored items. **Psychometrika**, v. 35, p. 179–197, 1970.
- BONAT, W. H.; RIBEIRO JR, P. J.; KRAINSKI, E. T.; ZEVIANI, W. M. **Métodos Computacionais em Inferência Estatística**. Curitiba: Associação Brasileira de Estatística-ABE, SINAPE, 2012.
- BOOMSMA, A.; DUIJN, M. A. J. V.; SNIJDERS, T. A. B. **Essays on item response theory. Lecture Notes in Statistics**. New York: Springer, 2000.
- CAI, L. High-dimensional exploratory item factor analysis by a metropolis-hastings robbins-monro algorithm. **Psychometrika**, v. 75, p. 33–57, 2010.
- CAMPBELL, D.; STANLEY, J. **Experimental and quasi-experimental designs for research**. Chicago: Rand McNally, 1973.
- CHACHAMOVICH, E. **Teoria de Resposta ao Item: Aplicação do modelo de Rasch em desenvolvimento e validação de instrumentos em saúde mental**. Tese (Tese de doutorado) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.
- CHALMERS, P. mirt: A multidimensional item response theory package for the r environment. **Journal of Statistical Software**, v. 48(6), p. 1–29, May 2012. Disponível em: <<http://www.jstatsoft.org/v48/i06/>>.
- COHEN, K. R.; HENIK, A.; LINDEN, D. E. J. Processing conflicting information: facilitation, interference, and functional connectivity. **J. Neuropsychologia**, v. 46, p. 2872–2879, Oct. 2008.
- CRONBACH, L. J. Coefficient alpha and the internal structure of tests. **Psychometrika**, v. 16, p. 297–334, 1951.
- DE ARAUJO, E. A. C.; ANDRADE, D. F.; BORTOLOTTI, S. L. V. Teoria da resposta ao item. **Revista da Escola de Enfermagem USP**, São Paulo, v. 43 (Esp), p. 1000–1008, 2009.
- DEMPSTER, A.; LAIRD, N.; RUBIN, D. Maximum likelihood from incomplete data via the em algorithm. **Journal of the Royal Statistical Society, Series B**, v. 39(1), p. 1–38, 1977.
- DEPRESBITERIS, L. Avaliação da aprendizagem na educação ambiental - uma relação muito delicada. In: SANTOS, J. E.; SATO, M. (Ed.). **A contribuição da Educação Ambiental à Esperança de Pandora**. São Carlos: RIMA, 2001.
- D'HAINAUT, L. **Conceitos e Métodos da Estatística. Volume II: Duas ou três variáveis segundo duas ou três dimensões**. Lisboa: Fundação Calouste Gulbenkian, 1992.

- DIVGI, D. R. Calculation of the tetrachoric correlation coefficient. **Psychometrika**, v. 44(2), p. 169–172, 1979.
- EDWARDS, M. A. Markov chain monte carlo approach to confirmatory item factor analysis. **Psychometrika**, v. 75(3), p. 474–497, 2010.
- EMBRETSON, S.; REISE, S. P. **Item Response Theory for Psychologists**. New Jersey: Lawrence Erlbaum Associates, 2000.
- FRAGOSO, T. M. **Modelos Multidimensionais da Teoria de Resposta ao Item. Dissertação de Mestrado**. Dissertação (Mestrado) — Instituto de Ciências Matemáticas e de Computação - USP, São Carlos - SP, 2010.
- FRANCISCO, R. **Aplicação da Teoria da Resposta ao Item (TRI) no Exame Nacional de Cursos (ENC) da Unicentro**. Dissertação (Mestrado) — Universidade Federal do Paraná - UFPR, Curitiba, 2005. Pós- Graduação em Métodos Numéricos em Engenharia.
- GIBBONS, R.; HEDEKER, D. Full-information item bi-factor analysis. **Psychometrika**, v. 57, p. 423–436, 1992.
- GONZALEZ, B. J. Bayesian methods in psychological research: The case of irt. **International Journal of Psychological Research**, v. 3 (1), p. 164–176, 2010.
- GREEN, B.; BOCK, R.; HUMPHREYS, L.; LINN, R.; RECKASE, M. Technical guidelines for assessing computerized adaptive tests. **Journal of Educational Measurement**, v. 21, p. 347–360, 1984.
- GULLIKSEN, H. **Theory of mental tests**. New York: Wiley, 1950.
- HAMBLETON, R. K.; JONES, R. W. Comparison of classical test theory and item response theory and their applications to test development. **Educational Measurement: Issues and Practice**, v. 12(3), 1993.
- HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. **Fundamentals of item response theory**. Newbury Park, CA: Sage, 1991.
- HOLZINGER, K.; HARMAN, H. **Factor analysis: A synthesis of factorial methods**. Chicago: The University of Chicago, 1941.
- HORN, J. A rationale and test for the number of factors in factor analysis. **Psychometrika**, v. 32, p. 179–185, 1965.
- INEP. **Exame Nacional de Desempenho dos Estudante - Manual do Enade**. [S.l.], 2004 a 2013. Acessado em: 23 Jan. 2012. Disponível em: <[http://portal.inep.gov.br/manual-do-enade.Manual\(2004-2013\).pdf](http://portal.inep.gov.br/manual-do-enade.Manual(2004-2013).pdf)>.
- INEP. **Exame Nacional de Desempenho dos Estudante - INEP - Relatório Síntese 2009**. [S.l.], 2009. Acessado: 12 nov. 2012. Disponível em: <<http://portal.inep.gov.br/web/guest/relatorio-sintese-2009.RelatórioSíntese:Estatística2009.pdf>>.
- KIM, S.-H. Bilog 3 for windows: Item analysis and test scoring with binary logistic models (software review). **Applied Psychological Measurement**, v. 21, p. 371–376, 1997.

- KLEIN, R. Utilização da teoria de resposta ao item no sistema nacional de avaliação da educação básica (saeb). **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 111, p. 283–296, 2003.
- KLEM, L. **Structural equation modeling**. Em L. G. Grimm & P. R. Yarnold (Orgs.), **Reading and understanding more multivariate statistics** (pp. 227-260). Washington: American Psychological Association, 2002.
- LAZERSFELD, P. **Latent structure analysis**. In S.Koch, **Psychology: A study of science**. New York: McGraw-Hill, 1959. 476-543 p.
- LEDESMA, R.; VALERO-MORA, P. . Determining the number of factors to retain in efa: an easy-to-use computer program for carrying out parallel analysis. **Practical Assessment, Research & Evaluation**, v. 12(1), p. 1–11, 2007.
- LINDEN, W. J. V. D.; HAMBLETON, R. K. **Handbook of Modern Item Response Theory**. New York: Springer-Verlag, 1997.
- LORD, F. A theory of test scores. iowa (ia): Psychometric society. **Psychometric Monograph**, v. 7, 1952.
- LORD, F. **Applications of item response theory to practical testing problems**. Hillsdale: Erlbaum, 1980.
- LORD, F. M.; NOVICK, M. R. **Statistical theories of mental test scores**. Reading, MA: Addison-Welsley, 1968.
- MARVELDE, J. M. e. a. Application of multidimensional item response theory models to longitudinal data. **Educational and Psychological Measurement**, v. 66, p. 5–34, 2006.
- MATOS, G. S. **Teoria da Resposta ao Item: Uma Proposta de Modelo Multivariado**. Dissertação (Mestrado) — Universidade Federal de Pernambuco (UFPE) - Departamento de Estatística (DE), Recife - PE, 2001.
- MATOS, G. S. **Modelos multidimensionais da TRI com distribuições assimétricas para os traços latentes**. Tese (Doutorado) — Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2008.
- Mc Donald, R. P. **Test Theory: A Unified Treatment**. Mahawah,1999: Lawrence Erlbaum, 1999.
- MOLENBERGHS, G.; VERBEKE, G. **Models for Discrete Longitudinal Data (Springer Series in Statistics)**. [S.l.]: Springer, 2005.
- MOREIRA, J. J. F. Aplicações da teoria da resposta ao item (tri) no brasil. **Rev. Bras. Biom.**, São Paulo, v. 28(4), p. 137–170, 2010.
- MULAIK, S. **The Foundations of factor analysis**. New York: McGraw-Hill, 1972.
- MURAKI, E. A generalized partial credit model: Application of an em algorithm. **Applied Psychological Measurement**, v. 16, p. 159–176, 1992.

- MURAKI, E.; CARLSON, J. E. Full-information factor analysis for polytomous item responses. **Applied Psychological Measurement**, v. 19, p. 73–90, 1995.
- MURAKI, E.; ENGELHARD, G. Full information item factor analysis: applications of eap scores. **Applied Psychological Measurement**, v. 9, p. 417–430, 1985.
- MUÑIZ, J.; HAMBLETON, R. K. Medio siglo de teoría de respuestas a los Ítems. **Anuario de Psicología**, v. 52, p. 41–66, 1992.
- NEYMAN, J.; SCOTT, E. L. Consistent estimates based on partially consistent observations. **Econometrica**, v. 16(1), p. 1–32, 1948.
- NOJOSA, R. T. **Modelos Multidimensionais para a Teoria de Resposta ao Item**. Dissertação (Mestrado) — Universidade Federal de Pernambuco - UFPE, Recife, 2001.
- NOJOSA, R. T. Teoria da resposta ao item(tri) - modelos multidimensionais. **Estudos em Avaliação Educacional**, v. 25, p. 123–166, 2002.
- PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis: Vozes, 2003.
- PASQUALI, L. **TRI - Teoria de Resposta ao Item: teoria, procedimentos e aplicações**. Brasília: LabPAM/UnB, 2007.
- PASQUALI, L. *Psicometria*. rev. esc. enferm. USP, São Paulo, v. 43(spe), 2009.
- PASQUALI, L. **Psicometria. Teoria dos testes na psicologia e na educação**. 4. ed. Petrópolis: Vozes, 2011.
- PASQUALI, L.; PRIMI, R. Fundamentos da teoria da resposta ao item - tri. **Avaliação Psicológica**, v. 2(2), dez 2003.
- PATZ, R. J.; JUNKER, B. W. A straightforward approach to markov chain monte carlo methods for item response theory models. **Journal of Educational and Behavioral Statistics**, v. 24, p. 146–178, 1999.
- PENIN, S.; MARTÍNEZ, M. **Profissão docente: pontos e contrapontos**. São Paulo: Summus, 2009.
- PITON-GONÇALVES, J. **Desafios e Perspectivas da Implementação Computacional de Testes Adaptativos Multidimensionais para Avaliações Educacionais**. Tese (Doutorado) — Universidade de São Paulo - USP, 2012.
- QUARESMA, S. E.; DIAS, S. T. C.; SARTORIO, D. S. **Avaliação da aprendizagem e das provas do centro de formação interdisciplinar/UFOPA via teoria da resposta ao item**. 2012. Disponível em: <<http://www.sbec.org.br/evt2012/trab16.pdf>>.
- R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2012. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org/>>.
- RASCH, G. **Probabilistic models for some intelligence and attainment tests**. Copenhagen: Danish Institute for Educational Research and St. Paul, 1960.

- RECKASE, M. D. **A linear logistic multidimensional model for dichotomous item response data.** In **W. J. van der Linden & R. K. Hambleton (Eds.), Handbook of modern item response theory.** New York: Springer, 1997. 271-286 p.
- RECKASE, M. D. **Multidimensional Item Response Theory. Statistical for Social and Behavioral Sciences.** [S.l.]: Springer Science Business Media: LLC, 2009.
- RECKASE, M. D.; ACKERMAN, T. A.; CARLSON, J. E. Building a unidimensional test using multidimensional items. **Journal of Educational Measurement**, v. 25, p. 193–203, 1988.
- REISE, S. P.; WALLER, N.; COMREY, A. Factor analysis and scale revision. **Psychological Assessment**, v. 12, p. 287–297, 2000.
- REQUENA, C. S. **Psicometria: Teoria y Práctica em la Construcción de Tests.** Madrid: Ediciones Norma, 1990.
- REVELLE, W. **Using the psych package to generate and test Structural Equation Models. (A vignette for the psych package).** 2013. Disponível em: <http://personality-project.org/r/psych/_for/_sem.pdf>.
- ROBERT, C.; CASELLA, G. **Introducing Monte Carlo Methods with R. Use R.** [S.l.]: Springer, 2010.
- SAMEJIMA, F. Normal ogive model on the continuous response level in the multi-dimensional latent space. **Psychometrika**, v. 39, p. 111–121, 1974.
- SCHILLING, S.; BOCK, R. High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. **Psychometrika**, v. 70, p. 533–555, 2005.
- SHENG, Y. Bayesian estimation of mirt models with general and specific latent traits in matlab. **Journal of Statistical Software**, v. 34(3), p. 1–27, 2010. Disponível em: <http://www.jstatsoft.org/v34/i03>, 2010.
- SOARES, T. M. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. **Pesquisa Operacional**, Rio de Janeiro, v. 25(1), p. 83–112, 2005.
- SOUZA, A. C.; TAVARES, H. R. A relação entre a teoria clássica de medidas e a teoria da resposta ao item. In: ENCONTRO REGIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL - ERMAC. Belém: Sociedade de Matemática Aplicada e Computacional - SBMAC, 2005.
- SPEARMAN, C. Demonstration of the formulae for true measurement of correlation. **American Journal of Psychology**, v. 18(2), p. 161–169, 1907.
- SPEARMAN, C. Correlations of sums and differences. **British Journal of Psychology**, v. 5, p. 417–426, 1913.
- THOMPSON, B. **Ten commandments of structural equation modeling.** Em **L. G. Grimm & P. R. Yarnold (Orgs.), Reading and understanding more multivariate statistics (pp. 261-283).** Washington: American Psychological, 2002.

THURSTONE, L. **Multiple-factor analysis: A development and expansion of The Vectors of Mind**. Chicago, 1947: The University of Chicago, 1947.

TRAVITZKI, R. Quantas competências o enem realmente avalia? In: ANPED (Ed.). **X Encontro de Pesquisa em Educação da Região Sudeste**. Rio de Janeiro: [s.n.], 2011.

ULLMAN, J. B. **Structural Equation Modeling**. Em **B. G. Tabachnick & L. S. Fidell (Orgs.), Using multivariate statistics (5 ed.)**. Boston: Pearson Education, 2007, 2007.

VALLE, R. d. C. Teoria da resposta ao item. **Estudos em Avaliação Educacional**, Fundação Carlos Chagas, São Paulo, v. 21, p. 7–88, 2000.

VENDRAMINI, C. M. M.; DIAS, A. S. Teoria de resposta ao item na análise de uma prova de estatística em universitários. **Psico-USF**, v. 10(2), p. 201–210, jul.-dez. 2005.

WINGERSKY, M. S.; BARTON, M. A.; LORD, F. M. **LOGIST user's guide**. Princeton: NJ, 1982.

WRIGHT, B. D.; MEAD, R. J.; BELL, S. R. **BICAL: Calibrating items with the Rasch model**. Department of Education, Statistical Laboratory, 1979. No. 23B.

APÊNDICE A – Índices de facilidade conforme a proporção de acertos, índices de discriminação de acordo com o ponto bisserial e o *Cronbach's alpha* na prova geral dos 443 respondentes.

