

**Equivalência da Análise Longitudinal de Dados
Multidimensionais através da Teoria da Resposta ao Item e de
Modelos de Equações Estruturais**

Hugo Kenji Pereira Harada

TESE APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA

OBTENÇÃO DO TÍTULO

DE

DOUTOR EM CIÊNCIAS

Programa: Doutorado em Estatística

Orientadora: Profa. Dra. Júlia Maria Pavan Soler

São Paulo, Dezembro de 2022

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da

Adaptativa Inteligência Educacional Ltda. e da Ânima Educação S.A.

Resumo

A *Teoria da Resposta ao Item (TRI)* e *Modelos de Equações Estruturais (MEE)* são técnicas bastante difundidas de mensuração de variáveis latentes através da observação de indicadores categóricos. A equivalência das duas metodologias em estudos transversais utilizando o modelo ogiva normal de dois parâmetros está bem estabelecida na literatura. Neste trabalho procura-se estender a equivalência dessas metodologias em modelos longitudinais multidimensionais buscando explorar e comparar propriedades de ajuste e implementação computacional das duas técnicas envolvidas. Primeiramente, para o caso uni e multidimensional sob dados de estudos transversais, as equivalências entre os modelos da TRI e de MEE são examinadas sob diferentes possibilidades de parametrizações e a implementação computacional é considerada por meio de dados simulados e reais, usando recursos dos pacotes MIRT e `lavaan` do R. Os resultados são então estendidos para propor modelos multidimensionais longitudinais cuja equivalência entre a TRI e MEE possa ser estabelecida.

Palavras-chave: Teoria da Resposta ao Item, Modelos de Equações Estruturais, Modelos Multidimensionais, Análise de Variáveis Latentes, Modelos Longitudinais, Análise Fatorial Binária.

Abstract

Item Response Theory (IRT) and *Structured Equation Models* (SEM) are well known techniques of latent variables measurement through the observation of categorical indicators. The equivalence of these two methodologies in cross-sectional studies using the two-parameter normal ogive model is well established in the literature. This work attempts to extend the equivalency of these methods in longitudinal and multidimensional studies while exploring and comparing model fit statistics and computational implementation of the techniques. As a first step, literature review was thoroughly conducted, and TRI and SEE uni- and multi-dimensional models were fitted to cross-sectional study data. Emphasis was put on showing the equivalency between the results. Proposal of TRI and SEE multi-dimensional longitudinal models whose equivalency is studied and verified are presented.

Keywords: Item Response Theory, Structural Equation Models, Latent Variable Analyses, Multidimensional Models, Longitudinal models, Binary Factor Analysis.

Não chores, meu filho;
Não chores, que a vida
É luta renhida:
Viver é lutar.
A vida é combate,
Que os fracos abate,
Que os fortes, os bravos
Só pode exaltar.

Antônio Gonçalves Dias (1864)

Copyright

© Hugo Harada 2022.

Sumário

Equivalência da Análise Longitudinal de Dados Multidimensionais através da Teoria da Resposta ao Item e de Modelos de Equações Estruturais	i
Resumo	ii
Abstract.....	iii
Copyright.....	v
Sumário	vi
Lista de figuras	x
Lista de tabelas.....	xiv
Lista de abreviações	xvii
Notação.....	xviii
Agradecimentos	xx
Capítulo 1 Introdução	1
1.1 Motivação	2
1.2 Objetivos	6
1.3 Contribuições esperadas.....	7
1.4 Organização do trabalho.....	7
Capítulo 2 Variáveis latentes	9
2.1 Introdução.....	10
2.2 Definições	10
2.2.1 Variáveis observáveis ou indicadores	10
2.2.2 Definições não formais de variáveis latentes.....	11
2.2.3 Definições formais de variáveis latentes	12
2.3 Propriedades de variáveis latentes	13
2.3.1 Tipos de variáveis latentes.....	14

2.3.2	Identificabilidade	14
2.3.3	Indeterminação fatorial.....	14
2.3.4	Variáveis observáveis como indicadores de causa ou efeito	16
2.4	Notação	17
Capítulo 3	A teoria da resposta ao item.....	23
3.1	Introdução.....	24
3.2	Curva característica do item.....	24
3.3	Modelos unidimensionais transversais	25
3.3.1	Modelo ogiva normal	25
3.3.2	Modelo logístico de um parâmetro - ML1P	27
3.3.3	Modelo logístico de dois parâmetros - ML2P	29
3.3.4	Modelo de três parâmetros - M3P	31
3.3.5	Modelo de quatro parâmetros - M4P	32
3.3.6	Modelo de cinco parâmetros - M5P	33
3.3.7	Relação entre os modelos logístico e ogiva normal.....	35
3.4	Modelos multidimensionais	37
3.4.1	Modelo multidimensional logístico.....	37
3.5	Estimação dos parâmetros e habilidades na TRI	41
3.5.1	Estimação do traço latente.....	43
3.5.2	Estimação dos parâmetros dos itens	44
3.5.3	Estimação conjunta do traço latente e dos parâmetros dos itens	44
3.5.4	Estimação por máxima verossimilhança marginal.....	44
3.6	Modelos de grupos múltiplos	46
3.6.1	Equalização em modelos de grupos múltiplos.....	46
3.6.2	Equalização via estimação simultânea	48
3.6.3	Equalização a posteriori.....	51
Capítulo 4	Modelos de equações estruturais.....	55

4.1	Introdução.....	56
4.2	Modelagem MEE.....	57
4.2.1	Cenário unidimensional transversal.....	57
4.2.2	Cenário multidimensional transversal.....	61
4.2.3	Cenário unidimensional longitudinal.....	64
4.2.4	Cenário multidimensional longitudinal.....	64
4.3	Estimação dos parâmetros dos modelos.....	64
4.4	Identificabilidade.....	66
4.4.1	Identificação com um indicador por variável latente.....	66
4.4.2	Identificação com dois indicadores por variável latente.....	67
4.4.3	Identificação com três ou mais indicadores por variável latente.....	67
4.5	Parametrização.....	68
4.6	Medidas de ajuste.....	68
4.6.1	Estatística χ^2	69
4.6.2	Raiz do erro quadrático médio de aproximação (RMSEA).....	70
4.6.3	<i>Comparative fit index</i> (CFI).....	72
4.6.4	<i>Tucker–Lewis index</i> (TLI) e <i>Non-normed fit index</i> (NNFI).....	72
4.7	Invariância fatorial ou de mensuração.....	73
Capítulo 5	Simulações.....	75
5.1	Modelos transversais.....	76
5.1.1	Metodologia.....	76
5.2	Discussão.....	77
5.3	Modelos longitudinais.....	78
5.3.1	Metodologia.....	78
5.3.2	Discussão.....	86
Capítulo 6	Aplicação.....	96
6.1	Origem dos dados.....	97

6.2	Aplicação em modelos transversais	98
6.2.1	Análise utilizando a TRI e a TCT.....	99
6.2.2	Análise utilizando a MEE.....	109
6.2.3	Comparação entre os resultados obtidos pela TRI e pela MEE	116
6.3	Aplicação em modelos longitudinais.....	121
6.3.1	Análise preliminar de seleção de itens	121
6.3.2	Análise utilizando a TRI e a TCT.....	130
6.3.3	Análise utilizando a MEE.....	136
6.3.4	Comparação entre os resultados obtidos pela TRI e pela MEE	145
Capítulo 7	Comentários e pesquisas futuras	153
7.1	Considerações finais	154
7.2	Sugestões para pesquisas futuras.....	158
7.2.1	Aplicabilidade dos modelos propostos a microdados multidimensionais.....	158
7.2.2	Estudo de sensibilidade dos resultados apresentados aos parâmetros das simulações 158	
7.2.3	Suporte para dados faltantes sistemáticos em modelos MEE	158
7.2.4	Otimização de algoritmos visando melhor uso de recursos computacionais.....	159
7.2.5	Equivalência entre o modelo de fator comum e o modelo de três parâmetros da TRI 159	
7.2.6	Desenvolvimento de critérios de seleção de itens comuns em cenários multidimensionais.....	160
7.2.7	Estudo sobre <i>Intensive Longitudinal Data Models</i>	160
7.2.8	Distribuições assimétricas de traços latentes	160
Referências Bibliográficas	161

Lista de figuras

Figura 2.1 Exemplo de indeterminação fatorial.....	15
Figura 2.2 Distribuição dos escores fatoriais preditos por diferentes métodos. SIM = valores simulados	16
Figura 2.3 Modelo unidimensional transversal. I=3, K=1, T=1.....	19
Figura 2.4 Modelo multidimensional transversal. I=6, K=2, T=1. Em (a) a resposta ao item 11 depende de um único traço latente enquanto que em (b) essa mesma resposta depende de múltiplos traços.	20
Figura 2.5 Modelo unidimensional longitudinal. I=6, K=1, T=2.....	21
Figura 2.6 Modelo multidimensional longitudinal. I=12, K=2, T=2.....	22
Figura 3.1 Exemplo de uma curva característica do item.....	25
Figura 3.2 MON1 - Comportamento da função de distribuição normal acumulada quando $\sigma_i = 1.00$ e μ_i varia. MON2 - Comportamento da função de distribuição normal acumulada quando $\mu_i = 0.00$ e σ_i varia.	27
Figura 3.3 Exemplo de como diferentes valores de b_i influenciam o formato da CCI do modelo ML1P ($D = 1.00$).....	28
Figura 3.4 Exemplo de como diferentes valores de a_i influenciam o formato da CCI do modelo ML2P quando $b_i = 0.00$ e $D = 1.0$	30
Figura 3.5 Exemplo de como diferentes valores de c_i influenciam o formato da CCI do modelo M3P quando $a_i = 1.00$, $b_i = 0.00$ e $D = 1.00$	32
Figura 3.6 Exemplo de como diferentes valores de γ_j influenciam o formato da CCI do modelo M4P quando $a_i = 1.00$, $b_i = 0$, $c_i = 0.20$ e $D = 1.00$	33
Figura 3.7 Exemplo de como diferentes valores de δ_i influenciam o formato da CCI do modelo M4P quando $a_i = 1.00$, $b_i = 0.00$, $c_i = 0.20$, $\gamma_i = 0.80$ e $D = 1.00$	34
Figura 3.8 Funções de distribuição acumulada das distribuições normal e logística ($D=1.000$)	36
Figura 3.9 Funções de distribuição acumulada das distribuições normal e logística ($D=1.702$)	36
Figura 3.10 Exemplo da superfície característica do item quando $a_{i1} = 1.288$, $a_{i2} = 0.567$, $d_i = -0.047$, $c_i = 0.253$	39

Figura 3.11 Curvas de nível da superfície característica do item quando $ai1 = 1.288$, $ai2 = 0.567$, $di = -0.047$, $ci = 0.253$.	40
Figura 3.12 Dimensionalidades intra-item e entre-item.	41
Figura 3.13 Equalização de grupos equivalentes ou equalização via população	47
Figura 3.14 Equalização de grupos não equivalentes via itens comuns.	48
Figura 3.15 Equalização de grupos não equivalentes via itens comuns – Equalização simultânea	51
Figura 4.1 Relação entre τ_{it} e Y_{it} * segundo o modelo de valor limiar / valor de corte	58
Figura 4.2 Como os modelos saturado e nulo servem de referência para cálculo de medidas de ajuste do modelo a ser testado	69
Figura 5.1 Equalização de grupos não equivalentes via itens comuns – Equalização Simultânea	84
Figura 5.2 Modelo Multidimensional Longitudinal. $K=2$, $T=2$.	85
Figura 5.3 Estrutura de dados para equalização simultânea de quatro testes – modelo de um grupo.	86
Figura 5.4 Recuperação do parâmetro de discriminação ait utilizando a TRI – Modelo unidimensional longitudinal de dois momentos. C – Item comum. NC – Item não-comum. sim – valores simulados, est – valores estimados.	87
Figura 5.5 Recuperação do parâmetro de dificuldade bit utilizando a TRI – Modelo unidimensional longitudinal de dois momentos. C – Item comum. NC – Item não-comum. sim – valores simulados, est – valores estimados.	88
Figura 5.6 Recuperação do parâmetro de discriminação ait utilizando a MEE – Modelo multidimensional longitudinal de dois momentos. C – Item comum. NC – Item não-comum.	92
Figura 5.7 Recuperação do parâmetro de dificuldade bit utilizando a MEE – Modelo multidimensional longitudinal de dois momentos. C – Item comum. NC – Item não-comum.	93
Figura 6.1 Histograma dos escores brutos. S01_LP - TRI	102
Figura 6.2 Curva de informação do teste. S01_LP - TRI	105
Figura 6.3 Curvas característica dos itens finais, S01_LP - TRI	106
Figura 6.4 Comparação de métodos de estimação dos escores latentes, S01_LP - TRI	108