443 respondentes						
Questões	Proporção % acertos	Ponto incluir	Bisserial excluir	<i>Cronbach's alpha</i> *Todas as questões e Excluindo a questão	Índice de Facilidade	Índice de Discriminação
				0,767*		
G1	10,30%	0,11	0,04	0,769	Muito difícil	Fraco
G2	58,30%	0,22	0,12	0,769	Médio	Médio
G3	54,50%	0,32	0,22	0,762	Médio	Bom
G4	34,90%	0,26	0,16	0,765	Difícil	Médio
G5	39,90%	0,32	0,23	0,762	Difícil	Bom
G6	53,50%	0,39	0,3	0,759	Médio	Bom
G7	84,90%	0,36	0,3	0,759	Fácil	Bom
G8	87,90%	0,34	0,28	0,758	Muito fácil	Bom
Q11	51,80%	0,36	0,27	0,761	Médio	Bom
Q12	56,40%	0,31	0,22	0,763	Médio	Bom
Q13	17,30%	0,41	0,34	0,757	Difícil	Muito bom
Q15	38,50%	0,35	0,26	0,761	Difícil	Bom
Q16	36,90%	0,44	0,35	0,757	Difícil	Muito bom
Q17	27,00%	0,35	0,27	0,761	Difícil	Bom
Q18	15,50%	0,29	0,22	0,764	Difícil	Médio
Q19	27,20%	0,4	0,32	0,758	Difícil	Muito bom
Q20	16,20%	0,26	0,19	0,765	Difícil	Médio
Q21	29,20%	0,43	0,35	0,758	Difícil	Muito Bom
Q23	26,80%	0,19	0,1	0,768	Difícil	Fraco
Q24	30,60%	0,3	0,21	0,764	Difícil	Bom
Q25	18,50%	0,41	0,34	0,759	Difícil	Muito bom
Q26	23,20%	0,42	0,34	0,757	Difícil	Muito bom
Q27	48,20%	0,49	0,4	0,753	Médio	Muito bom
Q28	31,80%	0,34	0,25	0,76	Difícil	Bom
Q29	30,20%	0,25	0,16	0,766	Difícil	Médio
Q30	49,60%	0,27	0,18	0,764	Médio	Médio
Q31	41,50%	0,47	0,39	0,754	Médio	Muito bom
Q32	51,50%	0,25	0,15	0,765	Médio	Médio
Q33	32,40%	0,49	0,41	0,754	Difícil	Muito bom
Q34	19,10%	0,38	0,31	0,76	Difícil	Bom
Q36	20,80%	0,39	0,32	0,76	Difícil	Bom
Q37	44,40%	0,43	0,34	0,756	Médio	Muito bom

Nota: Questão específica (Q) e Questão Geral (G).

O *Cronbach's alpha* para verificar a confiabilidade dos itens, a notação (*) significa que todas as questões são consideradas, excluindo as que não apresentam asterisco. As questões em negrito são as questões que apresentaram valores inadequados para a proporção de acerto, ponto bisserial e *Chonbach's alpha*.

APÊNDICE B – Índices de facilidade conforme a proporção de acertos, índices de discriminação de acordo com o ponto bisserial e o *Cronbach's alpha* na prova geral dos 436 respondentes com os 32 itens, excluindo as questões (G1, Q23) e as (G1, G2 e Q23)

436 respondentes						
Questões	Proporção % acertos	Ponto Bisserial incluir	Bisserial excluir	<i>Cronbach's alpha</i> *Todas as questões e Excluindo a questão	Índice de Facilidade	Índice de Discriminação
				0,746*		
G1	10,30%	0,11	0,04	0,749	Muito difícil	Fraco
G2	58,30%	0,22	0,12	0,75	Médio	Médio
G3	54,50%	0,32	0,22	0,741	Médio	Bom
G4	34,90%	0,26	0,16	0,744	Difícil	Médio
G5	39,90%	0,32	0,23	0,74	Difícil	Bom
G6	53,50%	0,39	0,3	0,737	Médio	Bom
G7	84,90%	0,36	0,3	0,74	Fácil	Bom
G8	87,90%	0,34	0,28	0,739	Muito fácil	Bom
Q11	51,80%	0,36	0,27	0,74	Médio	Bom
Q12	56,40%	0,31	0,22	0,743	Médio	Bom
Q13	17,30%	0,41	0,34	0,736	Difícil	Muito bom
Q15	38,50%	0,35	0,26	0,739	Difícil	Bom
Q16	36,90%	0,44	0,35	0,735	Difícil	Muito bom
Q17	27,00%	0,35	0,27	0,739	Difícil	Bom
Q18	15,50%	0,29	0,22	0,742	Difícil	Médio
Q19	27,20%	0,4	0,32	0,736	Difícil	Muito bom
Q20	16,20%	0,26	0,19	0,743	Difícil	Médio
Q21	29,20%	0,43	0,35	0,735	Difícil	Muito Bom
Q23	26,80%	0,19	0,1	0,747	Difícil	Fraco
Q24	30,60%	0,3	0,21	0,743	Difícil	Bom
Q25	18,50%	0,41	0,34	0,736	Difícil	Muito bom
Q26	23,20%	0,42	0,34	0,734	Difícil	Muito bom
Q27	48,20%	0,49	0,4	0,73	Médio	Muito bom
Q28	31,80%	0,34	0,25	0,739	Difícil	Bom
Q29	30,20%	0,25	0,16	0,745	Difícil	Médio
Q30	49,60%	0,27	0,18	0,744	Médio	Médio
Q31	41,50%	0,47	0,39	0,732	Médio	Muito bom
Q32	51,50%	0,25	0,15	0,744	Médio	Médio
Q33	32,40%	0,49	0,41	0,73	Difícil	Muito bom
Q34	19,10%	0,38	0,31	0,737	Difícil	Bom
Q36	20,80%	0,39	0,32	0,739	Difícil	Bom
Q37	44,40%	0,43	0,34	0,733	Médio	Muito bom

Nota: Questão específica (Q) e Questão Geral (G).

No *Cronbach's alpha* verifica-se a confiabilidade de sem o asterisco os itens com anotação (*) e significa, que todas questões são consideradas a questão é excluída.

As questões em negrito são as questões que apresentaram valores inadequadas.

436 respondentes - excluindo as questões G1 e Q23						
Questões	Proporção % acertos	Ponto Bisserial		<i>Cronbach's alpha</i>	Índice de Facilidade	Índice de Discriminação
		incluir	excluir	*Todas as questões e Excluindo a questão		
				0,763*		
G2	58,30%	0,22	0,12	0,766	Médio	Médio
G3	54,50%	0,32	0,23	0,759	Médio	Bom
G4	35,00%	0,26	0,16	0,763	Difícil	Médio
G5	39,90%	0,32	0,23	0,759	Difícil	Bom
G6	53,40%	0,39	0,3	0,754	Médio	Bom
G7	84,90%	0,36	0,29	0,757	Fácil	Bom
G8	87,80%	0,34	0,28	0,757	Muito fácil	Bom
Q11	51,80%	0,36	0,26	0,756	Médio	Bom
Q12	56,20%	0,32	0,22	0,759	Médio	Bom
Q13	17,30%	0,41	0,34	0,754	Difícil	Muito bom
Q15	38,50%	0,34	0,25	0,758	Difícil	Bom
Q16	36,70%	0,43	0,35	0,752	Difícil	Muito bom
Q17	27,00%	0,36	0,27	0,756	Difícil	Bom
Q18	15,50%	0,29	0,22	0,758	Difícil	Médio
Q19	27,20%	0,40	0,32	0,753	Difícil	Muito bom
Q20	16,20%	0,24	0,17	0,761	Difícil	Médio
Q21	29,20%	0,43	0,35	0,753	Difícil	Muito Bom
Q24	30,60%	0,31	0,22	0,758	Difícil	Fraco
Q25	18,50%	0,41	0,34	0,753	Difícil	Bom
Q26	23,20%	0,41	0,34	0,753	Difícil	Muito bom
Q27	48,20%	0,49	0,41	0,747	Difícil	Muito bom
Q28	31,80%	0,34	0,25	0,757	Médio	Muito bom
Q29	30,20%	0,25	0,16	0,762	Difícil	Bom
Q30	49,60%	0,28	0,18	0,762	Difícil	Médio
Q31	41,50%	0,48	0,39	0,75	Médio	Médio
Q32	51,50%	0,25	0,15	0,763	Médio	Muito bom
Q33	32,40%	0,5	0,42	0,749	Médio	Médio
Q34	19,10%	0,37	0,30	0,754	Difícil	Muito bom
Q36	20,80%	0,38	0,31	0,756	Difícil	Bom
Q37	44,40%	0,42	0,33	0,752	Difícil	Bom

Nota: Questão específica (Q) e Questão Geral (G).

No *Cronbach's alpha* verifica-se a confiabilidade de sem o asterisco os itens com anotação (*) e significa, que todas questões são consideradas a questão é excluída.

As questões em negrito são as questões que apresentaram valores inadequadas.

436 respondentes - excluindo as questões G1, G2, Q23						
Questões	Proporção % acertos	Ponto Bisserial		<i>Cronbach's alpha</i> *Todas as questões e Excluindo a questão	Índice de Facilidade	Índice de Discriminação
		incluir	excluir			
				0,774*		
G3	54,50%	0,32	0,22	0,772	Médio	Bom
G4	35,00%	0,28	0,18	0,774	Difícil	Médio
G5	39,90%	0,3	0,2	0,772	Difícil	Bom
G6	53,40%	0,37	0,28	0,769	Médio	Bom
G7	84,90%	0,34	0,27	0,769	Fácil	Bom
G8	87,80%	0,34	0,28	0,769	Muito fácil	Bom
Q11	51,80%	0,37	0,27	0,771	Médio	Bom
Q12	56,20%	0,32	0,22	0,773	Médio	Bom
Q13	17,30%	0,41	0,34	0,768	Difícil	Muito bom
Q15	38,50%	0,34	0,25	0,771	Difícil	Bom
Q16	36,70%	0,43	0,35	0,767	Difícil	Muito bom
Q17	27,00%	0,36	0,27	0,77	Difícil	Bom
Q18	15,50%	0,29	0,22	0,773	Difícil	Médio
Q19	27,20%	0,41	0,32	0,768	Difícil	Muito bom
Q20	16,20%	0,24	0,17	0,773	Difícil	Médio
Q21	29,20%	0,44	0,35	0,767	Difícil	Muito Bom
Q24	30,60%	0,31	0,22	0,773	Difícil	Fracó
Q25	18,50%	0,41	0,34	0,768	Difícil	Bom
Q26	23,20%	0,43	0,35	0,766	Difícil	Muito bom
Q27	48,20%	0,49	0,41	0,763	Difícil	Muito bom
Q28	31,80%	0,35	0,26	0,77	Médio	Muito bom
Q29	30,20%	0,26	0,17	0,775	Difícil	Bom
Q30	49,60%	0,29	0,19	0,773	Difícil	Médio
Q31	41,50%	0,48	0,4	0,764	Médio	Médio
Q32	51,50%	0,26	0,16	0,774	Médio	Muito bom
Q33	32,40%	0,5	0,42	0,762	Médio	Médio
Q34	19,10%	0,38	0,31	0,769	Difícil	Muito bom
Q36	20,80%	0,38	0,31	0,77	Difícil	Bom
Q37	44,40%	0,43	0,34	0,765	Difícil	Bom

Nota: Questão específica (Q) e Questão Geral (G).

No *Cronbach's alpha* verifica-se a confiabilidade de sem o asterisco os itens com anotação (*) e significa, que todas questões são consideradas a questão é excluída.

As questões em negrito são as questões que apresentaram valores inadequadas.

**APÊNDICE C – Código fonte R - estimação dos
parâmetros dos itens com métodos
para acerto ao acaso**

Estimação dos parâmetros dos itens com o acerto ao acaso estimado
pelo método MVM

```
# dados com 32 itens
dados<- read.csv2(" ")
mod1<-mirt(dados,1,itemtype="2PL")
coef(mod1)
par<-t(sapply(head(coef(mod1),32), function(x) x))
par
```

Estimação dos parâmetros dos itens com o acerto ao acaso estimado
pelo método da verossimilhança perfilhada

```
ce <- seq(0,1,l=32)
ll <- c()

for(i in 1:32){
  (mod1<-mirt(dados,1,guess=ce[i]))
  ll[i] <- mod1@logLik
  print(i)}
max(ll)
data.frame(ll,ce)

#MLD1-2PLf
(mod1.2<-mirt(dados,1,guess= ,itemtype="2PL"))
mod1.2
coef(mod1.2)
par<-t(sapply(head(coef(mod1.2),32), function(x) x))
par
```

Estimação dos parâmetros dos itens com o acerto ao acaso estimado
pelo método bayesiano

```
mpri <- mirt.model('F = 1-32
```

```
PRIOR = (1-32, a1, lnorm, -0.6550, 1.1445),
        (1-32, g, norm, -1.36, 0.53)')
modpri <- mirt(dados,mpri, itemtype = '3PL')
modpri
par<-t(sapply(head(coef(modpri),32), function(x) x))
par
# Para comparar com modelos de mais dimensão, usar o vetor c fixo:
v <- mod2values(m)
c <- v$value[v$name == 'g']
c
mod1 <- mirt(dados, 1, guess=c)
mod1
par<-t(sapply(head(coef(mod1),32), function(x) x))
par
```

**APÊNDICE D - Código fonte R - transformação dos
escores dos modelos da TRI para
escala padronizada entre 0 a 1**

```
-----
ESCALA PADRONIZADA
```

```
-----
names(dados)
newdata=rep(1,40)
newdata
names(newdata)=names(dados)
newdata
pad.1<-fscores(mod1, full.scores=TRUE
               , response.pattern = newdata)

newdata=rep(0,40)
newdata
names(newdata)=names(dados)
newdata
pad.0 <-fscores(mod1, full.scores=TRUE
               , response.pattern = newdata)
cbind(pad.0[,41],pad.1[,41])
escore <- fscores(mod1, full.scores=TRUE)
head(escore[,41])
padro <- (escore[,41] - pad.0[,41])/(pad.1[,41] - pad.0[,41])
padro
```

```
-----
ESCALA PADRONIZADA MULTIDIMENSIONAL
```

```
-----
names(dados)
newdata=rep(1,40)
newdata
names(newdata)=names(dados)
newdata
pad.1 <- fscores(mod2.2, full.scores=TRUE
               , response.pattern = newdata)

newdata=rep(0,40)
newdata
```

```

names(newdata)=names(dados)
newdata
pad.0 <- fscores(mod2.2, full.scores=TRUE
                 , response.pattern = newdata)
cbind(pad.0[,41:42],pad.1[,41:42])

escore <- fscores(mod2.2, full.scores=TRUE)
head(escore[,41:42])
pad.0[,41]
padro1 <- (escore[,41] - pad.0[,41])/(pad.1[,41] - pad.0[,41])

padro2<- (escore[,42] - pad.0[,42])/(pad.1[,42] - pad.0[,42])
escor<-escore[,41:42]
padro<-data.frame(escor,padro1,padro2)
padro

fat<-summary(mod2.2)
fat <- cbind(F=mod2.2@F, h2=mod2.2@h2)
fat
s.f1 <- sum(fat[,1]^2)
s.f1
s.f2 <-sum(fat[,2]^2)
s.f2
esc.padronizado <- (s.f1/(s.f1 + s.f2))*padro1 + (s.f2/(s.f1 + s.f2))*padro2
esc.padronizado
data.frame(esc.padronizado)
padromult<-data.frame(escor,padro1,padro2,esc.padronizado)
padromult

```

APÊNDICE E – Escore dos 94 respondentes das habilidades conforme os modelos propostos pela TRI.

Al.	H.geral	Ac.	Al	H.QGI	Ac	Al.	H.IEQ	Ac	Al	H.GAA	Ac	Al	H.CDI	Ac.	Al	H.FGE	Ac.	Al.	H.EST	A
2	-1,565750465	0	2	-0,666940599	0	2	-0,854792865	0	2	-0,451355247	0	2	-0,779934789	0	2	-0,644210844	0	2	-0,7992725	0
10	-1,565750465	0	10	-0,666940599	0	10	-0,854792865	0	10	-0,451355247	0	4	-0,779934789	0	10	-0,644210844	0	10	-0,7992725	0
11	-1,565750465	0	11	-0,666940599	0	11	-0,854792865	0	11	-0,451355247	0	8	-0,779934789	0	11	-0,644210844	0	11	-0,7992725	0
15	-1,565750465	0	15	-0,666940599	0	15	-0,854792865	0	15	-0,451355247	0	10	-0,779934789	0	15	-0,644210844	0	15	-0,7992725	0
36	-1,565750465	0	23	-0,666940599	0	36	-0,854792865	0	18	-0,451355247	0	11	-0,779934789	0	18	-0,644210844	0	36	-0,7992725	0
44	-1,565750465	0	31	-0,666940599	0	44	-0,854792865	0	20	-0,451355247	0	15	-0,779934789	0	22	-0,644210844	0	44	-0,7992725	0
45	-1,565750465	0	36	-0,666940599	0	45	-0,854792865	0	31	-0,451355247	0	24	-0,779934789	0	29	-0,644210844	0	45	-0,7992725	0
51	-1,565750465	0	37	-0,666940599	0	47	-0,854792865	0	32	-0,451355247	0	36	-0,779934789	0	33	-0,644210844	0	47	-0,7992725	0
57	-1,565750465	0	44	-0,666940599	0	51	-0,854792865	0	36	-0,451355247	0	44	-0,779934789	0	35	-0,644210844	0	51	-0,7992725	0
58	-1,565750465	0	45	-0,666940599	0	53	-0,854792865	0	38	-0,451355247	0	45	-0,779934789	0	36	-0,644210844	0	57	-0,7992725	0
60	-1,565750465	0	51	-0,666940599	0	57	-0,854792865	0	44	-0,451355247	0	51	-0,779934789	0	44	-0,644210844	0	58	-0,7992725	0
64	-1,565750465	0	52	-0,666940599	0	58	-0,854792865	0	45	-0,451355247	0	56	-0,779934789	0	45	-0,644210844	0	60	-0,7992725	0
66	-1,565750465	0	53	-0,666940599	0	60	-0,854792865	0	47	-0,451355247	0	57	-0,779934789	0	46	-0,644210844	0	61	-0,7992725	0
80	-1,565750465	0	54	-0,666940599	0	64	-0,854792865	0	48	-0,451355247	0	58	-0,779934789	0	48	-0,644210844	0	64	-0,7992725	0
81	-1,565750465	0	55	-0,666940599	0	66	-0,854792865	0	49	-0,451355247	0	60	-0,779934789	0	51	-0,644210844	0	66	-0,7992725	0
82	-1,565750465	0	57	-0,666940599	0	75	-0,854792865	0	51	-0,451355247	0	63	-0,779934789	0	52	-0,644210844	0	72	-0,7992725	0
83	-1,565750465	0	58	-0,666940599	0	80	-0,854792865	0	56	-0,451355247	0	64	-0,779934789	0	57	-0,644210844	0	76	-0,7992725	0
93	-1,565750465	0	60	-0,666940599	0	81	-0,854792865	0	57	-0,451355247	0	66	-0,779934789	0	58	-0,644210844	0	80	-0,7992725	0
94	-1,565750465	0	61	-0,666940599	0	82	-0,854792865	0	58	-0,451355247	0	69	-0,779934789	0	60	-0,644210844	0	81	-0,7992725	0
69	-0,476715395	4	64	-0,666940599	0	83	-0,854792865	0	59	-0,451355247	0	80	-0,779934789	0	64	-0,644210844	0	82	-0,7992725	0
61	-0,476715395	4	66	-0,666940599	0	93	-0,854792865	0	60	-0,451355247	0	81	-0,779934789	0	66	-0,644210844	0	83	-0,7992725	0
92	-0,309494814	5	69	-0,666940599	0	94	-0,854792865	0	61	-0,451355247	0	82	-0,779934789	0	68	-0,644210844	0	84	-0,7992725	0
52	-0,309494814	5	74	-0,666940599	0	28	-0,296340275	1	64	-0,451355247	0	83	-0,779934789	0	69	-0,644210844	0	90	-0,7992725	0
84	-0,309494814	5	76	-0,666940599	0	32	-0,296340275	1	66	-0,451355247	0	88	-0,779934789	0	80	-0,644210844	0	93	-0,7992725	0
59	-0,160206267	6	80	-0,666940599	0	46	-0,296340275	1	71	-0,451355247	0	92	-0,779934789	0	81	-0,644210844	0	94	-0,7992725	0
68	-0,160206267	6	81	-0,666940599	0	59	-0,296340275	1	78	-0,451355247	0	93	-0,779934789	0	82	-0,644210844	0	18	-0,241138912	1
47	-0,160206267	6	82	-0,666940599	0	88	-0,296340275	1	80	-0,451355247	0	94	-0,779934789	0	83	-0,644210844	0	29	-0,241138912	1
90	-0,023815981	7	83	-0,666940599	0	89	-0,296340275	1	81	-0,451355247	0	28	-0,216301476	1	84	-0,644210844	0	69	-0,241138912	1
88	-0,023815981	7	84	-0,666940599	0	38	-0,296340275	1	82	-0,451355247	0	39	-0,216301476	1	86	-0,644210844	0	77	-0,241138912	1
89	-0,023815981	7	87	-0,666940599	0	50	-0,296340275	1	83	-0,451355247	0	43	-0,216301476	1	93	-0,644210844	0	59	-0,241138912	1
74	-0,023815981	7	88	-0,666940599	0	54	-0,296340275	1	86	-0,451355247	0	59	-0,216301476	1	94	-0,644210844	0	75	-0,241138912	1
76	-0,023815981	7	92	-0,666940599	0	3	-0,296340275	1	87	-0,451355247	0	71	-0,216301476	1	47	-0,040528432	1	88	-0,241138912	1
87	-0,023815981	7	93	-0,666940599	0	34	-0,296340275	1	93	-0,451355247	0	73	-0,216301476	1	61	-0,040528432	1	92	-0,241138912	1
86	-0,023815981	7	94	-0,666940599	0	35	-0,296340275	1	94	-0,451355247	0	91	-0,216301476	1	70	-0,040528432	1	3	-0,241138912	1
71	-0,023815981	7	40	0,000554026	1	68	-0,296340275	1	54	-0,451351769	1	34	-0,216301476	1	74	-0,040528432	1	50	-0,241138912	1
20	0,102507357	8	49	0,000554026	1	71	-0,296340275	1	88	-0,451351769	1	52	-0,216301476	1	77	-0,040528432	1	74	-0,241138912	1
50	0,102507357	8	42	0,000554028	1	74	-0,296340275	1	12	-0,419588608	2	61	-0,216301476	1	85	-0,040528432	1	85	-0,241138912	1
54	0,102507357	8	86	0,000554028	1	77	-0,296340275	1	77	-0,419588608	2	65	-0,216301476	1	5	-0,040528432	1	87	-0,241138912	1
18	0,102507357	8	59	0,000554028	1	31	-0,296340275	1	1	-0,408153	1	68	-0,216301476	1	7	-0,040528432	1	89	-0,241138912	1
32	0,102507357	8	67	0,000554028	1	52	-0,296340275	1	21	-0,408153	1	75	-0,216301476	1	8	-0,040528432	1	32	-0,241138912	1
55	0,102507357	8	72	0,000554028	1	55	-0,296340275	1	52	-0,408153	1	76	-0,216301476	1	13	-0,040528432	1	38	-0,241138912	1
46	0,102507357	8	75	0,000554028	1	92	-0,296340275	1	85	-0,408153	1	84	-0,216301476	1	17	-0,040528432	1	67	-0,241138912	1
34	0,102507357	8	89	0,000554028	1	61	0,207482404	2	90	-0,408153	1	3	-0,216301476	1	20	-0,040528432	1	13	0,25360343	2
75	0,102507357	8	1	0,000554029	1	65	0,207482404	2	72	-0,408148952	2	12	-0,216301476	1	28	-0,040528432	1	49	0,25360343	2
72	0,221141984	9	4	0,000554029	1	14	0,207482404	2	55	0,053204812	1	20	-0,216301476	1	32	-0,040528432	1	62	0,25360343	2
67	0,221141984	9	17	0,000554029	1	19	0,207482404	2	65	0,053204812	1	42	-0,216301476	1	34	-0,040528432	1	63	0,25360343	2
48	0,221141984	9	20	0,000554029	1	37														

continua

Al.	H.geral	Ac.	Al	H.QGI	Ac	Al.	H.IEQ	Ac	Al	H.GAA	Ac	Al	H.CDI	Ac.	Al	H.FGE	Ac.	Al.	H.EST	A
49	0,221141984	9	21	0,000554029	1	86	0,207482404	2	53	0,182213562	2	86	-0,216301476	1	65	-0,040528432	1	20	0,25360343	2
65	0,221141984	9	22	0,000554029	1	29	0,207482404	2	29	0,233386047	2	89	-0,216301476	1	67	-0,040528432	1	42	0,25360343	2
37	0,221141984	9	29	0,000554029	1	49	0,207482404	2	75	0,233386047	2	7	0,28270222	2	76	-0,040528432	1	48	0,25360343	2
56	0,334093011	10	34	0,000554029	1	72	0,207482404	2	6	0,369163711	1	14	0,28270222	2	89	-0,040528432	1	73	0,25360343	2
85	0,334093012	10	41	0,000554029	1	85	0,207482404	2	7	0,369163711	1	54	0,28270222	2	92	-0,040528432	1	52	0,25360343	2
78	0,334093012	10	46	0,000554029	1	6	0,207482404	2	9	0,369163711	1	90	0,28270222	2	50	-0,040528432	1	41	0,25360343	2
31	0,334093013	10	78	0,000554029	1	40	0,207482404	2	13	0,369163711	1	19	0,28270222	2	71	-0,040528432	1	54	0,25360343	2
38	0,334093013	10	79	0,000554029	1	69	0,207482404	2	16	0,369163711	1	22	0,28270222	2	79	-0,040528432	1	71	0,25360343	2
35	0,334093016	10	90	0,000554029	1	73	0,207482404	2	24	0,369163711	1	55	0,28270222	2	90	-0,040528432	1	24	0,25360343	2
29	0,334093017	10	50	0,00055403	1	16	0,207482404	2	35	0,369163711	1	67	0,28270222	2	9	0,491756384	2	37	0,25360343	2
3	0,442420486	11	65	0,00055403	1	17	0,207482404	2	37	0,369163711	1	72	0,28270222	2	3	0,491756384	2	43	0,25360343	2
28	0,442420486	11	68	0,00055403	1	18	0,207482404	2	39	0,369163711	1	9	0,28270222	2	6	0,491756384	2	25	0,25360343	2
22	0,442420486	11	48	0,512767429	2	27	0,207482404	2	40	0,369163711	1	37	0,28270222	2	21	0,491756384	2	22	0,25360343	2
77	0,442420486	11	5	0,512767429	2	30	0,207482404	2	63	0,369163711	1	35	0,28270222	2	23	0,491756384	2	28	0,25360343	2
24	0,547316834	12	77	0,512767429	2	41	0,207482404	2	67	0,369163711	1	50	0,28270222	2	27	0,491756384	2	34	0,25360343	2
42	0,547316834	12	12	0,512767434	2	87	0,207482404	2	68	0,369163711	1	85	0,28270222	2	30	0,491756384	2	55	0,25360343	2
73	0,547316834	12	16	0,512767434	2	90	0,207482404	2	69	0,369163711	1	21	0,28270222	2	38	0,491756384	2	1	0,25360343	2
4	0,547316836	12	24	0,512767434	2	1	0,207482404	2	70	0,369163711	1	46	0,28270222	2	42	0,491756384	2	19	0,25360343	2
5	0,547316836	12	27	0,512767434	2	9	0,207482404	2	74	0,369163711	1	62	0,28270222	2	55	0,491756384	2	33	0,25360343	2
1	0,547316836	12	56	0,512767434	2	56	0,207482404	2	23	0,36917755	2	87	0,28270222	2	62	0,491756384	2	5	0,25360343	2
53	0,547316836	12	71	0,512767438	2	76	0,207482404	2	43	0,36917755	2	5	0,28270222	2	53	0,491756384	2	46	0,25360343	2
8	0,64969249	13	18	0,512767439	2	84	0,207482404	2	25	0,499384619	2	48	0,28270222	2	59	0,491756384	2	68	0,25360343	2
21	0,64969249	13	47	0,512767439	2	91	0,207482404	2	46	0,499384619	2	79	0,28270222	2	31	0,491756384	2	70	0,25360343	2
40	0,64969249	13	62	0,512767439	2	5	0,682588228	3	62	0,499384619	2	1	0,745909415	3	40	0,491756384	2	78	0,25360343	2
70	0,64969249	13	6	0,51276744	2	21	0,682588228	3	84	0,499384619	2	16	0,745909415	3	4	0,491756384	2	4	0,713622707	3
62	0,64969249	13	14	0,51276744	2	23	0,682588228	3	73	0,546351306	2	18	0,745909415	3	16	0,491756384	2	7	0,713622707	3
7	0,64969249	13	32	0,51276744	2	22	0,682588228	3	28	0,841066545	3	25	0,745909415	3	25	0,491756384	2	8	0,713622707	3
63	0,64969249	13	35	0,51276744	2	70	0,682588228	3	91	0,841066545	3	40	0,745909415	3	26	0,491756384	2	17	0,713622707	3
19	0,750335125	14	38	0,51276744	2	20	0,682588228	3	4	1,00849064	2	47	0,745909415	3	37	0,491756384	2	79	0,713795122	3
6	0,750335125	14	26	0,51276744	2	7	0,682588228	3	5	1,00849064	2	49	0,74590942	3	41	0,491756384	2	91	0,713795122	3
79	0,750335125	14	30	0,958363725	3	25	0,682588228	3	8	1,00849064	2	32	0,74590942	3	54	0,491756384	2	6	0,713795122	3
23	0,750335125	14	63	0,958363725	3	43	0,682588228	3	26	1,00849064	2	30	0,74590942	3	72	0,491756384	2	56	0,713795122	3
16	0,750335125	14	73	0,958363725	3	62	0,682588228	3	27	1,00849064	2	70	0,74590942	3	87	0,491756384	2	9	0,713795122	3
41	0,75033513	14	91	0,958363725	3	39	0,682588228	3	30	1,00849064	2	74	0,74590942	3	73	0,491756384	2	23	0,713795122	3
17	0,750335131	14	70	0,958363725	3	48	0,682588228	3	34	1,00849064	2	13	1,198453805	4	78	0,491756384	2	26	0,713795122	3
91	0,849952429	15	13	0,958363728	3	63	0,682588228	3	42	1,00849064	2	26	1,198453805	4	24	0,979265202	3	31	0,713795122	3
13	0,849952429	15	3	0,958363728	3	67	0,682588228	3	50	1,00849064	2	29	1,198453805	4	43	0,979265202	3	39	0,713795122	3
9	0,849952437	15	8	0,958363728	3	13	1,151831236	4	89	1,00849064	2	53	1,198453805	4	91	0,979265203	3	86	0,713795122	3
30	0,949205581	16	28	0,958363728	3	78	1,151831236	4	92	1,00849064	2	6	1,198453805	4	19	0,979265207	3	65	0,713795122	3
14	0,949205585	16	7	0,958363734	3	42	1,151831236	4	14	1,008540253	3	23	1,198453805	4	14	0,979265212	3	14	1,164913972	4
39	0,949205586	16	19	0,958363745	3	4	1,151831236	4	3	1,418666062	3	31	1,198453805	4	1	0,979265212	3	35	1,164913972	4
43	0,949205588	16	85	0,958363746	3	8	1,151831236	4	17	1,418666062	3	33	1,198453805	4	56	0,979265212	3	40	1,164913972	4
12	0,949205588	16	25	1,389430409	4	24	1,151831236	4	22	1,418666062	3	38	1,198453805	4	39	0,979265212	3	21	1,164913972	4
25	0,949205598	16	9	1,854677273	5	33	1,151831236	4	33	1,486802357	3	17	1,198453805	4	75	0,979265212	3	30	1,164913972	4
27	1,048718998	17	33	1,854677273	5	79	1,151831236	4	41	1,486802357	3	27	1,198453805	4	12	1,447645613	4	16	1,164913972	4
33	1,149201127	18	39	1,854677273	5	12	1,635405243	5	76	1,486802357	3	41	1,198453805	4	63	1,447645613	4	53	1,164913972	4
26	1,149201143	18	43	1,854677273	5	26	1,635405243	5	79	1,486802357	3	77	1,198453805	4	88	1,447645613	4	27	1,630561302	5