Figura 6.5 Distribuição dos escores latentes nos métodos de estimação. S01_LP - TRI.....	108
Figura 6.6 Comparação de métodos de estimação dos escores latentes. S01_LP - MEE. N=10 000.	115
Figura 6.7 Distribuição dos escores latentes nos métodos de estimação. S01_LP - MEE.	115
Figura 6.8 - Equivalência entre os valores de <i>ait</i> obtidos com modelos TRI e MEE. S01_LP..	116
Figura 6.9 - Equivalência entre os valores de <i>bit</i> obtidos com modelos TRI e MEE. S01_LP..	117
Figura 6.10 Comparação de métodos de estimação dos escores latentes – TRI e MEE. S01_LP.	119
Figura 6.11 Distribuição dos escores latentes nos métodos de estimação – TRI e MEE. S01_LP - MEE.....	120
Figura 6.12 Histograma dos escores brutos. S04_LP - TRI	122
Figura 6.13 Histograma dos escores brutos. S01_MT.....	125
Figura 6.14 Histograma dos escores brutos. S04_MT.....	128
Figura 6.15 Estrutura de grupos múltiplos de dados para equalização simultânea dos testes S01_LP e S04_LP – Modelo unidimensional longitudinal TRI.	131
Figura 6.16 Estrutura de grupos múltiplos de dados para equalização simultânea dos testes S01_MT e S04_MT – Modelo unidimensional longitudinal TRI.....	133
Figura 6.17 Estrutura de dados para equalização simultânea de quatro testes: S01_LP, S04_LP, S01_MT E S04_MT – modelo de um grupo.	138
Figura 6.18 Comparação das estimativas do parâmetro de discriminação <i>ait</i> . Modelos longitudinais. TRI vs. MEE.....	146
Figura 6.19 Comparação das estimativas do parâmetro de dificuldade <i>bit</i> . Modelos longitudinais. TRI vs. MEE.....	147
Figura 6.20 Estudo de sensibilidade do processo de estimação dos escores latentes utilizando parâmetros de itens estimados pela TRI e pela MEE. Método de estimação: MAP/EBM.	148
Figura A.1 Recuperação dos valores do parâmetro a utilizando a TRI. ML2P. 30 itens.....	171
Figura A.2 Recuperação dos valores do parâmetro b utilizando a TRI. ML2P. 30 itens.	171

Figura A.3 Recuperação dos valores do parâmetro a utilizando a MEE. ML2P. 30 itens.	174
Figura A.4 Recuperação dos valores do parâmetro b utilizando a MEE. ML2P. 30 itens.....	175
Figura A.5 Exemplo de organização de itens em desenho longitudinal de 3 momentos.....	177
Figura A.6 Estrutura de entrada de dados para modelos de grupos múltiplos	179
Figura A.7 - Recuperação de a - modelo longitudinal de grupos múltiplos	180
Figura A.8 - Recuperação de b - Modelo longitudinal de grupos múltiplos.....	181
Figura A.9 Modelo longitudinal MEE de dois momentos.....	183
Figura A.10 Recuperação do parâmetro de dificuldade a . Modelo longitudinal MEE.	187
Figura A.11 Recuperação do parâmetro de dificuldade b . Modelo longitudinal MEE.	188
Figura B.1 – Número de itens versus a correlação entre traço latente estimado e simulado.....	190
Figure D.1 Threshold model for a latent response variable with a binary indicator.....	198
Figura E.1 – Autorização para o uso de dados do Programa Especial de Aumento de Rendimento Escolar / Programa Escola + – Município de Cariacica/ES.	221

Lista de tabelas

Tabela 3.1 Valores de referência para o parâmetro de discriminação ai segundo Baker [49, p. 34]	31
Tabela 3.2 - Métodos de equalização a posteriori.....	54
Tabela 4.1 Valores de referência para interpretação do RMSEA.....	71
Tabela 4.2 Valores de referência para interpretação do CFI/TLI/NNFI.....	72
Tabela 4.3 Invariância fatorial - Resumo	74
Tabela 5.1 Valores simulados para parâmetros dos itens	82
Tabela 5.2 Recuperação dos parâmetros da distribuição das variáveis latentes – Modelo unidimensional ($K = 1$) longitudinal de dois momentos ($T = 2$). TRI.....	89
Tabela 5.3 Recuperação dos parâmetros da distribuição das variáveis latentes calculados a partir dos escores latentes individuais – Modelo unidimensional ($K = 1$) longitudinal de dois momentos ($T = 2$). TRI.	90
Tabela 5.4 Medidas de ajuste. Modelo multidimensional longitudinal. MEE.....	91
Tabela 5.5 Recuperação dos parâmetros de distribuição das variáveis latentes– Modelo multidimensional longitudinal de dois momentos. Valores brutos.....	94
Tabela 5.6 Recuperação dos parâmetros de distribuição das variáveis latentes– Modelo multidimensional longitudinal de dois momentos. Valores padronizados.	95
Tabela 6.1 Testes aplicados no Programa Especial de Aumento de Rendimento Escola – Escola +	98
Tabela 6.2 Estatísticas do teste, S01_LP - TRI.....	101
Tabela 6.3 Estatísticas de todos os itens. S01_LP - TRI.....	103
Tabela 6.4 Estatísticas dos itens finais, S01_LP - TRI.....	107
Tabela 6.5 - Carga fatorial dos itens em três diferentes parametrizações. S01_LP – MEE.	110
Tabela 6.6 – R^2 dos itens em três diferentes parametrizações. S01_LP – MEE.....	111
Tabela 6.7 - Medidas de ajuste. S01_LP - MEE.	113
Tabela 6.8 Parâmetros da distribuição da variável latente. S01_LP - MEE.	113