**APÊNDICE F – Escores e escala padronizada dos 94
e 75 respondentes do MLD1
indicados pela TRIU e TRIM**

Al	Ac	TCT	94 respondentes		75 respondentes		94 respondentes		75 respondentes	
			Esc.M2	E.P	Esc.M4	E.P	Esc.2PL	E.P	Esc.2PLf	E.P
1	12	4	0,547	0,47	0,186	0,29	0,501	0,45	-0,367	0,30
2	0	0	-1,566	0	-	-	-1,639	0	-	-
3	11	3,7	0,442	0,45	-0,005	0,25	0,519	0,45	0,525	0,50
4	12	4	0,547	0,47	0,572	0,36	0,657	0,48	0,512	0,50
5	12	4	0,547	0,47	0,516	0,35	0,515	0,45	0,292	0,45
6	14	4,7	0,750	0,52	0,809	0,41	0,831	0,51	0,736	0,55
7	13	4,3	0,650	0,49	0,645	0,38	0,888	0,53	1,333	0,68
8	13	4,3	0,650	0,49	0,716	0,39	0,79	0,51	0,745	0,55
9	15	5	0,850	0,54	0,860	0,42	1,22	0,60	1,83	0,79
10	0	0	-1,566	0	-	-	-1,639	0	-	-
11	0	0	-1,566	0	-	-	-1,639	0	-	-
12	16	5,3	0,949	0,56	0,697	0,39	0,75	0,50	-0,26	0,33
13	15	5	0,850	0,54	1,019	0,45	0,904	0,53	0,761	0,55
14	16	5,3	0,949	0,56	0,855	0,42	1,004	0,55	0,837	0,57
15	0	0	-1,566	0	-	-	-1,639	0	-	-
16	14	4,7	0,750	0,52	1,056	0,46	0,887	0,53	0,664	0,53
17	14	4,7	0,750	0,52	0,808	0,41	0,836	0,52	0,591	0,52
18	8	2,7	0,103	0,37	-0,255	0,2	0,258	0,39	-0,147	0,35
19	14	4,7	0,750	0,52	0,362	0,32	0,757	0,50	-0,22	0,33
20	8	2,7	0,103	0,37	-0,428	0,17	0,049	0,35	-0,9	0,18
21	13	4,3	0,650	0,49	0,418	0,33	0,587	0,46	-0,07	0,37
22	11	3,7	0,442	0,45	0,141	0,28	0,526	0,45	0,397	0,47
23	14	4,7	0,750	0,52	0,788	0,41	0,695	0,49	0,159	0,42
24	12	4	0,547	0,47	0,454	0,34	0,695	0,49	0,743	0,55
25	16	5,3	0,949	0,56	1,101	0,47	1,014	0,55	1,303	0,68
26	18	6	1,149	0,61	1,641	0,57	1,231	0,60	1,28	0,67
27	17	5,7	1,049	0,58	1,585	0,56	1,149	0,58	0,799	0,56
28	11	3,7	0,442	0,45	0,010	0,26	0,479	0,44	0,498	0,50
29	10	3,3	0,334	0,42	-0,054	0,24	0,223	0,39	-0,677	0,23
30	16	5,3	0,949	0,56	1,216	0,49	1,166	0,58	1,521	0,72
31	10	3,3	0,334	0,42	0,054	0,26	0,333	0,41	-0,582	0,26
32	8	2,7	0,103	0,37	-0,352	0,19	0,211	0,39	-0,671	0,24
33	18	6	1,149	0,61	1,573	0,56	1,349	0,62	1,532	0,73
34	8	2,7	0,103	0,37	-0,237	0,21	0,284	0,40	0,303	0,45
35	10	3,3	0,334	0,42	0,163	0,28	0,601	0,47	0,635	0,53
36	0	0	-1,566	0	-	-	-1,639	0	-	-
37	9	3	0,221	0,4	-0,086	0,24	0,239	0,39	0,123	0,41
38	10	3,3	0,334	0,42	-0,137	0,23	0,31	0,41	-1,01	0,16
39	16	5,3	0,949	0,56	1,124	0,47	1,331	0,62	1,893	0,80
40	13	4,3	0,650	0,49	0,625	0,37	0,739	0,49	0,532	0,50
41	14	4,7	0,750	0,52	0,871	0,42	0,825	0,51	0,577	0,51
42	12	4	0,547	0,47	0,283	0,31	0,495	0,44	-0,457	0,28
43	16	5,3	0,949	0,56	0,986	0,44	1,26	0,60	2,478	0,94
44	0	0	-1,566	0	-	-	-1,639	0	-	-
45	0	0	-1,566	0	-	-	-1,639	0	-	-
46	8	2,7	0,103	0,37	-0,385	0,18	0,186	0,38	0,068	0,40
47	6	2	-0,160	0,31	-0,791	0,1	-0,174	0,31	-0,931	0,18
48	9	3	0,221	0,4	-0,178	0,22	0,375	0,42	-0,599	0,25
49	9	3	0,221	0,4	-0,101	0,23	0,258	0,39	-0,759	0,22

continua

Al	Ac	TCT	94 respondentes		75 respondentes		94 respondentes		75 respondentes	
			Esc.M2	E.P	Esc.M4	E.P	Esc.2PL	E.P	Esc.2PLf	E.P
50	8	2,7	0,103	0,37	-0,625	0,13	-0,024	0,34	-0,697	0,23
51	0	0	-1,566	0	-	-	-1,639	0,00	-	-
52	5	1,7	-0,309	0,28	-1,165	0,03	-0,509	0,24	-1,286	0,10
53	12	4	0,547	0,47	0,156	0,28	0,25	0,39	-0,989	0,17
54	8	2,7	0,103	0,37	-0,793	0,1	-0,044	0,33	-0,882	0,19
55	8	2,7	0,103	0,37	-0,343	0,19	0,075	0,36	-0,638	0,24
56	10	3,3	0,334	0,42	-0,237	0,21	0,298	0,40	-0,305	0,32
57	0	0	-1,566	0	-	-	-1,639	0	-	-
58	0	0	-1,566	0	-	-	-1,639	0	-	-
59	6	2	-0,160	0,31	-0,860	0,09	-0,262	0,29	-0,92	0,18
60	0	0	-1,566	0	-	-	-1,639	0	-	-
61	4	1,3	-0,477	0,24	-1,312	0	-0,828	0,17	-1,592	0,03
62	13	4,3	0,650	0,49	0,600	0,37	0,744	0,50	0,556	0,51
63	13	4,3	0,650	0,49	0,538	0,36	0,818	0,51	0,697	0,54
64	0	0	-1,566	0	-	-	-1,639	0,00	-	-
65	9	3	0,221	0,4	-0,492	0,16	0,128	0,37	-0,767	0,21
66	0	0	-1,566	0	-	-	-1,639	0	-	-
67	9	3	0,221	0,4	-0,260	0,2	0,257	0,39	-0,13	0,36
68	6	2	-0,160	0,31	-0,739	0,11	-0,025	0,34	0,003	0,39
69	4	1,3	-0,477	0,24	-1,074	0,05	-0,392	0,26	-0,581	0,26
70	13	4,3	0,650	0,49	0,638	0,38	0,708	0,49	0,545	0,51
71	7	2,3	-0,024	0,34	-0,650	0,13	0,209	0,38	-0,307	0,31
72	9	3	0,221	0,4	-0,432	0,17	0,088	0,36	-0,654	0,24
73	12	4	0,547	0,47	-0,008	0,25	0,662	0,48	0,389	0,47
74	7	2,3	-0,024	0,34	-0,496	0,16	-0,029	0,34	-0,421	0,29
75	8	2,7	0,103	0,37	-1,111	0,04	-0,192	0,30	-1,725	0
76	7	2,3	-0,024	0,34	-0,690	0,12	0,018	0,35	-0,197	0,34
77	11	3,7	0,442	0,45	-0,175	0,22	0,35	0,41	-0,199	0,34
78	10	3,3	0,334	0,42	-0,174	0,22	0,335	0,41	-0,483	0,28
79	14	4,7	0,750	0,52	0,824	0,41	0,686	0,48	0,071	0,40
80	0	0	-1,566	0	-	-	-1,639	0	-	-
81	0	0	-1,566	0	-	-	-1,639	0	-	-
82	0	0	-1,566	0	-	-	-1,639	0	-	-
83	0	0	-1,566	0	-	-	-1,639	0	-	-
84	5	1,7	-0,309	0,28	-1,203	0,02	-0,427	0,25	-1,086	0,14
85	10	3,3	0,334	0,42	-0,196	0,22	0,342	0,41	-0,428	0,29
86	7	2,3	-0,024	0,34	-0,511	0,15	0,089	0,36	-0,825	0,20
87	7	2,3	-0,024	0,34	-0,511	0,15	-0,047	0,33	-0,629	0,25
88	7	2,3	-0,024	0,34	-1,037	0,05	-0,347	0,27	-1,345	0,09
89	7	2,3	-0,024	0,34	-0,676	0,12	-0,085	0,32	-0,544	0,26
90	7	2,3	-0,024	0,34	-0,740	0,11	0,014	0,34	-0,586	0,25
91	15	5	0,850	0,54	0,512	0,35	0,832	0,51	0,502	0,49
92	5	1,7	-0,309	0,28	-0,923	0,08	-0,274	0,28	-0,512	0,27
93	0	0	-1,566	0	-	-	-1,639	0	-	-
94	0	0	-1,566	0	-	-	-1,639	0	-	-

NOTA:Escore (Esc.) dos modelos e Escala Padronizada (EP).

**APÊNDICE G - Escores e escala padronizada dos
189 e 154 respondentes do conjunto
de dados II**

Al	Total Acertos	Nota TCT	189 respondentes		154 respondentes				
			Escore-MLD1	EP	Escore MLD2-2PLf		EP	EPM	
			F1-2PL		F1	F2	F1	F2	
1	27	6,8	1,34	0,64	-1,875	-1,026	0,83	0,71	0,78
2	8	2	-0,33	0,31	1,41	0,93	0,16	0,22	0,18
3	20	5	0,66	0,5	-1,344	-0,046	0,72	0,47	0,62
4	16	4	0,28	0,43	0,394	0,165	0,37	0,41	0,39
5	22	5,5	0,84	0,54	-0,85	-0,647	0,62	0,62	0,62
6	23	5,8	0,77	0,52	-0,203	-1,223	0,49	0,76	0,6
7	21	5,3	0,66	0,5	-0,564	-0,4	0,56	0,56	0,56
8	22	5,5	0,79	0,53	-1,639	-0,494	0,78	0,58	0,7
9	14	3,5	0,13	0,4	0,518	0,135	0,34	0,42	0,37
10	13	3,3	0,05	0,38	0,694	0,218	0,31	0,4	0,34
11	21	5,3	0,61	0,49	-0,078	-0,306	0,46	0,53	0,49
12	22	5,5	0,7	0,51	-1,039	0,688	0,66	0,28	0,51
13	0	-	-1,87	0	-	-	-	-	-
14	0	-	-1,87	0	-	-	-	-	-
15	17	4,3	0,35	0,44	-0,428	0,495	0,54	0,33	0,45
16	12	3	-0,16	0,34	1,528	1,12	0,14	0,17	0,15
17	22	5,5	0,63	0,5	-0,541	0,656	0,56	0,29	0,45
18	23	5,8	0,97	0,57	-1,64	-1,02	0,78	0,71	0,75
19	20	5	0,54	0,48	-0,605	0,059	0,57	0,44	0,52
20	21	5,3	0,71	0,51	-0,993	0,286	0,65	0,38	0,54
21	19	4,8	0,49	0,47	-0,203	0,191	0,49	0,41	0,46
22	23	5,8	0,79	0,53	0,12	-2,458	0,42	1,08	0,69
23	19	4,8	0,45	0,46	-0,527	1,005	0,56	0,2	0,41
24	0	-	-1,87	0	-	-	-	-	-
25	16	4	0,22	0,42	-0,246	0,596	0,5	0,3	0,42
26	0	-	-1,87	0	-	-	-	-	-
27	25	6,3	1,17	0,6	-2,511	-0,717	0,96	0,64	0,83
28	25	6,3	1,06	0,58	-1,884	-0,577	0,83	0,6	0,74
29	0	-	-1,87	0	-	-	-	-	-
30	0	-	-1,87	0	-	-	-	-	-
31	21	5,3	0,64	0,5	-0,26	-0,358	0,5	0,55	0,52
32	14	3,5	0,26	0,42	-0,059	-0,179	0,46	0,5	0,48
33	12	3	-0,12	0,35	1,399	0,821	0,16	0,25	0,2
34	19	4,8	0,55	0,48	-0,786	-0,088	0,61	0,48	0,56
35	16	4	0,11	0,39	1,082	1,048	0,23	0,19	0,21
36	23	5,8	0,91	0,55	-1,372	-0,762	0,73	0,65	0,7
37	7	1,8	-0,79	0,22	2,945	1,784	-0,15	0	-0,09
38	22	5,5	0,7	0,51	-0,796	-0,205	0,61	0,51	0,57
39	21	5,3	0,64	0,5	-0,646	0,081	0,58	0,43	0,52
40	24	6	0,97	0,57	-1,971	-0,859	0,85	0,67	0,78
41	19	4,8	0,49	0,47	-0,442	0,155	0,54	0,42	0,49
42	29	7,3	1,26	0,62	-1,209	-0,868	0,69	0,67	0,69
43	19	4,8	0,54	0,48	-0,525	0,517	0,56	0,32	0,46
44	15	3,8	-0,07	0,36	1,41	1,45	0,16	0,09	0,13
45	22	5,5	0,84	0,54	-1,538	0,024	0,76	0,45	0,64
46	0	-	-1,87	0	-	-	-	-	-
47	17	4,3	0,41	0,45	-0,321	0,247	0,51	0,39	0,46
48	19	4,8	0,52	0,48	-1,529	0,336	0,76	0,37	0,6
49	0	-	-1,87	0	-	-	-	-	-

continua

Al	Total Acertos	Nota TCT	189 respondentes		154 respondentes				
			Escore-MLD1	EP	Escore MLD2-2PLf		EP	EPM	
			F1-2PL		F1	F2	F1	F2	
50	25	6,3	0,88	0,55	-0,917	-0,431	0,64	0,56	0,61
51	0	-	-1,87	0	-	-	-	-	-
52	0	-	-1,87	0	-	-	-	-	-
53	15	3,8	0,14	0,4	0,888	0,18	0,27	0,41	0,32
54	23	5,8	0,79	0,53	-2,53	1,254	0,96	0,14	0,63
55	18	4,5	0,44	0,46	-0,286	0,078	0,51	0,43	0,48
56	19	4,8	0,56	0,48	-0,292	-0,487	0,51	0,58	0,54
57	18	4,5	0,32	0,44	-0,19	1,061	0,49	0,19	0,37
58	25	6,3	1,08	0,59	-0,914	-1,691	0,63	0,88	0,73
59	18	4,5	0,5	0,47	-0,634	-0,038	0,58	0,46	0,53
60	12	3	-0,08	0,36	0,946	0,703	0,26	0,28	0,26
61	15	3,8	0,19	0,41	-0,025	1,13	0,45	0,17	0,34
62	14	3,5	0,1	0,39	0,61	0,989	0,32	0,2	0,28
63	27	6,8	1,24	0,62	-0,976	-1,836	0,65	0,92	0,76
64	16	4	0,29	0,43	0,047	0,338	0,44	0,37	0,41
65	21	5,3	0,51	0,47	-0,508	1,367	0,55	0,11	0,37
66	7	1,8	-0,52	0,27	0,903	1,917	0,26	-0,03	0,15
67	0	-	-1,87	0	-	-	-	-	-
68	16	4	0,1	0,39	1,232	-0,384	0,2	0,55	0,34
69	0	-	-1,87	0	-	-	-	-	-
70	13	3,3	-0,1	0,35	1,284	1,143	0,19	0,16	0,18
71	0	-	-1,87	0	-	-	-	-	-
72	17	4,3	0,36	0,44	-0,276	0,682	0,5	0,28	0,42
73	13	3,3	0,01	0,37	0,227	1,278	0,4	0,13	0,29
74	14	3,5	-0,13	0,35	1,978	0,857	0,05	0,24	0,12
75	17	4,3	0,31	0,43	0,718	-0,876	0,3	0,68	0,45
76	25	6,3	1,02	0,58	-1,401	-0,186	0,73	0,5	0,64
77	14	3,5	0,12	0,4	0,679	0,452	0,31	0,34	0,32
78	12	3	-0,15	0,34	1,108	1,27	0,22	0,13	0,19
79	15	3,8	0,15	0,4	0,105	1,306	0,43	0,12	0,31
80	21	5,3	0,55	0,48	-1,11	1,135	0,67	0,17	0,47
81	19	4,8	0,52	0,48	-0,988	0,742	0,65	0,27	0,5
82	25	6,3	0,88	0,55	-1,446	-0,04	0,74	0,46	0,63
83	20	5	0,67	0,51	-0,994	-0,626	0,65	0,61	0,64
84	24	6	0,95	0,56	-1,074	-1,112	0,67	0,74	0,69
85	24	6	0,84	0,54	-0,361	-0,959	0,52	0,7	0,59
86	18	4,5	0,4	0,45	-0,957	1,021	0,64	0,2	0,46
87	16	4	0,16	0,4	-0,099	1,834	0,47	-0,01	0,28
88	9	2,3	-0,32	0,31	1,22	1,444	0,2	0,09	0,16
89	15	3,8	0,22	0,42	-0,031	-0,12	0,46	0,48	0,47
90	18	4,5	0,46	0,46	-0,23	-0,352	0,5	0,54	0,51
91	17	4,3	0,27	0,43	0,251	0,805	0,4	0,25	0,34
92	11	2,8	-0,14	0,34	1,009	1,113	0,24	0,17	0,21
93	18	4,5	0,38	0,45	-0,35	0,357	0,52	0,36	0,46
94	0	-	-1,87	0	-	-	-	-	-
95	17	4,3	0,38	0,45	-0,446	0,914	0,54	0,22	0,41
96	0	-	-1,87	0	-	-	-	-	-
97	16	4	0,29	0,43	0,419	-0,417	0,36	0,56	0,44
98	19	4,8	0,31	0,43	0,519	0,639	0,34	0,29	0,32
99	13	3,3	-0,05	0,36	0,793	1,477	0,29	0,08	0,2

continua

Al	Total Acertos	Nota TCT	189 respondentes		154 respondentes				
			Escore-MLD1	EP	Escore MLD2-2PLf		EP	EPM	
			F1-2PL		F1	F2	F1	F2	
100	18	4,5	0,18	0,41	1,117	0,597	0,22	0,3	0,25
101	0	-	-1,87	0	-	-	-	-	-
102	16	4	0,29	0,43	-0,429	0,978	0,54	0,21	0,4
103	17	4,3	0,33	0,44	0,462	0,173	0,35	0,41	0,38
104	17	4,3	0,17	0,41	0,647	1,251	0,32	0,14	0,24
105	10	2,5	-0,33	0,31	0,889	1,956	0,27	-0,04	0,14
106	0	-	-1,87	0	-	-	-	-	-
107	0	-	-1,87	0	-	-	-	-	-
108	0	-	-1,87	0	-	-	-	-	-
109	0	-	-1,87	0	-	-	-	-	-
110	0	-	-1,87	0	-	-	-	-	-
111	18	4,5	0,2	0,41	0,912	0,885	0,26	0,23	0,25
112	0	-	-1,87	0	-	-	-	-	-
113	0	-	-1,87	0	-	-	-	-	-
114	0	-	-1,87	0	-	-	-	-	-
115	17	4,3	0,34	0,44	0,028	0,156	0,44	0,41	0,43
116	0	-	-1,87	0	-	-	-	-	-
117	0	-	-1,87	0	-	-	-	-	-
118	15	3,8	0,1	0,39	1,193	0,126	0,21	0,42	0,29
119	17	4,3	0,33	0,44	0,197	-0,223	0,41	0,51	0,45
120	15	3,8	0,28	0,43	0,716	-0,995	0,3	0,71	0,47
121	16	4	0,3	0,43	-0,246	0,842	0,5	0,24	0,4
122	20	5	0,64	0,5	-0,308	-0,776	0,51	0,65	0,57
123	22	5,5	0,7	0,51	0,594	-2,226	0,33	1,02	0,61
124	21	5,3	0,66	0,5	-0,692	0,376	0,59	0,36	0,5
125	9	2,3	-0,4	0,29	1,63	1,27	0,12	0,13	0,12
126	20	5	0,58	0,49	-0,152	0,224	0,48	0,4	0,45
127	15	3,8	0,29	0,43	-0,599	1,139	0,57	0,17	0,41
128	21	5,3	0,59	0,49	-0,177	-0,26	0,48	0,52	0,5
129	17	4,3	0,27	0,43	0,553	0,433	0,34	0,34	0,34
130	21	5,3	0,64	0,5	-0,273	-0,672	0,5	0,62	0,55
131	0	-	-1,87	0	-	-	-	-	-
132	18	4,5	0,43	0,46	1,252	-2,239	0,19	1,02	0,53
133	16	4	0,32	0,44	0,422	-0,59	0,36	0,6	0,46
134	0	-	-1,87	0	-	-	-	-	-
135	17	4,3	0,28	0,43	-0,015	1,181	0,45	0,16	0,33
136	16	4	0,24	0,42	1,547	-1,847	0,13	0,92	0,45
137	16	4	0,41	0,45	0,297	-0,632	0,39	0,61	0,48
138	21	5,3	0,59	0,49	0,938	-1,725	0,26	0,89	0,51
139	17	4,3	0,36	0,44	1,161	-1,74	0,21	0,9	0,49
140	17	4,3	0,38	0,45	0,999	-1,917	0,25	0,94	0,52
141	17	4,3	0,37	0,45	0,8	-0,839	0,29	0,67	0,44
142	27	6,8	1,37	0,64	-1,647	-1,348	0,78	0,8	0,79
143	0	-	-1,87	0	-	-	-	-	-
144	16	4	0,16	0,4	1,046	-0,192	0,24	0,5	0,34
145	18	4,5	0,39	0,45	0,352	-0,247	0,38	0,52	0,43
146	10	2,5	-0,18	0,34	0,532	1,642	0,34	0,04	0,22
147	15	3,8	0,22	0,42	0,291	0,202	0,39	0,4	0,4
148	15	3,8	0,23	0,42	0,495	0,562	0,35	0,31	0,33
149	19	4,8	0,48	0,47	1,212	-2,313	0,2	1,04	0,54

continua

Al	Total Acertos	Nota TCT	189 respondentes		154 respondentes				
			Escore-MLD1 F1-2PL	EP	Escore MLD2-2PLf F1	EP F2	EPM F1 F2		
150	0	-	-1,87	0	-	-	-	-	-
151	18	4,5	0,36	0,44	0,833	-0,398	0,28	0,56	0,39
152	21	5,3	0,62	0,5	-0,281	-0,497	0,51	0,58	0,54
153	14	3,5	0,15	0,4	0,153	0,779	0,42	0,26	0,35
154	21	5,3	0,56	0,48	1,18	-2,126	0,21	0,99	0,52
155	16	4	0,33	0,44	0,801	-0,978	0,29	0,7	0,45
156	25	6,3	1,07	0,59	-1,59	-0,409	0,77	0,56	0,69
157	16	4	0,24	0,42	1,654	-1,821	0,11	0,92	0,44
158	24	6	0,97	0,57	0,134	-2,823	0,42	1,17	0,72
159	17	4,3	0,38	0,45	0,193	0,211	0,41	0,4	0,41
160	0	-	-1,87	0	-	-	-	-	-
161	20	5	0,59	0,49	-0,782	-0,312	0,61	0,53	0,58
162	0	-	-1,87	0	-	-	-	-	-
163	27	6,8	1,3	0,63	-1,467	-2,555	0,75	1,1	0,89
164	19	4,8	0,55	0,48	-0,62	0,505	0,57	0,33	0,47
165	19	4,8	0,45	0,46	0,993	-0,631	0,25	0,61	0,39
166	0	-	-1,87	0	-	-	-	-	-
167	16	4	0,29	0,43	0,686	-0,597	0,31	0,61	0,43
168	16	4	0,26	0,42	0,312	-0,053	0,39	0,47	0,42
169	0	-	-1,87	0	-	-	-	-	-
170	18	4,5	0,48	0,47	0,222	-0,525	0,4	0,59	0,48
171	0	-	-1,87	0	-	-	-	-	-
172	18	4,5	0,42	0,46	-0,014	-0,261	0,45	0,52	0,48
173	20	5	0,54	0,48	0,162	-1,255	0,42	0,77	0,56
174	13	3,3	-0,04	0,36	1,675	0,294	0,11	0,38	0,22
175	24	6	0,88	0,55	-0,919	-0,186	0,64	0,5	0,58
176	17	4,3	0,26	0,42	0,535	0,004	0,34	0,45	0,39
177	10	2,5	-0,24	0,32	0,809	1,426	0,28	0,09	0,21
178	18	4,5	0,49	0,47	0,049	-0,625	0,44	0,61	0,51
179	17	4,3	0,27	0,43	0,44	-0,186	0,36	0,5	0,42
180	17	4,3	0,29	0,43	0,813	-0,177	0,28	0,5	0,37
181	15	3,8	0,22	0,42	0,856	-0,346	0,27	0,54	0,38
182	19	4,8	0,46	0,46	-0,146	0,293	0,48	0,38	0,44
183	20	5	0,48	0,47	-0,799	0,816	0,61	0,25	0,46
184	19	4,8	0,5	0,47	-0,204	-0,235	0,49	0,51	0,5
185	17	4,3	0,34	0,44	0,765	-0,239	0,29	0,51	0,38
186	27	6,8	1,34	0,64	-2,071	-0,981	0,87	0,7	0,8
187	22	5,5	0,59	0,49	0,142	0,056	0,42	0,44	0,43
188	20	5	0,51	0,47	-0,463	0,772	0,54	0,26	0,43
189	22	5,5	0,82	0,54	-0,634	-0,763	0,58	0,65	0,61

**APÊNDICE H - Estimativas dos parâmetros dos
modelos do conjunto de dados III -
excluindo G1;Q23 e G1;G2;Q23**

436 respondentes - excluindo as questões G1 e Q23

Q	MLDI						MLD2														
	2PL		3PL		2PL		2PLf		3PL												
	α_{i1}	d_i	c_i	α_{i1}	d_i	c_i	α_{i1}	α_{i2}	d_i	c_i	α_{i1}	α_{i2}	d_i	c_i							
G2	0,185	0,338	0	0,211	0,142	0,103	0,203	0,237	0,055	-0,021	-0,291	0,34	0	-0,168	-0,502	0,144	0,103	-0,175	-0,62	-0,521	0,325
G3	0,591	0,196	0	0,666	-0,035	0,103	0,61	0,187	0,001	0,106	-0,733	0,199	0	0,168	-0,739	-0,036	0,103	-0,005	-0,745	0,195	0
G4	0,407	-0,646	0	0,501	-1,03	0,103	1,119	-2,334	0,254	0,493	-0,011	-0,656	0	0,622	0,023	-1,049	0,103	0,458	-0,036	-0,66	0,001
G5	0,578	-0,44	0	0,709	-0,792	0,103	0,895	-1,157	0,175	0,297	-0,468	-0,444	0	0,282	-0,697	-0,818	0,103	0,339	-1,056	-1,54	0,218
G6	0,759	0,102	0	0,869	-0,09	0,103	0,822	0,062	0,04	0,241	-0,837	0,105	0	0,244	-1,112	-0,107	0,103	0,069	-1,38	-0,419	0,189
G7	1,193	2,137	0	1,404	2,133	0,103	1,324	2,24	0,001	-0,337	-2,599	3,183	0	-0,544	-4,588	4,574	0,103	-1,085	-3,799	3,67	0,133
G8	1,427	2,594	0	1,525	2,552	0,103	1,587	2,764	0	0,115	-1,923	3,061	0	0,175	-1,944	2,933	0,103	0,039	-1,896	3,03	0
Q11	0,653	0,084	0	0,733	-0,169	0,103	0,682	0,06	0,007	0,425	-0,392	0,083	0	0,584	-0,277	-0,167	0,103	0,626	-0,237	0,077	0
Q12	0,565	0,278	0	0,662	0,064	0,103	0,658	0,275	0	-0,056	-0,946	0,296	0	0,083	-0,884	0,061	0,103	-0,192	-0,954	0,289	0,001
Q13	0,967	-1,839	0	2,474	-4,093	0,103	2,413	-3,787	0,079	1,612	0,417	-2,191	0	3,956	1,011	-5,566	0,103	2,121	0,464	-2,66	0,018
Q15	0,588	-0,502	0	0,804	-0,909	0,103	2,34	-3,194	0,286	0,649	-0,082	-0,51	0	0,751	-0,198	-0,907	0,103	1,672	-1,839	-3,866	0,293
Q16	0,861	-0,622	0	1,19	-1,128	0,103	1,584	-1,708	0,174	0,665	-0,414	-0,621	0	0,847	-0,555	-1,105	0,103	0,949	-0,736	-1,538	0,162
Q17	0,692	-1,099	0	1,102	-1,866	0,103	2,781	-4,405	0,197	0,767	-0,087	-1,124	0	0,999	-0,268	-1,85	0,103	2,022	-1,544	-4,748	0,2
Q18	0,63	-1,829	0	5,009	-8,78	0,103	5,77	-9,799	0,112	0,976	0,299	-1,976	0	10,647	0,804	-17,548	0,103	15,588	1,976	-24,193	0,108
Q19	0,853	-1,126	0	1,784	-2,284	0,103	2,537	-3,31	0,151	0,846	-0,177	-1,139	0	1,648	-0,205	-2,211	0,103	2,185	-0,582	-3,281	0,152
Q20	0,492	-1,722	0	2,175	-4,433	0,103	2,563	-5,052	0,118	0,769	0,201	-1,818	0	1,716	-0,757	-4,317	0,103	1,399	-1,019	-4,298	0,113
Q21	0,963	-1,039	0	1,513	-1,832	0,103	1,438	-1,685	0,085	0,801	-0,372	-1,037	0	1,472	-0,232	-1,853	0,103	1,494	-0,335	-1,943	0,105
Q24	0,482	-0,863	0	0,68	-1,348	0,103	0,531	-0,891	0,005	0,295	-0,31	-0,861	0	0,512	-0,255	-1,337	0,103	1,523	-0,502	-3,441	0,241
Q25	1,017	-1,77	0	2,159	-3,614	0,103	1,822	-2,953	0,073	0,865	-0,375	-1,771	0	1,779	-1,58	-4,344	0,103	1,115	-0,858	-2,71	0,063
Q26	1,073	-1,453	0	1,73	-2,555	0,103	1,356	-1,887	0,047	0,81	-0,56	-1,459	0	1,289	-0,923	-2,615	0,103	0,88	-0,925	-2,059	0,058
Q27	1,263	-0,074	0	1,509	-0,451	0,103	1,375	-0,232	0,035	0,774	-0,841	-0,083	0	1,065	-0,83	-0,443	0,103	0,824	-0,798	-0,116	0
Q28	0,699	-0,844	0	1,011	-1,408	0,103	1,61	-2,32	0,184	0,667	-0,168	-0,846	0	1,051	-0,025	-1,428	0,103	4,233	0,349	-5,807	0,243
Q29	0,364	-0,861	0	0,618	-1,372	0,103	1,645	-3,214	0,234	0,467	0,042	-0,874	0	0,827	0,187	-1,417	0,103	2,01	0,397	-3,268	0,222
Q30	0,463	-0,012	0	0,509	-0,263	0,103	0,711	-0,845	0,261	0,439	-0,129	-0,011	0	0,53	-0,061	-0,263	0,103	0,403	-0,152	-0,018	0,001
Q31	1,193	-0,419	0	1,37	-0,828	0,103	1,245	-0,469	0	0,707	-0,857	-0,432	0	1,006	-0,663	-0,817	0,103	0,734	-0,767	-0,46	0
Q32	0,428	0,066	0	0,482	-0,171	0,103	0,432	0,056	0,003	0,226	-0,326	0,065	0	0,3	-0,32	-0,171	0,103	5,544	-11,001	-21,05	0,48
Q33	1,193	-0,919	0	1,908	-1,806	0,103	2,463	-2,487	0,146	1,005	-0,475	-0,92	0	1,521	-0,727	-1,788	0,103	1,954	-1,36	-2,978	0,17
Q34	0,844	-1,646	0	2,636	-4,271	0,103	3,422	-5,353	0,12	1,285	0,276	-1,837	0	2,725	0,333	-4,072	0,103	1,868	0,234	-2,675	0,057
Q36	0,738	-1,476	0	1,836	-3,075	0,103	2,117	-3,503	0,121	0,947	0,058	-1,552	0	1,728	-0,099	-2,938	0,103	1,89	-0,105	-3,153	0,107
Q37	0,972	-0,247	0	1,348	-0,681	0,103	2,361	-1,904	0,249	0,825	-0,396	-0,247	0	1,328	-0,287	-0,679	0,103	1,854	-1,261	-2,109	0,257