Tabela 6.9 Parâmetros dos itens. S01_LP - MEE.....	114
Tabela 6.10 - Valores dos parâmetros dos itens com modelagem TRI e MEE. S01_LP.....	118
Tabela 6.11 Estatísticas do teste, S04_LP - TRI.....	121
Tabela 6.12 Estatísticas de todos os itens. S04_LP - TRI.....	123
Tabela 6.13 Estatísticas do teste, S01_MT.....	124
Tabela 6.14 Estatísticas de todos os itens. S01_MT.....	126
Tabela 6.15 Estatísticas do teste, S04_MT.....	127
Tabela 6.16 Estatísticas de todos os itens. S04_MT.....	129
Tabela 6.17 Parâmetros dos itens - Modelo unidimensional longitudinal - TRI - S01_LP e S04_LP	132
Tabela 6.18 Parâmetros da variável latente - Modelo unidimensional longitudinal - TRI - S01_LP e S04_LP. Erro padrão entre parêntesis (não calculado para momento de referência).....	132
Tabela 6.19 Parâmetros dos itens - Modelo unidimensional longitudinal - TRI - S01_MT e S04_MT	134
Tabela 6.20 Parâmetros da variável latente - Modelo unidimensional longitudinal - TRI - S01_MP e S04_MP. Erro padrão entre parêntesis (não calculado para momento de referência).....	135
Tabela 6.21 Parâmetros e correspondentes estimativas da distribuição das variáveis latentes. TRI. Valores estimados. Erro padrão entre parêntesis (não calculado para momento de referência).	135
Tabela 6.22 Parâmetros e correspondentes estimativas da distribuição das variáveis latentes. TRI. Valores calculados a partir dos escores latentes individuais.....	136
Tabela 6.23 Medidas de ajuste. Modelo multidimensional longitudinal. MEE.....	139
Tabela 6.24 Parâmetros e correspondentes estimativas da distribuição das variáveis latentes. Modelo multidimensional longitudinal. MEE. Valores não padronizados.....	139
Tabela 6.25 Parâmetros e correspondentes estimativas da distribuição das variáveis latentes. Modelo multidimensional longitudinal. MEE. Valores padronizados.	140
Tabela 6.26 Parâmetros dos itens do teste S01_LP. Modelo multidimensional longitudinal MEE.	141

Tabela 6.27 Parâmetros dos itens do teste S04_LP. Modelo multidimensional longitudinal MEE.	142
Tabela 6.28 Parâmetros dos itens do teste S01_MT. Modelo multidimensional longitudinal MEE.	143
Tabela 6.29 Parâmetros dos itens do teste S04_MT. Modelo multidimensional longitudinal MEE.	144
Tabela 6.30 Comparação das estimativas do parâmetro dos itens do teste S01_LP. Modelos longitudinais. TRI vs. MEE.....	149
Tabela 6.31 Comparação das estimativas do parâmetro dos itens do teste S04_LP. Modelos longitudinais. TRI vs. MEE.....	150
Tabela 6.32 Comparação das estimativas do parâmetro dos itens do teste S01_MT. Modelos longitudinais. TRI vs. MEE.....	151
Tabela 6.33 Comparação das estimativas do parâmetro dos itens do teste S04_MT. Modelos longitudinais. TRI vs. MEE.....	152
Tabela A.1 Recuperação de parâmetros populacionais.	182
Tabela A.2 Estatísticas de Ajuste de Modelo Longitudinal MEE.....	187
Tabela A.3 Recuperação de parâmetros populacionais.	189

Lista de abreviações

Algoritmo EM	Algoritmo estimação-maximização/expectação-maximização (<i>expectation-maximization</i>)
AFIC	Análise Fatorial de Informação Completa.
CH	Ciências Humanas e suas Tecnologias
CN	Ciências da Natureza e suas Tecnologias
DIC	Critério de Informação da Deviance
DIF	Funcionamento Diferencial de Itens (<i>Differential Item Functioning</i>)
DRIFT	Deslocamento ou desvio dos parâmetros dos itens associado ao tempo
EAP	Estimação pela Média da Distribuição a Posteriori (<i>Expected a posteriori</i>)
ENEM	Exame Nacional do Ensino Médio
ENADE	Exame Nacional de Desempenho de Estudantes
EP	Erro padrão
IES	Instituição de Educação Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
LC	Linguagens, Códigos e suas Tecnologias
MVM	Máxima Verossimilhança Marginal
MEE	Modelo de Equações Estruturadas (<i>Structural Equation Modeling</i>)
MEC	Ministério da Educação
TCT	Teoria Clássica dos Testes
TRI	Teoria da Resposta ao Item (<i>Item Response Theory</i>)
TRIM	Teoria da Resposta ao Item Multidimensional
ML1P	Modelo Logístico de Um Parâmetro
ML2P	Modelo Logístico de Dois Parâmetros
M3P	Modelo de Três Parâmetros. Modelo de Birnbaum de Três Parâmetros
M4P	Modelo de Quatro Parâmetros
MON	Modelo Ogiva Normal
MT	Matemática e suas Tecnologias
NNFI	Non-normed fit index
QMCEM	Quasi-Monte Carlo EM
RD	Redação
RB	Razão de Bayes
RMSEA	Raiz do erro médio quadrático de aproximação (Root mean square error of approximation)
TLI	Tucker-Lewis index

Notação

i	Índice referente aos itens. $I = 1, \dots, I$.
h	Índice referente ao indivíduo (<i>human</i>). $h = 1, \dots, H$.
k	Índice referente ao traço latente. $k = 1, \dots, K$. Pode ser abreviado na notação de modelos unidimensionais.
t	Índice referente momento t . $t = 1, \dots, T$. Pode ser abreviado na notação de modelos transversais.
g	Índice referente ao grupo g . $g = 1, \dots, G$. Pode ser abreviado na notação de modelos com um grupo.
m	Índice referente a medidas ou experimentos. $M = 1, \dots, M$.
η_{hktg}	Observação do k -ésimo traço latente do indivíduo h pertencente ao grupo g no momento t .
$\boldsymbol{\eta}_{.ktg}$	O vetor $\boldsymbol{\eta}_{.ktg} = (\eta_{1ktg}, \eta_{2ktg}, \dots, \eta_{Hktg})^t$ tem comprimento H e contém o k -ésimo traço latente de H indivíduos do grupo g no momento t .
$\boldsymbol{\eta}_{..tg}$	O vetor $\boldsymbol{\eta}_{..tg} = (\boldsymbol{\eta}_{.1tg}^t, \boldsymbol{\eta}_{.2tg}^t, \dots, \boldsymbol{\eta}_{.Ktg}^t)^t$ tem comprimento K e contém os valores de todos os traços latentes de H indivíduos do grupo g no momento t .
η_{ktg}	Variável latente associada k -ésimo traço latente do indivíduo h do grupo g no momento t .
α_{ktg}	Média de η_{ktg} .
ψ_{ktg}	Variância de η_{ktg} .
$\boldsymbol{\pi}_{ktg}$	Vetor de parâmetros que definem a distribuição de uma variável aleatória. $\boldsymbol{\pi}_{ktg} = (\alpha_{ktg}, \psi_{ktg})^t$ quando η_{ktg} tem distribuição normal.
$\eta_{ktg} \mid \boldsymbol{\pi}_{ktg}$	Variável aleatória associada ao k -ésimo traço latente do indivíduo h pertencente ao grupo g no momento t com distribuição definida pelo vetor de parâmetros $\boldsymbol{\pi}_{ktg}$.
Y_{iht}	Variável aleatória relacionada a resposta do indivíduo h ao item i no momento t .
y_{iht}	Observação de Y_{iht} . Resposta ao item i do indivíduo h , no momento t .
$\mathbf{y}_{i.t}$	O vetor $\mathbf{y}_{i.t} = (y_{i1t}, y_{i2t}, \dots, y_{iHt})^t$ tem comprimento H e contém as respostas de H indivíduos ao item i no momento t .
$\mathbf{y}_{..t}$	O vetor $\mathbf{y}_{..t} = (\mathbf{y}_{.1t}^t, \mathbf{y}_{.2t}^t, \dots, \mathbf{y}_{.It}^t)^t$ tem comprimento HI e contém as respostas de todos os indivíduos que responderam aos I itens no momento t .
$\mathbf{y}_{...}$	O vetor $\mathbf{y}_{...} = (\mathbf{y}_{..1}^t, \mathbf{y}_{..2}^t, \dots, \mathbf{y}_{..T}^t)^t$ tem comprimento HIT e contém as respostas de H indivíduos à I itens em T momentos.
$\mathbf{y}_{.ht}$	O vetor $\mathbf{y}_{.ht} = (y_{1ht}, y_{2ht}, \dots, y_{Iht})^t$ tem comprimento I e contém o vetor de respostas do indivíduo h a um conjunto de I itens no momento t .

$f(X)$	Função densidade de probabilidade da variável aleatória X .
$f(X \boldsymbol{\pi})$	Função densidade de probabilidade da variável aleatória X com distribuição definida pelo conjunto $\boldsymbol{\pi}$ de parâmetros.
$\boldsymbol{\eta}_{(KT)}$	Variável aleatória KT -dimensional multivariada $\boldsymbol{\eta}_{(KT)} = (\eta_{1t}, \eta_{2t}, \dots, \eta_{KT})^t$.
I_g	Número total de itens apresentados ao grupo g .
$I_{g',g''}$	Número total de itens comuns entre os grupos g' e g'' .
a_{it}	Parâmetro de discriminação do i -ésimo item no momento t , com valor proporcional à inclinação da Curva Característica do Item (CCI) no ponto b_{it} . $a_{it'} = a_{it''}$ se houver invariância de mensuração entre os momentos t' e t'' .
b_{it}	Parâmetro de dificuldade do i -ésimo item no momento t , medido na mesma escala das habilidades. $b_{it'} = b_{it''}$ se houver invariância de mensuração entre os momentos t' e t'' .
c_{it}	Parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o i -ésimo item no momento t . $c_{it'} = c_{it''}$ se houver invariância de mensuração entre os momentos t' e t'' .
γ_{it}	Parâmetro de descuido do i -ésimo item. Representa a probabilidade de indivíduos com alta habilidade responderem incorretamente o item i no momento t . $\gamma_{it'} = \gamma_{it''}$ se houver invariância de mensuração entre os momentos t' e t'' .
δ_{it}	Parâmetro de assimetria do i -ésimo item no momento t . $\delta_{it'} = \delta_{it''}$ se houver invariância de mensuração entre os momentos t' e t'' .
$\boldsymbol{\zeta}_{it}$	Vetor de parâmetros associados ao item i no momento t .
$\boldsymbol{\zeta}$	O conjunto $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \dots, \boldsymbol{\zeta}_I)$ contém os parâmetros do I itens associados a uma aplicação.
λ_{it}	Carga do i -ésimo item no momento t no modelo de fator comum. $\lambda_{it'} = \lambda_{it''}$ se houver invariância de mensuração entre os momentos t' e t'' .
ν_{it}	Intercepto do i -ésimo item no momento t no modelo de fator comum. $\nu_{it'} = \nu_{it''}$ se houver invariância de mensuração entre os momentos t' e t'' .
ε_{iht}	Erro de mensuração do i -ésimo item associado ao indivíduo h no momento t no modelo de fator comum.
$\boldsymbol{\varepsilon}_{i.t}$	O vetor $\boldsymbol{\varepsilon}_{i.t} = (\varepsilon_{i1t}, \varepsilon_{i2t}, \dots, \varepsilon_{iHt})^t$ em comprimento H e contém os erros de mensuração observados nas respostas de H indivíduos ao item i no momento t .
ε_{it}	Variável aleatória associada à $\boldsymbol{\varepsilon}_{i.t}$. É comum assumir que a distribuição dos erros de mensuração tenha média zero.
θ_{it}	Variância de $\varepsilon_{i.t}$.
D	Fator de escala. Utiliza-se o valor 1,702 de forma que a função logística forneça resultados semelhantes aos da função ogiva normal.