continua

436 respondentes - excluindo as questões G1, G2, Q23

Q	MLD1						MLD2														
	2PL		2PLf		3PL		2PL		2PLf		3PL										
	α_i	d_i	c_i																		
G3	0,59	0,2	0	0,67	-0,05	0,11	0,61	0,19	0	0,09	0,74	0,2	0	0,14	0,79	-0,05	0,11	0,25	0,64	0,2	0
G4	0,41	-0,65	0	0,52	-1,05	0,11	1,09	-2,26	0,25	0,48	0,04	-0,66	0	0,59	0,03	-1,06	0,11	0,44	0,04	-0,67	0,01
G5	0,57	-0,44	0	0,71	-0,81	0,11	0,95	-1,27	0,19	0,31	0,44	-0,44	0	0,35	0,61	-0,82	0,11	1,17	0,86	-2,21	0,27
G6	0,75	0,16	0	0,87	-0,1	0,11	0,78	0,15	0	0,24	0,81	0,16	0	0,29	0,99	-0,11	0,11	0,51	0,66	0,03	0,06
G7	1,17	2,12	0	1,39	2,12	0,11	1,3	2,23	0	-0,35	2,41	3,03	0	-0,48	3,89	3,97	0,11	0,29	2,43	3,15	0
G8	1,41	2,58	0	1,52	2,54	0,11	1,57	2,76	0	0,07	1,9	3,03	0	0,11	1,98	2,94	0,11	0,73	1,62	3,04	0
Q11	0,65	0,08	0	0,74	-0,18	0,11	0,68	0,06	0,01	0,41	0,41	0,08	0	0,59	0,29	-0,18	0,11	0,75	0,11	0,08	0
Q12	0,57	0,28	0	0,67	0,05	0,11	0,66	0,28	0	-0,1	1	0,3	0	0	1,04	0,05	0,11	-1,66	10,33	-4,16	0,33
Q13	0,96	-1,84	0	2,57	-4,24	0,11	2,43	-3,81	0,08	1,68	-0,44	-2,24	0	4,61	-1,37	-6,34	0,11	1,91	-0,78	-2,44	0
Q15	0,59	-0,5	0	0,81	-0,93	0,11	2,37	-3,23	0,29	0,65	0,09	-0,51	0	0,77	0,2	-0,93	0,11	2,37	1,08	-3,7	0,29
Q16	0,86	-0,62	0	1,21	-1,16	0,11	1,58	-1,7	0,17	0,66	0,42	-0,62	0	0,9	0,5	-1,12	0,11	1,43	0,42	-1,79	0,19
Q17	0,7	-1,1	0	1,14	-1,92	0,11	2,76	-4,36	0,2	0,76	0,12	-1,12	0	1,03	0,27	-1,9	0,11	2,5	0,55	-4,25	0,19
Q18	0,63	-1,83	0	5,38	-9,4	0,11	5,24	-8,9	0,11	0,99	-0,29	-1,98	0	8,99	-0,45	-15,1	0,11	5,91	-1,41	-9,58	0,1
Q19	0,86	-1,13	0	1,85	-2,37	0,11	2,54	-3,3	0,15	0,83	0,21	-1,14	0	1,71	0,21	-2,29	0,11	2,64	0,08	-3,53	0,16
Q20	0,49	-1,72	0	2,28	-4,63	0,11	2,57	-5,06	0,12	0,78	-0,2	-1,82	0	1,81	0,79	-4,5	0,11	2,03	1,12	-5,02	0,12
Q21	0,97	-1,04	0	1,55	-1,87	0,11	1,37	-1,58	0,07	0,78	0,42	-1,04	0	1,49	0,27	-1,9	0,11	1,82	-0,03	-2,13	0,12
Q24	0,48	-0,86	0	0,69	-1,37	0,11	0,52	-0,87	0	0,29	0,31	-0,86	0	0,55	0,22	-1,36	0,11	2,21	-0,08	-4,1	0,25
Q25	1,02	-1,77	0	2,21	-3,71	0,11	1,81	-2,92	0,07	0,85	0,41	-1,77	0	1,83	1,54	-4,38	0,11	1,9	0,95	-3,63	0,09
Q26	1,09	-1,46	0	1,77	-2,61	0,11	1,33	-1,82	0,04	0,77	0,65	-1,47	0	1,22	1,24	-2,79	0,11	1,24	0,94	-2,21	0,06
Q27	1,25	-0,08	0	1,51	-0,47	0,11	1,36	-0,22	0,03	0,76	0,85	-0,08	0	1,08	0,8	-0,46	0,11	1,18	0,45	-0,11	0
Q28	0,7	-0,85	0	1,03	-1,44	0,11	1,61	-2,3	0,18	0,65	0,2	-0,85	0	1,01	0,09	-1,44	0,11	9,48	-2,85	-13,5	0,26
Q29	0,37	-0,86	0	0,64	-1,4	0,11	1,61	-3,15	0,23	0,46	-0,02	-0,87	0	0,86	-0,2	-1,45	0,11	2,15	-0,83	-3,58	0,23
Q30	0,47	-0,01	0	0,52	-0,27	0,11	0,7	-0,79	0,25	0,42	0,16	-0,01	0	0,49	0,12	-0,27	0,11	0,39	0,18	-0,02	0
Q31	1,2	-0,42	0	1,38	-0,85	0,11	1,25	-0,47	0	0,67	0,92	-0,44	0	0,96	0,78	-0,84	0,11	0,99	0,65	-0,47	0
Q32	0,44	0,07	0	0,49	-0,18	0,11	0,44	0,06	0	0,2	0,36	0,07	0	0,27	0,38	-0,18	0,11	11,21	7,76	-22	0,48
Q33	1,19	-0,92	0	1,96	-1,86	0,11	2,5	-2,51	0,15	0,99	0,51	-0,92	0	1,53	0,81	-1,85	0,11	2,56	0,96	-3,09	0,17
Q34	0,85	-1,65	0	2,78	-4,48	0,11	3,43	-5,34	0,12	1,27	-0,23	-1,83	0	2,74	-0,26	-4,17	0,11	1,72	-0,29	-2,66	0,06
Q36	0,74	-1,48	0	1,91	-3,19	0,11	2,11	-3,48	0,12	0,95	-0,05	-1,56	0	1,78	0,07	-3	0,11	2,14	-0,34	-3,4	0,11
Q37	0,98	-0,25	0	1,37	-0,7	0,11	2,34	-1,87	0,25	0,8	0,44	-0,25	0	1,28	0,37	-0,7	0,11	2,46	1,03	-2,32	0,26

**APÊNDICE I – Cargas fatoriais dos modelos para as
32 questões e retirando a G1 e Q23**

Q	436 respondentes e 32 questões						436 respondentes e excluindo G1 e Q23						436 respondentes e excluindo G1, G2 e Q23							
	MLD1		MLD2		3PL		MLD1		MLD2		3PL		MLD1		MLD2		3PL			
	2PLf	3PL	2PLf	3PL	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2		
G1	0,23	0,05	0,09	0,01	0,37	-0,67	0,59	0,07	-0,16	0,03	-	-	-	-	-	-	-	-	-	
G2	0,12	0,01	0,12	0,01	0,08	-0,22	0,06	0,1	-0,18	0,04	0,12	0,02	0,12	0,01	0,08	-0,26	0,07	0,16	-0,26	0,09
G3	0,36	0,13	0,34	0,11	0,32	-0,28	0,18	0,29	-0,29	0,17	0,32	0,13	0,34	0,11	0,32	-0,28	0,18	0,29	-0,27	0,16
G4	0,28	0,08	0,53	0,28	0,31	0,12	0,11	0,25	0,08	0,07	0,29	0,08	0,55	0,3	0,32	0,12	0,12	0,25	0,09	0,07
G5	0,38	0,15	0,48	0,23	0,37	-0,2	0,18	0,54	-0,29	0,37	0,39	0,15	0,47	0,22	0,36	-0,24	0,19	0,52	-0,27	0,35
G6	0,45	0,2	0,44	0,19	0,44	-0,37	0,33	0,48	-0,42	0,41	0,46	0,21	0,43	0,19	0,43	-0,38	0,34	0,49	-0,41	0,4
G7	0,63	0,4	0,61	0,37	0,47	-0,78	0,84	0,41	-0,76	0,75	0,64	0,41	0,61	0,38	0,48	-0,81	0,88	0,46	-0,76	0,8
G8	0,66	0,44	0,68	0,46	0,5	-0,57	0,58	0,54	-0,51	0,55	0,67	0,45	0,68	0,47	0,51	-0,56	0,58	0,56	-0,49	0,56
Q11	0,39	0,15	0,38	0,14	0,38	-0,04	0,15	0,4	0,02	0,16	0,4	0,16	0,37	0,14	0,39	-0,02	0,15	0,4	0,05	0,16
Q12	0,35	0,13	0,36	0,13	0,31	-0,38	0,24	0,26	-0,39	0,22	0,36	0,13	0,36	0,13	0,32	-0,35	0,22	0,27	-0,38	0,22
Q13	0,81	0,66	0,82	0,68	0,73	0,57	0,86	0,63	0,43	0,58	0,83	0,69	0,82	0,67	0,76	0,51	0,84	0,61	0,45	0,57
Q15	0,42	0,18	0,8	0,63	0,43	0,07	0,19	0,85	-0,22	0,77	0,43	0,19	0,81	0,65	0,44	0,05	0,19	0,85	-0,16	0,75
Q16	0,57	0,32	0,68	0,46	0,55	-0,08	0,31	0,63	-0,08	0,41	0,58	0,34	0,68	0,46	0,55	-0,08	0,31	0,64	-0,05	0,41
Q17	0,52	0,27	0,86	0,73	0,53	0,05	0,29	0,91	-0,13	0,84	0,56	0,31	0,85	0,73	0,54	0,06	0,3	0,86	-0,05	0,75
Q18	0,94	0,88	0,97	0,94	0,89	0,41	0,96	0,86	0,49	0,99	0,95	0,9	0,96	0,92	0,9	0,39	0,96	0,85	0,5	0,98
Q19	0,7	0,49	0,83	0,69	0,67	0,13	0,47	0,82	0,13	0,68	0,73	0,54	0,83	0,69	0,69	0,16	0,51	0,81	0,17	0,68
Q20	0,76	0,57	0,81	0,66	0,68	0,16	0,48	0,71	0,01	0,5	0,8	0,64	0,83	0,69	0,77	-0,01	0,6	0,77	-0,06	0,6
Q21	0,65	0,43	0,66	0,44	0,65	0,08	0,44	0,71	0,12	0,52	0,67	0,45	0,65	0,42	0,66	0,13	0,46	0,68	0,16	0,49
Q23	0,23	0,05	0,92	0,85	0,28	0,22	0,13	0,93	0,19	0,9	-	-	-	-	-	-	-	-	-	-
Q24	0,36	0,13	0,37	0,14	0,34	-0,06	0,12	0,8	0,08	0,65	0,38	0,14	0,3	0,09	0,35	-0,02	0,12	0,72	0,12	0,53
Q25	0,76	0,58	0,7	0,49	0,8	-0,19	0,68	0,54	0,01	0,29	0,79	0,63	0,73	0,53	0,82	-0,21	0,72	0,71	-0,07	0,51
Q26	0,7	0,49	0,63	0,39	0,7	-0,1	0,5	0,64	-0,14	0,43	0,72	0,52	0,62	0,39	0,72	-0,12	0,53	0,66	-0,12	0,45
Q27	0,65	0,42	0,62	0,38	0,65	-0,15	0,44	0,61	-0,1	0,39	0,67	0,44	0,63	0,39	0,66	-0,14	0,45	0,62	-0,09	0,39
Q28	0,5	0,25	0,74	0,54	0,49	0,16	0,26	0,84	0,41	0,87	0,52	0,27	0,69	0,47	0,5	0,16	0,28	0,82	0,44	0,86
Q29	0,33	0,11	0,72	0,52	0,34	0,19	0,15	0,64	0,42	0,59	0,35	0,12	0,7	0,48	0,36	0,22	0,18	0,61	0,43	0,55
Q30	0,28	0,08	0,36	0,13	0,29	0,08	0,09	0,28	0,04	0,08	0,29	0,08	0,38	0,15	0,3	0,07	0,1	0,27	0,03	0,07
Q31	0,62	0,39	0,59	0,35	0,62	-0,09	0,39	0,59	-0,13	0,36	0,63	0,4	0,59	0,35	0,62	-0,09	0,39	0,58	-0,11	0,35
Q32	0,27	0,07	0,24	0,06	0,26	-0,09	0,08	0,91	-0,4	0,98	0,27	0,07	0,25	0,06	0,27	-0,09	0,08	0,92	-0,35	0,98
Q33	0,73	0,54	0,82	0,67	0,72	-0,03	0,53	0,85	-0,06	0,73	0,75	0,57	0,82	0,68	0,74	-0,03	0,55	0,85	-0,04	0,73
Q34	0,82	0,67	0,9	0,81	0,72	0,41	0,68	0,64	0,36	0,54	0,85	0,72	0,9	0,8	0,75	0,37	0,71	0,62	0,36	0,51
Q36	0,72	0,51	0,79	0,63	0,67	0,23	0,49	0,71	0,26	0,57	0,74	0,55	0,78	0,61	0,69	0,19	0,51	0,7	0,27	0,56
Q37	0,62	0,38	0,81	0,66	0,64	0,11	0,42	0,83	-0,05	0,7	0,63	0,39	0,81	0,66	0,64	0,09	0,42	0,84	-0,04	0,71