Agradecimentos

Não sou uma exceção e desconfio que muitos saíram mais humildes diante do desafio intelectual, físico e emocional associado à obtenção de um doutorado no IME/USP. Seria tolice não admitir que se o sucesso foi alcançado foi porque sempre estive cercado de pessoas generosas e dispostas a compartilhar comigo os custos desta jornada. Precisei e, por vezes, abusei de muita ajuda para conseguir chegar até aqui. Poderia até tentar, mas não vou conseguir expressar melhor que Gonzaguinha¹ esse sentimento de dependência e gratidão que sinto tão fortemente.

[...]

E aprendi que se depende sempre
De tanta, muita, diferente gente
Toda pessoa sempre é as marcas
Das lições diárias de outras tantas pessoas

E é tão bonito quando a gente entende
Que a gente é tanta gente onde quer que a gente vá
E é tão bonito quando a gente sente
Que nunca está sozinho, por mais que pense estar

[...]

Com esse coração, gostaria de, primeiramente, expressar minha sincera gratidão à minha amada esposa, Ana Paula, e ao meu filho, Eduardo, por toda a paciência e compreensão que tiveram comigo durante as horas em que precisei me ausentar para me dedicar a esse projeto. O apoio, amor e a compreensão incondicional de vocês foram fundamentais para conseguir terminá-lo. Foram muitos sacrifícios e renúncias! Muito obrigado!! Amo vocês!

Tive o privilégio e honra de ter não uma, mas duas fabulosas orientadoras. Não posso ser mais grato à Professora Júlia Pavan e à Professora Lúcia Barroso pelo sincero apoio que me foi concedido nestes longos sete anos. O conhecimento, paciência e dedicação das senhoras foram fundamentais para meu crescimento acadêmico e pessoal. Suas palavras de incentivo e olhar crítico

¹ Trecho de Caminhos do Coração. 1982.

me permitiram evoluir e superar meus limites. Foi um privilégio ter a oportunidade de aprender com pessoas tão dedicadas e experientes quanto as senhoras.

Gostaria de expressar minha sincera gratidão ao professor Heliton Tavares por ter me apresentado a Teoria da Resposta ao Item (TRI) em 2010. Graças à sua paixão e entusiasmo pela área, pude descobrir um novo mundo de possibilidades e oportunidades de pesquisa que me levaram a buscar um aprofundamento maior nesse campo.

Não posso esquecer meus sócios, Clecio Lima e Carlos Aquiles Siqueira por acreditarem no valor prático e comercial do meu doutorado e por incondicionalmente me apoiarem ao longo desta empreitada. Essa parceria sólida e comprometida é uma grande prova do nossa força e um sinal de que temos muitas oportunidades emocionantes pela frente.

Também tirei muita energia de todos os membros dos times da Adaptativa Inteligência Educacional e da Anima Educação. Sou grato de todo o coração por todo o apoio, orientação e incentivo que me proporcionaram ao longo desta jornada. Sempre estive imerso em uma cultura de colaboração e excelência, que me inspirou a dar o meu melhor todos os dias. Sem a flexibilidade que encontrei nesses ambientes, certamente este projeto teria outro final não tanto feliz.

Sinto-me profundamente grato a todos os professores do Departamento de Estatística do IME-USP por todo o apoio, ensinamentos e conselhos que me proporcionaram durante minha trajetória acadêmica. Graças à sua dedicação e comprometimento, pude adquirir um conhecimento sólido e consistente na área de estatística e avançar em minha carreira profissional. Agradeço também a todos os funcionários do departamento, que sempre estiveram disponíveis para ajudar e orientar quando precisei, e que contribuem diariamente para o funcionamento eficiente do departamento.

Tenho enorme gratidão aos professores Dalton Francisco de Andrade, Jorge Luis Bazán Guzmán, e Caio Lucidius Naberezny Azevedo por aceitarem o convite para compor minha banca de avaliação.

Não consigo expressar a contento minha profunda gratidão à minha mãe, Maria das Neves, meus irmãos de sangue, Henry, Harumi e Rodolfo, meus irmãos de papel, Christian, Dedeia e Amanda, minhas sogras, Wania e Edna, meus tios e tias, primos e primas, à toda família Harada, família Leal e família Azevedo e agregados, pelo amor, apoio e acolhimento que me concederam em nossa convivência. Desde os primeiros dias até hoje, cada um de vocês contribuiu para o meu desenvolvimento e sucesso de uma forma única e significativa. Vocês são minha fonte de força e inspiração, e sou imensamente grato por ter uma família tão amorosa e dedicada como a minha.

E por fim, como não ficar agradecido por toda essa energia cósmica positiva que nos cerca? Ela sempre me convence que tudo sempre acaba bem e que, quando acaba mal, é porque ainda não acabou. E vamo que vamo! Ainda há muito por acabar ...

Hugo Harada
Fevereiro/2023

Capítulo 1 Introdução

Not everything that can be counted counts and
not everything that counts can be counted.

William Bruce Cameron (1963)

1.1 Motivação

A relevância da Psicometria aplicada a testes de larga escala em Educação aumentou drasticamente no Brasil nos últimos trinta anos dada a crescente preocupação em se obter métricas quantitativas para se comparar diferentes sistemas de ensino e metodologias de aprendizado aplicadas em todo o território nacional [1]. O desempenho favorável de instituições de ensino em exames como a Prova Brasil, o Exame Nacional do Ensino Médio (ENEM) e Exame Nacional de Desempenho de Estudantes (ENADE), se tornou fator determinante e de referência na reputação das mesmas, na facilidade de obtenção de recursos públicos e na captura de alunos. Nesse contexto de construção de métricas, a Psicometria tenta estabelecer correspondências entre as observações ou manifestações do traço latente de interesse e suas posições em uma escala numérica relacionada a este. No caso de uma prova de linguagens do ENEM, por exemplo, o traço latente seria a habilidade em linguagens do indivíduo e as manifestações dessa habilidade seriam o padrão de respostas desse aluno aos itens de uma prova de linguagens. Quanto maior o número de acertos e a consistência do padrão de respostas de um aluno nessa prova, maior seria a habilidade do aluno em linguagens.