**APÊNDICE J - Escores e escala padronizada do
Enade no MLD1-3PL para 436
graduados de estatística - 2009**

Al	T. A	TCT	MLD1-32 itens		MLD1-G1,G2 e Q23		Al	T. A	TCT	MLD1-32 itens		MLD1-G1,G2 e Q23	
			3PL	Escala P.	3PL	Escala P.				3PL	Escala P.	3PL	Escala P.
1	22	0,69	1,45	0,71	1,38	0,61	46	20	0,63	1,31	0,68	1,34	0,6
2	16	0,5	0,4	0,5	0,3	0,42	47	15	0,47	0,91	0,6	0,93	0,53
3	12	0,38	0,24	0,47	0,21	0,41	48	8	0,25	-0,39	0,35	-0,35	0,31
4	18	0,56	1	0,62	0,96	0,54	49	16	0,5	1,03	0,63	1,04	0,55
5	10	0,31	-0,57	0,32	-0,58	0,27	50	7	0,22	-0,78	0,28	-0,78	0,24
6	19	0,59	1,77	0,77	1,67	0,66	51	12	0,38	0,38	0,5	0,38	0,44
7	13	0,41	0,31	0,49	0,35	0,43	52	12	0,38	0,03	0,43	0,03	0,38
8	21	0,66	1,52	0,72	1,6	0,65	53	14	0,44	0,42	0,51	0,39	0,44
9	12	0,38	0,34	0,49	0,34	0,43	54	22	0,69	1,34	0,69	1,38	0,61
10	24	0,75	2,03	0,82	1,96	0,71	55	6	0,19	-2,06	0,03	-2,04	0,02
11	18	0,56	1,53	0,72	1,37	0,61	56	13	0,41	0,35	0,49	0,33	0,43
12	10	0,31	0,18	0,46	0,1	0,39	57	7	0,22	-0,82	0,27	-0,76	0,24
13	18	0,56	1,14	0,65	1,17	0,57	58	6	0,19	-1,69	0,1	-1,71	0,08
14	13	0,41	0,4	0,5	0,39	0,44	59	12	0,38	0,29	0,48	0,28	0,42
15	22	0,69	1,58	0,73	1,64	0,65	60	15	0,47	0,48	0,52	0,47	0,45
16	14	0,44	0,66	0,56	0,69	0,49	61	9	0,28	-0,25	0,38	-0,21	0,34
17	17	0,53	0,74	0,57	0,67	0,49	62	11	0,34	0,3	0,49	0,29	0,42
18	15	0,47	0,26	0,48	0,26	0,42	63	10	0,31	-0,03	0,42	0,01	0,37
19	24	0,75	1,58	0,73	1,64	0,65	64	10	0,31	-0,22	0,38	-0,25	0,33
20	17	0,53	1,11	0,64	1,16	0,57	65	9	0,28	-1,22	0,19	-1,27	0,15
21	19	0,59	1,26	0,67	1,15	0,57	66	7	0,22	-1,81	0,07	-1,76	0,07
22	28	0,88	2,15	0,84	2,08	0,73	67	11	0,34	-0,06	0,41	-0,04	0,36
23	10	0,31	-0,3	0,37	-0,32	0,32	68	16	0,5	0,81	0,58	0,86	0,52
24	5	0,16	-1,83	0,07	-1,79	0,06	69	8	0,25	-0,67	0,3	-0,56	0,28
25	24	0,75	1,78	0,77	1,9	0,7	70	13	0,41	0,47	0,52	0,46	0,45
26	8	0,25	-1,85	0,07	-1,95	0,04	71	18	0,56	0,92	0,61	0,88	0,52
27	14	0,44	0,65	0,55	0,63	0,48	72	3	0,09	-2,18	0	-2,15	0
28	18	0,56	1,16	0,65	1,17	0,57	73	11	0,34	0,01	0,43	0	0,37
29	13	0,41	-0,08	0,41	-0,09	0,36	74	6	0,19	-1,46	0,14	-1,4	0,13
30	9	0,28	-0,31	0,37	-0,27	0,33	75	15	0,47	0,74	0,57	0,81	0,51
31	7	0,22	-1,17	0,2	-1,14	0,18	76	12	0,38	0,17	0,46	0,15	0,4
32	12	0,38	-0,54	0,32	-0,57	0,27	77	13	0,41	0,28	0,48	0,3	0,42
33	6	0,19	-1,33	0,17	-1,32	0,15	78	7	0,22	-0,2	0,39	-0,18	0,34
34	11	0,34	-0,37	0,35	-0,4	0,3	79	12	0,38	0,3	0,48	0,29	0,42
35	9	0,28	-0,07	0,41	-0,09	0,36	80	15	0,47	0,48	0,52	0,46	0,45
36	6	0,19	-0,95	0,24	-0,97	0,2	81	16	0,5	0,68	0,56	0,68	0,49
37	12	0,38	-0,43	0,34	-0,31	0,32	82	15	0,47	0,76	0,57	0,79	0,51
38	12	0,38	0,15	0,46	0,15	0,4	83	19	0,59	1,6	0,74	1,52	0,63
39	9	0,28	-0,26	0,38	-0,24	0,33	84	7	0,22	-0,33	0,36	-0,34	0,31
40	15	0,47	1,07	0,63	1,08	0,56	85	9	0,28	-0,28	0,37	-0,28	0,32
41	10	0,31	0,02	0,43	0,05	0,38	86	13	0,41	0,21	0,47	0,2	0,41
42	17	0,53	0,89	0,6	0,88	0,52	87	10	0,31	-0,03	0,42	0,01	0,37
43	12	0,38	0,17	0,46	0,22	0,41	88	6	0,19	-0,87	0,26	-0,76	0,24
44	17	0,53	0,5	0,52	0,49	0,46	89	10	0,31	-0,32	0,36	-0,25	0,33
45	8	0,25	-0,48	0,33	-0,5	0,29	90	12	0,38	0,65	0,55	0,65	0,48

continua

Al	T. A	TCT	MLD1-32 itens		MLD1-G1,G2 e Q23		Al	T. A	TCT	MLD1-32 itens		MLD1-G1,G2 e Q23	
			3PL	Escala P.	3PL	Escala P.				3PL	Escala P.	3PL	Escala P.
91	19	0,59	1,11	0,64	1,05	0,55	136	16	0,5	0,44	0,51	0,41	0,44
92	17	0,53	0,95	0,61	0,93	0,53	137	7	0,22	-1,61	0,11	-1,64	0,09
93	14	0,44	0,26	0,48	0,26	0,42	138	26	0,81	1,94	0,8	2,13	0,74
94	10	0,31	-0,23	0,38	-0,23	0,33	139	10	0,31	-0,14	0,4	-0,17	0,34
95	18	0,56	0,96	0,61	0,9	0,53	140	16	0,5	0,76	0,57	0,76	0,5
96	20	0,63	1,25	0,67	1,3	0,6	141	17	0,53	0,9	0,6	0,84	0,52
97	12	0,38	0,22	0,47	0,23	0,41	142	8	0,25	-0,35	0,36	-0,34	0,31
98	4	0,13	-2,19	0	-2,16	0	143	9	0,28	-0,54	0,32	-0,57	0,27
99	8	0,25	-0,24	0,38	-0,21	0,33	144	12	0,38	-0,12	0,4	-0,15	0,35
100	14	0,44	0,47	0,52	0,5	0,46	145	6	0,19	-0,9	0,25	-0,88	0,22
101	18	0,56	1,15	0,65	1,18	0,57	146	10	0,31	-0,01	0,43	-0,01	0,37
102	8	0,25	-1,69	0,1	-1,7	0,08	147	11	0,34	-0,37	0,35	-0,33	0,31
103	4	0,13	-0,15	0,4	-0,07	0,36	148	6	0,19	-0,91	0,25	-0,86	0,22
104	18	0,56	0,52	0,53	0,41	0,44	149	16	0,5	0,65	0,55	0,65	0,48
105	9	0,28	-0,2	0,39	-0,15	0,35	150	8	0,25	-0,95	0,24	-0,97	0,21
106	16	0,5	0,61	0,54	0,61	0,48	151	12	0,38	0,3	0,48	0,29	0,42
107	14	0,44	0,56	0,54	0,56	0,47	152	9	0,28	-0,31	0,37	-0,27	0,32
108	14	0,44	0,49	0,52	0,54	0,46	153	5	0,16	-1,85	0,07	-1,9	0,05
109	10	0,31	-0,37	0,35	-0,39	0,31	154	9	0,28	-0,98	0,24	-0,93	0,21
110	12	0,38	0,25	0,47	0,27	0,42	155	7	0,22	-0,75	0,28	-0,71	0,25
111	10	0,31	-0,37	0,36	-0,35	0,31	156	13	0,41	0,22	0,47	0,22	0,41
112	15	0,47	0,74	0,57	0,75	0,5	157	9	0,28	-0,87	0,26	-0,9	0,22
113	9	0,28	-0,46	0,34	-0,5	0,29	158	22	0,69	1,35	0,69	1,31	0,6
114	12	0,38	0,13	0,45	0,12	0,39	159	9	0,28	-0,76	0,28	-0,76	0,24
115	10	0,31	-0,55	0,32	-0,56	0,28	160	9	0,28	-0,52	0,33	-0,58	0,27
116	12	0,38	-0,22	0,38	-0,19	0,34	161	10	0,31	0,19	0,46	0,18	0,4
117	14	0,44	0,52	0,53	0,56	0,47	162	10	0,31	-0,08	0,41	-0,05	0,36
118	11	0,34	0,25	0,47	0,28	0,42	163	13	0,41	0,1	0,45	0,11	0,39
119	15	0,47	0,69	0,56	0,69	0,49	164	9	0,28	-0,04	0,42	-0,04	0,36
120	12	0,38	-0,29	0,37	-0,23	0,33	165	15	0,47	0,4	0,5	0,4	0,44
121	15	0,47	0,74	0,57	0,76	0,5	166	13	0,41	-0,25	0,38	-0,25	0,33
122	11	0,34	-0,06	0,42	-0,06	0,36	167	17	0,53	0,86	0,59	0,86	0,52
123	5	0,16	-1,75	0,09	-1,76	0,07	168	10	0,31	-0,33	0,36	-0,32	0,32
124	11	0,34	0,02	0,43	0	0,37	169	14	0,44	0,49	0,52	0,49	0,46
125	12	0,38	-0,33	0,36	-0,34	0,31	170	8	0,25	-0,68	0,29	-0,7	0,25
126	11	0,34	-0,16	0,39	-0,2	0,34	171	8	0,25	-0,41	0,35	-0,42	0,3
127	10	0,31	-0,33	0,36	-0,33	0,31	172	10	0,31	-0,41	0,35	-0,38	0,31
128	16	0,5	0,82	0,58	0,8	0,51	173	10	0,31	-0,13	0,4	-0,12	0,35
129	10	0,31	0,06	0,44	0,11	0,39	174	16	0,5	0,92	0,6	0,93	0,53
130	23	0,72	1,63	0,74	1,56	0,64	175	7	0,22	-0,41	0,35	-0,43	0,3
131	7	0,22	-0,84	0,26	-0,88	0,22	176	11	0,34	1,22	0,66	1,26	0,59
132	13	0,41	0,61	0,54	0,61	0,48	177	12	0,38	0,11	0,45	0,1	0,39
133	15	0,47	0,51	0,52	0,57	0,47	178	5	0,16	-0,62	0,31	-0,65	0,26
134	9	0,28	-0,52	0,32	-0,46	0,29	179	17	0,53	1,03	0,63	1,05	0,55
135	16	0,5	0,69	0,56	0,68	0,49	180	4	0,13	-1,79	0,08	-1,76	0,07

continua

Al	T. A	TCT	MLD1-32 itens		MLD1-G1,G2 e Q23		Al	T. A	TCT	MLD1-32 itens		MLD1-G1,G2 e Q23	
			3PL	Escala P.	3PL	Escala P.				3PL	Escala P.	3PL	Escala P.
181	10	0,31	-0,26	0,38	-0,2	0,34	226	7	0,22	-1,61	0,11	-1,65	0,09
182	11	0,34	-0,05	0,42	-0,04	0,37	227	5	0,16	-1,43	0,15	-1,46	0,12
183	9	0,28	-0,44	0,34	-0,49	0,29	228	12	0,38	0,02	0,43	0,01	0,37
184	14	0,44	0,2	0,47	0,19	0,4	229	8	0,25	-0,33	0,36	-0,31	0,32
185	11	0,34	-0,13	0,4	-0,14	0,35	230	9	0,28	-0,18	0,39	-0,17	0,34
186	20	0,63	1,07	0,63	0,99	0,54	231	8	0,25	-0,23	0,38	-0,17	0,34
187	8	0,25	-1,03	0,23	-1,05	0,19	232	8	0,25	-0,42	0,35	-0,4	0,3
188	9	0,28	-0,34	0,36	-0,32	0,32	233	10	0,31	-0,34	0,36	-0,32	0,32
189	9	0,28	-0,31	0,37	-0,35	0,31	234	13	0,41	0,2	0,46	0,17	0,4
190	7	0,22	-0,4	0,35	-0,39	0,3	235	9	0,28	-0,45	0,34	-0,47	0,29
191	8	0,25	-0,7	0,29	-0,64	0,26	236	8	0,25	-0,74	0,28	-0,74	0,24
192	8	0,25	-0,35	0,36	-0,38	0,31	237	11	0,34	-0,13	0,4	-0,1	0,35
193	12	0,38	0,27	0,48	0,25	0,42	238	6	0,19	-0,47	0,34	-0,44	0,3
194	10	0,31	-0,46	0,34	-0,48	0,29	239	6	0,19	-0,8	0,27	-0,75	0,24
195	10	0,31	-0,37	0,35	-0,38	0,31	240	5	0,16	-1,54	0,13	-1,52	0,11
196	10	0,31	-0,26	0,38	-0,23	0,33	241	4	0,13	-1,4	0,16	-1,37	0,14
197	13	0,41	0,25	0,47	0,25	0,42	242	12	0,38	0,18	0,46	0,21	0,41
198	10	0,31	-0,06	0,42	-0,07	0,36	243	9	0,28	-0,96	0,24	-0,92	0,21
199	10	0,31	-0,26	0,38	-0,26	0,33	244	7	0,22	-0,75	0,28	-0,76	0,24
200	14	0,44	0,04	0,43	-0,02	0,37	245	11	0,34	0,02	0,43	0,02	0,38
201	8	0,25	-0,51	0,33	-0,53	0,28	246	7	0,22	-0,82	0,27	-0,78	0,24
202	8	0,25	-0,55	0,32	-0,59	0,27	247	11	0,34	-0,45	0,34	-0,4	0,3
203	12	0,38	0,3	0,48	0,33	0,43	248	7	0,22	-1,55	0,13	-1,51	0,11
204	10	0,31	-0,16	0,39	-0,17	0,34	249	3	0,09	-1,85	0,07	-1,76	0,07
205	9	0,28	-0,34	0,36	-0,35	0,31	250	8	0,25	-0,32	0,37	-0,29	0,32
206	22	0,69	1,66	0,75	1,7	0,66	251	7	0,22	-0,9	0,25	-0,81	0,23
207	11	0,34	-0,04	0,42	0	0,37	252	5	0,16	-0,89	0,25	-0,93	0,21
208	15	0,47	0,44	0,51	0,41	0,44	253	7	0,22	-0,7	0,29	-0,72	0,25
209	5	0,16	-1,39	0,16	-1,33	0,14	254	10	0,31	-0,87	0,26	-0,9	0,22
210	9	0,28	-0,25	0,38	-0,27	0,33	255	7	0,22	-0,49	0,33	-0,52	0,28
211	8	0,25	-0,61	0,31	-0,63	0,26	256	10	0,31	-0,18	0,39	-0,2	0,34
212	12	0,38	-0,17	0,39	-0,17	0,34	257	15	0,47	0,43	0,51	0,52	0,46
213	9	0,28	-0,72	0,29	-0,75	0,24	258	6	0,19	-0,82	0,27	-0,8	0,23
214	13	0,41	-0,35	0,36	-0,32	0,32	259	12	0,38	0,23	0,47	0,26	0,42
215	15	0,47	0,51	0,53	0,51	0,46	260	11	0,34	-0,07	0,41	-0,1	0,35
216	9	0,28	-1,15	0,2	-1,16	0,17	261	11	0,34	0,04	0,44	-0,02	0,37
217	4	0,13	-1,42	0,15	-1,35	0,14	262	14	0,44	0,66	0,55	0,65	0,48
218	11	0,34	-0,05	0,42	-0,05	0,36	263	9	0,28	-0,4	0,35	-0,43	0,3
219	3	0,09	-1,46	0,14	-1,5	0,11	264	7	0,22	-1,17	0,2	-1,26	0,16
220	7	0,22	-1,23	0,19	-1,27	0,15	265	10	0,31	-0,73	0,28	-0,69	0,25
221	10	0,31	0,1	0,45	0,14	0,4	266	9	0,28	-0,59	0,31	-0,61	0,27
222	11	0,34	0	0,43	-0,06	0,36	267	9	0,28	-0,84	0,26	-0,89	0,22
223	7	0,22	-0,71	0,29	-0,71	0,25	268	7	0,22	-0,85	0,26	-0,89	0,22
224	7	0,22	-1,13	0,21	-1,16	0,17	269	10	0,31	-0,44	0,34	-0,48	0,29
225	11	0,34	0,19	0,46	0,17	0,4	270	10	0,31	-1,54	0,13	-1,63	0,09

continua

Al	T. A	TCT	MLD1-32 itens		MLD1-G1,G2 e Q23		Al	T. A	TCT	MLD1-32 itens		MLD1-G1,G2 e Q23	
			3PL	Escala P.	3PL	Escala P.				3PL	Escala P.	3PL	Escala P.
271	10	0,31	-0,36	0,36	-0,35	0,31	316	8	0,25	-0,59	0,31	-0,62	0,26
272	9	0,28	-0,8	0,27	-0,76	0,24	317	16	0,5	0,32	0,49	0,21	0,41
273	10	0,31	-1	0,23	-0,99	0,2	318	13	0,41	0,24	0,47	0,22	0,41
274	12	0,38	-0,4	0,35	-0,44	0,3	319	11	0,34	-0,21	0,39	-0,18	0,34
275	12	0,38	0,24	0,47	0,23	0,41	320	18	0,56	0,82	0,58	0,8	0,51
276	10	0,31	-1,16	0,2	-1,14	0,18	321	14	0,44	0,39	0,5	0,4	0,44
277	11	0,34	-0,21	0,39	-0,19	0,34	322	20	0,63	1,29	0,68	1,32	0,6
278	13	0,41	0	0,43	-0,02	0,37	323	14	0,44	0,6	0,54	0,63	0,48
279	10	0,31	-0,39	0,35	-0,43	0,3	324	10	0,31	-0,33	0,36	-0,34	0,31
280	7	0,22	-0,65	0,3	-0,69	0,25	325	14	0,44	0,18	0,46	0,18	0,4
281	6	0,19	-0,58	0,31	-0,61	0,27	326	8	0,25	-0,59	0,31	-0,53	0,28
282	13	0,41	0,39	0,5	0,39	0,44	327	7	0,22	-0,6	0,31	-0,54	0,28
283	12	0,38	0	0,43	-0,01	0,37	328	9	0,28	-0,61	0,31	-0,56	0,28
284	13	0,41	0,14	0,45	0,19	0,4	329	11	0,34	0,01	0,43	-0,01	0,37
285	7	0,22	-0,78	0,28	-0,73	0,25	330	8	0,25	-0,69	0,29	-0,67	0,26
286	12	0,38	-0,64	0,3	-0,64	0,26	331	13	0,41	0,55	0,53	0,54	0,46
287	21	0,66	1,45	0,71	1,42	0,62	332	9	0,28	-0,39	0,35	-0,42	0,3
288	5	0,16	-1,16	0,2	-1,14	0,18	333	17	0,53	0,78	0,58	0,78	0,5
289	4	0,13	-2,1	0,02	-2,07	0,02	334	17	0,53	0,74	0,57	0,72	0,5
290	14	0,44	0,38	0,5	0,38	0,44	335	9	0,28	-0,45	0,34	-0,42	0,3
291	9	0,28	-0,78	0,28	-0,82	0,23	336	10	0,31	-0,06	0,41	-0,07	0,36
292	10	0,31	0,05	0,44	0,09	0,39	337	18	0,56	1,5	0,72	1,56	0,64
293	9	0,28	-1,24	0,19	-1,33	0,14	338	12	0,38	0,23	0,47	0,22	0,41
294	13	0,41	0,05	0,44	0,04	0,38	339	10	0,31	-1,14	0,21	-1,06	0,19
295	11	0,34	-0,06	0,41	-0,06	0,36	340	12	0,38	-0,49	0,33	-0,53	0,28
296	15	0,47	0,66	0,56	0,72	0,5	341	16	0,5	0,91	0,6	0,9	0,53
297	12	0,38	-0,09	0,41	-0,06	0,36	342	8	0,25	-0,67	0,3	-0,64	0,26
298	10	0,31	-0,03	0,42	-0,03	0,37	343	15	0,47	0,58	0,54	0,58	0,47
299	12	0,38	0,24	0,47	0,21	0,41	344	13	0,41	0,12	0,45	0,1	0,39
300	18	0,56	1,16	0,65	1,2	0,58	345	14	0,44	0,62	0,55	0,58	0,47
301	19	0,59	1,26	0,67	1,29	0,59	346	16	0,5	0,66	0,55	0,64	0,48
302	13	0,41	0,39	0,5	0,38	0,44	347	17	0,53	0,84	0,59	0,84	0,52
303	11	0,34	-0,03	0,42	-0,05	0,36	348	12	0,38	0,18	0,46	0,15	0,4
304	5	0,16	-1,11	0,21	-1,17	0,17	349	18	0,56	1,15	0,65	1,18	0,57
305	8	0,25	-0,5	0,33	-0,47	0,29	350	5	0,16	-1,58	0,12	-1,62	0,09
306	10	0,31	-0,32	0,36	-0,29	0,32	351	22	0,69	1,54	0,73	1,63	0,65
307	14	0,44	0,25	0,47	0,22	0,41	352	27	0,84	2,3	0,87	2,27	0,76
308	7	0,22	-1,38	0,16	-1,42	0,13	353	20	0,63	1,26	0,67	1,31	0,6
309	14	0,44	0,2	0,46	0,18	0,4	354	25	0,78	1,7	0,76	1,67	0,66
310	5	0,16	-0,85	0,26	-0,84	0,23	355	17	0,53	0,97	0,62	0,93	0,53
311	13	0,41	0,09	0,44	0,11	0,39	356	18	0,56	1,35	0,69	1,28	0,59
312	5	0,16	-0,67	0,3	-0,68	0,25	357	29	0,91	2,47	0,9	2,44	0,79
313	10	0,31	-0,19	0,39	-0,14	0,35	358	19	0,59	1,31	0,68	1,29	0,59
314	8	0,25	-0,62	0,31	-0,66	0,26	359	21	0,66	1,25	0,67	1,2	0,58
315	11	0,34	-0,01	0,42	0,01	0,37	360	17	0,53	0,92	0,61	0,88	0,52

continua

Al	T. A	TCT	MLD1-32 itens		MLD1-G1,G2 e Q23		Al	T. A	TCT	MLD1-32 itens		MLD1-G1,G2 e Q23	
			3PL	Escala P.	3PL	Escala P.				3PL	Escala P.	3PL	Escala P.
361	15	0,47	0,75	0,57	0,78	0,51	399	26	0,81	1,92	0,8	1,85	0,69
362	25	0,78	1,61	0,74	1,56	0,64	400	22	0,69	1,91	0,8	2,09	0,73
363	23	0,72	1,48	0,71	1,53	0,63	401	17	0,53	1,05	0,63	1,06	0,55
364	18	0,56	1,29	0,68	1,24	0,58	402	12	0,38	0,24	0,47	0,23	0,41
365	19	0,59	1,23	0,67	1,29	0,59	403	12	0,38	0,34	0,49	0,33	0,43
366	27	0,84	2,16	0,85	2,1	0,73	404	12	0,38	0,09	0,44	0,09	0,39
367	22	0,69	1,67	0,75	1,74	0,67	405	9	0,28	-0,82	0,27	-0,92	0,21
368	19	0,59	1,28	0,67	1,3	0,6	406	22	0,69	1,78	0,77	1,67	0,66
369	9	0,28	-0,85	0,26	-0,84	0,23	407	20	0,63	1,18	0,66	1,2	0,58
370	16	0,5	0,98	0,62	1,02	0,55	408	8	0,25	-0,25	0,38	-0,33	0,31
371	16	0,5	1	0,62	1,05	0,55	409	6	0,19	-1,15	0,2	-1,15	0,17
372	19	0,59	0,93	0,61	0,94	0,53	410	2	0,06	-2,12	0,01	-2,17	0
373	14	0,44	0,65	0,55	0,64	0,48	411	8	0,25	-0,6	0,31	-0,63	0,26
374	10	0,31	-0,63	0,3	-0,61	0,27	412	11	0,34	0,11	0,45	0,13	0,39
375	9	0,28	-0,38	0,35	-0,33	0,31	413	8	0,25	-0,36	0,36	-0,37	0,31
376	10	0,31	0,02	0,43	0,01	0,37	414	9	0,28	-0,21	0,39	-0,17	0,34
377	16	0,5	0,75	0,57	0,74	0,5	415	11	0,34	-0,13	0,4	-0,1	0,35
378	11	0,34	-0,02	0,42	0,04	0,38	416	7	0,22	-0,88	0,26	-0,91	0,22
379	14	0,44	-0,01	0,42	-0,04	0,36	417	10	0,31	-0,18	0,39	-0,19	0,34
380	19	0,59	1,36	0,69	1,45	0,62	418	14	0,44	0,57	0,54	0,56	0,47
381	9	0,28	-0,41	0,35	-0,38	0,31	419	10	0,31	-0,19	0,39	-0,17	0,34
382	17	0,53	0,98	0,62	1,02	0,55	420	9	0,28	-0,69	0,29	-0,68	0,26
383	11	0,34	0,16	0,46	0,18	0,4	421	15	0,47	0,44	0,51	0,39	0,44
384	14	0,44	0,73	0,57	0,76	0,5	422	12	0,38	0,26	0,48	0,31	0,42
385	5	0,16	-1,34	0,17	-1,38	0,13	423	10	0,31	-0,27	0,37	-0,33	0,32
386	15	0,47	0,83	0,59	0,81	0,51	424	8	0,25	-1,08	0,22	-0,98	0,2
387	8	0,25	-0,86	0,26	-0,82	0,23	425	7	0,22	-1,13	0,21	-1,17	0,17
388	14	0,44	0,53	0,53	0,52	0,46	426	6	0,19	-1,71	0,1	-1,74	0,07
389	16	0,5	0,64	0,55	0,64	0,48	427	8	0,25	-0,77	0,28	-0,8	0,23
390	16	0,5	0,82	0,59	0,81	0,51	428	7	0,22	-0,94	0,24	-0,87	0,22
391	18	0,56	1,01	0,62	1,03	0,55	429	6	0,19	-1,89	0,06	-1,84	0,06
392	23	0,72	1,85	0,79	1,78	0,68	430	6	0,19	-0,87	0,26	-0,91	0,21
393	16	0,5	0,73	0,57	0,74	0,5	431	16	0,5	0,39	0,5	0,34	0,43
394	5	0,16	-0,82	0,27	-0,86	0,22	432	12	0,38	0,02	0,43	0,06	0,38
395	9	0,28	-0,26	0,38	-0,3	0,32	433	10	0,31	0,04	0,43	0,06	0,38
396	14	0,44	0,21	0,47	0,18	0,4	434	7	0,22	-0,89	0,25	-0,95	0,21
397	8	0,25	-0,73	0,28	-0,78	0,24	435	13	0,41	-0,51	0,33	-0,59	0,27
398	20	0,63	1,63	0,74	1,52	0,63	436	9	0,28	0,22	0,47	0,25	0,41

**ANEXO A – Dados do exame multidisciplinar do
conjunto de dados I.**

**ANEXO B – Dados do exame multidisciplinar do
conjunto de dados II.**

Nota	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Ac.			
3	0	1	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0	1	0	1	0	0	0	12		
3.8	1	1	0	0	1	1	1	1	1	0	0	0	0	1	0	0	0	1	1	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	15	
5.3	1	1	0	0	1	1	1	0	1	1	0	1	1	0	1	0	0	1	1	1	0	0	1	1	0	0	1	0	0	1	0	0	1	0	1	0	1	1	0	0	1	21		
4.8	1	0	0	1	1	1	0	1	1	1	0	0	0	1	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	0	1	0	1	0	0	0	19		
6.3	0	1	0	1	1	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	0	0	0	1	1	0	1	25		
5	1	0	1	0	1	1	1	0	1	1	1	1	0	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	20		
6	1	0	0	1	1	0	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	1	1	0	0	0	1	0	0	24	
6	0	0	1	1	1	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	0	1	0	1	0	1	1	0	1	1	1	1	0	1	0	0	0	0	1	0	24		
4.5	1	0	0	1	1	1	0	0	1	0	1	1	1	1	1	1	0	1	0	1	1	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	18		
4	0	1	1	0	0	1	1	1	1	1	0	0	0	0	1	0	1	0	1	1	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	16		
2.3	0	0	0	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9		
3.8	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	1	1	0	0	1	1	0	0	1	0	0	1	0	0	0	1	0	0	15	
4.5	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	1	0	1	1	1	0	1	0	1	0	0	1	0	1	0	1	1	0	0	0	1	0	0	0	0	0	18		
4.3	1	1	1	0	1	1	1	1	1	0	0	0	1	0	0	0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	17		
2.8	0	0	1	1	1	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	11		
4.5	1	0	0	1	1	1	1	0	1	1	0	0	1	0	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	18		
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
4.3	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17		
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
4	1	1	0	0	1	1	1	0	1	0	1	0	0	1	0	0	1	0	1	0	1	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	16	
4.8	0	1	1	0	1	0	1	0	1	0	1	1	1	0	1	1	1	1	0	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	
3.3	0	1	0	0	1	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13		
4.5	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1	1	1	0	0	0	0	0	1	0	1	0	0	18		
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
4	0	1	0	0	0	1	1	1	0	0	0	0	1	0	0	1	0	1	1	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	16	
4.3	1	1	1	0	1	1	0	1	1	0	0	1	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	17	
4.3	1	1	0	0	1	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	17	
2.5	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	10		
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
4.5	1	1	0	0	1	0	0	0	1	1	0	1	1	0	0	1	0	1	1	0	1	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4.3	1	0	1	1	1	1	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3.8	0	0	1	1	0	0	0	0	1	1	1	0	1	0	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	
4.3	0	1	0	0	1	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0	1	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	17
3.8	1	0	1	0	1	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	15	
4	1	0	0	0	1	1	1	1	0	1	1	0	1	0	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	
5	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	20	
5.5	1	1	1	1	1	0	0	1	1	1	0	0	0	1	0	1	1	0	0	1	0	0	0	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	22		
5.3	1	1	0	1	1	1	0	1	1	0	1	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	0													

Nota	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Ac.			
3.8	0	0	0	1	1	0	0	0	1	1	1	0	1	0	0	0	0	1	0	1	1	1	0	1	0	1	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	15		
3.8	0	1	0	1	1	1	0	0	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	15	
4.8	1	1	0	0	1	1	1	0	0	0	1	0	1	0	0	1	1	1	1	0	0	1	0	0	0	0	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	19		
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
4.5	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	0	1	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1	0	1	0	1	0	1	0	0	0	18	
5.3	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	1	0	1	0	1	1	0	1	1	0	1	1	0	0	1	1	0	1	1	0	1	0	1	0	0	0	21		
3.5	0	0	0	0	1	1	1	1	1	0	0	0	1	0	0	1	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	14	
5.3	1	1	0	1	1	1	1	0	0	0	0	0	1	1	0	1	0	1	1	1	0	1	1	0	1	0	0	0	0	1	1	1	1	1	1	0	1	1	0	0	0	21		
4	1	0	1	0	1	1	1	0	0	0	0	0	1	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0	1	0	1	0	1	1	0	1	0	0	0	0	0	16		
6.3	1	0	0	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	0	1	1	1	0	1	1	0	1	1	0	1	1	0	1	0	1	0	0	1	25	
4	1	1	0	1	1	0	0	0	1	0	0	1	1	0	0	0	1	1	0	0	1	1	0	0	1	0	0	1	0	1	1	0	1	1	0	0	1	0	0	0	0	16		
6	1	1	0	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	1	0	1	0	1	1	1	1	1	1	1	0	1	0	0	0	0	24		
4.3	1	0	0	1	1	1	1	0	1	1	1	0	1	0	0	1	1	0	1	1	0	1	1	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	17		
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
5	1	0	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	1	20
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6.8	1	0	1	0	1	1	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	27		
4.8	1	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	1	0	1	0	1	19
4.8	1	1	0	1	1	1	1	0	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	19	
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	1	0	0	1	0	1	1	0	1	1	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0	0	0	16	
4	1	0	0	1	1	0	1	0	0	1	1	0	0	0	0	1	1	0	0	1	1	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	0	1	0	1	0	0	16	
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4.5	0	1	1	1	1	1	0	0	1	0	0	1	0	1	0	1	1	1	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	0	0	18		
—	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4.5	1	0	1	1	1	1	0	0	1	0	1	1	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	18	
5	1	0	1	0	1	1	1	1	0	1	1	1	1	0	0	0	1	1	1	1	0	1	0	1	0	0	1	1	0	0	1	1	0	1	1	0	0	0	0	0	0	0	20	
3.3	0	1	0	0	1	1	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	13	
6	1	1	1	1	1	1	0	1	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	24	
4.3	1	0	0	0	1	1	0	1	0	0	1	0	1	0	1	0	1	1	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	1	0	17	
2.5	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	10		
4.5	0	1	0	0	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	1	1	1	0	0	1	1	0	0	1	1	0	0	1	0	0	1	0	0	0	18		
4.3	1	0	0	1	0	0	1	0	1	0	1	0	0	1	1	1	0	0	0	1	1	0	0	0	1	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	17		
4.3	0	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	1	0	0	1	1	0	0	1	1	0	1	1	0	1	1	0	0	0	17		
3.8	0	0	0	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0	15		
4.8	1	0	0	1	1	1	0	0	1	1	1	0	1	0	0	0	1	1	1	0	1	1	0	1	1	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	19		
5	0	0	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	1	0	0	1	1	0	0	1	0	0	0	1	0	0	1	1	0	0	1	1	0	0	0	1	20	
4.8	1	0	0	1	0	1	0	1	1	1	0	1	0	1	1	1	1	0	1	1	0	1	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	19		
4.3	1	1	0	1	0	1	1	0	1	1	0	1	0	0	0	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	17		
6.8	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	1	1	0	1	0	0	1	0	0	0	27		
5.5	1	1	1	1	1	1	0	0	1	1	0	1	1	1	0	1	0	1	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1	1	0	1	22	
5	0	1	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	0	1	1	0	0	1	0	0	1	0	0	0	0	1	1	0	1	0	0	0	1	0	1	0	1	20	
5.5	1	1	0	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	1	0	1	1	0	1	0	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	22	

Nota: Total de acertos (A.c)