As primeiras tentativas de aferir traços latentes de indivíduos na Psicometria foram feitas por Spearman [2]–[4] no início do século XX com a introdução da definição de escores empíricos (T), escores verdadeiros (V), seus erros (E) e sua relação na clássica equação:

$$T = V + E . \quad (1.1)$$

O modelo de Spearman serviu de alicerce para o que hoje se chama de Teoria Clássica dos Testes (TCT). A TCT infere uma maior ou menor aptidão de um indivíduo unicamente através da quantidade de itens corretamente respondidos por este a um teste (seu escore). Quanto aos itens, toda a informação sobre um específico item é correlacional e depende de como este item se relaciona aos outros itens que compõem o teste. Com isso, parâmetros clássicos dos itens, como a discriminação e dificuldade, ficam invariavelmente dependentes da amostra ou do grupo de respondentes ao teste. Além disso, talvez a maior crítica à TCT seja que as avaliações dos indivíduos dependem dos testes aplicados e não são comparáveis entre testes a não ser que estes sejam paralelos, o que, na prática, é muito difícil de se construir. O fato de toda a mensuração se basear unicamente no escore final do teste e minimizar a contribuição de seus componentes estruturais, os itens, no parecer final não soa para muitos como uma ideia intuitivamente aceitável.

Em contrapartida, a Teoria da Resposta ao Item (TRI) foi proposta e seus primórdios podem ser rastreados nos trabalhos de Binet e Simon [5], [6] na França, nos quais ele tabula a percentagem de

acertos de um item em função da idade do indivíduo, e de Terman [7] cujo livro contém os primeiros gráficos do que hoje se conhece por Curva Característica do Item. A formalização matemática dessa teoria se consolida com os trabalhos de Lawley em 1943 [8], nos quais ele obtém os parâmetros de um item através de métodos de máxima verossimilhança e por Lord [9] que demonstrou que muitas das construções da TCT podem ser escritas como funções dos parâmetros dos itens de um teste.

Dentre as vantagens da TRI sobre a TCT estão as enumeradas abaixo citadas por Hambleton e Jones [10]:

- a independência dos parâmetros dos itens da amostra dos indivíduos avaliados;
- a independência da mensuração dos traços latentes dos testes utilizados;
- a independência da característica de um item das características dos demais itens do teste;
- o fato de oferecer um nível de precisão para cada valor do traço latente;
- o fato de o modelo não exigir testes paralelos para avaliar a fidedignidade;
- o fato de oferecer suporte para variáveis latentes multidimensionais.

Apesar de suas inúmeras vantagens frente à TCT, a TRI somente se consolidou como metodologia de avaliação a partir dos anos 80 quando *softwares* e *hardwares* capazes de lidar com a complexidade de seus modelos estatísticos propostos apareceram no mercado e se tornaram financeiramente acessíveis. Com sua popularização, já são raros os testes de larga escala que não utilizam a TRI. Em diversas aplicações, entender como os parâmetros da distribuição do traço latente variam com o tempo é uma pergunta de grande interesse prático. Um exemplo na área educacional seria avaliar a efetividade de uma interferência pedagógica no aumento das habilidades dos alunos no entendimento e uso da Língua Portuguesa, comparando a aptidão dos alunos nessas habilidades nos momentos pré- e pós-intervenção.

A estimação dos parâmetros da distribuição do traço latente para uma única população (de um único grupo) foi primeiramente abordada por Andersen [11] assumindo como conhecidos os parâmetros dos itens que seguem o modelo Rasch. Sanathanan and Blumenthal [12], De Leeuw e Verhelst [13] e Lindsay [14] expandiram esse trabalho possibilitando a estimação simultânea dos parâmetros dos itens e da distribuição do traço latente também adotando o modelo Rasch. Mislevy [15] propôs um método de estimação que não requer que todos os indivíduos respondam a todos os itens e que também não se restringe ao modelo Rasch. Bock e Zimowski [16] aplicaram os

conceitos desenvolvidos por Mislevy em cenários em que indivíduos de diferentes populações (grupos múltiplos) respondem ao mesmo teste ou a testes com itens comuns.

Na TRI, análises com grupos múltiplos podem ser utilizadas tanto transversalmente para comparar dois grupos com medidas tomadas no mesmo momento, quanto longitudinalmente para comparar o mesmo grupo de indivíduos com medidas tomadas em momentos diferentes. Neste último caso, é de interesse saber se os parâmetros populacionais do grupo inicial mudaram o suficiente para se considerar que este grupo agora pertence a uma nova população com parâmetros populacionais estatisticamente distintos. É importante ressaltar a importância de termos as medidas realizadas em diferentes testes e momentos colocadas em uma única escala para efeitos de comparação. Kolen e Brennan [17] discutem diferentes métodos de equalização dos escores de diferentes testes utilizando os modelos da TRI que se baseiam na presença de itens comuns nos diferentes testes aplicados. Exemplos práticos de aplicações dessas técnicas podem ser encontrados em Andrade et al. [18], [19]. Uma alternativa ao uso de modelos de grupos múltiplos para análises longitudinais é o uso de modelos unidimensionais da TRI para dados longitudinais como estudados por Tavares [20] e Azevedo [21], que incorporam estruturas de covariância entre os traços latentes aferidos em diferentes momentos.

Extensões dos modelos unidimensionais de 1-, 2- e 3- parâmetros foram derivadas para a modelagem multidimensional transversal [22]. Esses modelos têm natureza compensatória, ou seja, um aumento de valor em um traço latente é compensado por uma diminuição em outro, o que limita a aplicabilidade desses modelos em cenários em que esse tipo de relação não se sustenta. Simpson [23] propôs modelos parcialmente compensatórios em que um aumento em um traço latente não necessariamente acarreta diminuição em outro, esse pode aumentar junto ou ficar constante. A literatura comparando esses modelos é escassa e não há consenso sobre quando um é superior ao outro [24]. Nesta tese, esses modelos não serão utilizados na modelagem da multidimensionalidade transversal.

Os modelos de equações estruturais ou a teoria dos modelos de equações estruturais (MEE) são tradicionalmente utilizados na pesquisa da dimensionalidade dos fatores e, caso múltiplos existam, no estudo sobre a estrutura latente que melhor explica a interação existente entre eles, enquanto a TRI explora em mais detalhe a interação entre as respostas dos itens e o traço latente associado [25], [26]. Evidência disso é que na TRI existem pelo menos seis diferentes maneiras de se modelar essa interação, enquanto nos MEE somente os modelos mais simples (modelos de um ou dois parâmetros e o modelo Ogiva Normal) são utilizados. Nos MEE, a multidimensionalidade transversal é modelada sem nenhuma restrição compensatória entre os traços latentes e o mesmo pode ser dito quanto à multidimensionalidade longitudinal. Esta tese explora o potencial de uso

dos MEE em análises longitudinais tradicionalmente feitas com modelos unidimensionais da TRI ligados longitudinalmente via uma estrutura de grupos múltiplos. Esta última metodologia não é livre de críticas pois estudos sobre a multidimensionalidade dos dados utilizando os microdados do INEP, por exemplo, sugerem que um modelo que suporte dois traços latentes - um associado ao raciocínio lógico e outro à leitura e interpretação de textos - seja o mais apropriado para a avaliação dos alunos [27]. Essa estrutura de dados induz à formulação de uma análise de traços latentes dos estudantes que leve em consideração (a) a evidente correlação presente entre os distintos traços latentes de um mesmo indivíduo e (b) a correlação existente entre os traços avaliados em diferentes momentos. Assim, no contexto da TRI e de MEE, esta tese aborda a multidimensionalidade transversal e longitudinal dos traços latentes em um único modelo que permite uma análise acurada do comportamento desses traços.

A maneira mais óbvia de se garantir que duas medidas são comparáveis, que estão na mesma escala, é utilizar o mesmo instrumento de mensuração. Em análises com medidas repetidas, dados de uma mesma unidade amostral são coletados com um mesmo instrumento sob diferentes condições de avaliação como diferentes momentos, tipos de solo, ou medicamentos com o intuito de avaliar o comportamento desses dados frente ao efeito dos fatores. Quando tais condições não podem ser aleatorizadas entre si, tais experimentos são ditos longitudinais [28, p. 67]. É bastante comum estudos que utilizam “modelos de equações estruturais longitudinais” com o uso de medidas repetidas [29], [30]. O uso de medidas repetidas em contextos de avaliação educacional pode ser problemático, pois pode induzir vieses nas respostas coletadas de acordo com o objetivo do instrumento utilizado na coleta de dados. O respondente, ao ser exposto aos mesmos itens múltiplas vezes e, portanto, conhecê-los, pode calibrar suas respostas para obter um resultado socialmente desejado ou aceitável. Por exemplo, se o objetivo do instrumento for o diagnóstico de uma doença ou um teste de autoconhecimento, o indivíduo pouco se beneficia ao utilizar seu conhecimento prévio dos itens para direcionar o resultado. Por outro lado, o mesmo já não pode ser dito quando o objetivo for selecionar os ingressantes de um curso de graduação, pois nesses contextos competitivos, os indivíduos invariavelmente utilizarão de todos os seus conhecimentos sobre os itens repetidos para obter uma melhor colocação. Assim, existem situações em que é necessário utilizarmos diferentes itens em momentos diferentes na MEE, assim como é corriqueiramente feito na TRI. Nesta tese, também abordaremos como a calibração simultânea de itens é feita na MEE.

Em termos transversais, a comparabilidade entre os modelos da TRI e da MEE está bem estabelecida. Takane e De Leeuw [31] provaram a equivalência entre modelos da TRI e MEE para variáveis dicotômicas baseando-se em trabalhos de Muthén [32], [33] e David John [34]. Glockner-

Rist e Hoijtink [35] mostraram que ambas as teorias podem ser utilizadas na verificação da dimensionalidade de dados e em testes de invariância da medida. Por fim, Kamata e Bauer [36] resumiram as diferentes opções de parametrizações envolvidas na determinação das escalas das variáveis latentes e como essas parametrizações resultam em diferentes equações de equivalência entre a TRI e MEE.

É um equívoco comum assumir que a TRI se limita a problemas com muitos itens e poucas variáveis latentes e a MEE a problemas com poucos itens e muitas variáveis latentes. A causa disso provavelmente reside no fato de que os pacotes computacionais disponíveis no mercado refletem essas características. No entanto, todos os modelos discutidos nesta tese, a contar por restrições de identificabilidade, são facilmente estendidos para permitir qualquer número de itens e variáveis latentes bem como tamanho amostral. A grande limitação no desenvolvimento de *softwares* não está na definição dos modelos, mas sim na dificuldade de estimação dos parâmetros desses modelos quando eles têm muitos itens e muitas variáveis latentes. Wirth e Edwards [37] relatam que a performance e convergência dos estimadores varia de acordo com número de itens, número de variáveis latentes, número de dados faltantes, tamanho da amostra, número de categorias em respostas politômicas e valores iniciais; e recomendam a utilização de múltiplos estimadores para obter estimativas sob diferentes metodologias no processo de elucidação do modelo e na validação das estimativas.

1.2 Objetivos

O objetivo deste trabalho de tese é comparar a aplicação de modelos multidimensionais longitudinais utilizando duas distintas abordagens na construção de variáveis latentes em um contexto educacional. Multidimensional porque admite a possibilidade de que diferentes traços latentes podem interagir positivamente ou negativamente na definição de um perfil latente “geral” do indivíduo em um determinado momento, e longitudinal, porque o passado e o futuro andam juntos, e é inquestionável conhecer as tendências dessa relação. Para isso, far-se-á uso de técnicas da MEE em conjunto com métodos da TRI, esclarecer-se-á as equivalências entre suas formulações, seus benefícios e suas inconveniências, e assim ganhar-se-á flexibilidade na modelagem desse contexto combinando técnicas de áreas independentes e complementares da estatística.

1.3 Contribuições esperadas

Nesse contexto, inicialmente, pode-se listar as seguintes contribuições esperadas deste trabalho:

- Em análises transversais, propor modelos multidimensionais de correlação entre as proficiências de um indivíduo nas diferentes áreas do conhecimento e entender como essas proficiências podem ser agrupadas para medir proficiência “geral” do estudante;
- Estender a metodologia para o caso longitudinal em que são introduzidas correlações entre as proficiências avaliadas nos diferentes momentos. Esses modelos permitirão avaliar a evolução do estudante ao longo do ano nas diferentes áreas;
- Tanto na abordagem de cada corte transversal como no caso longitudinal, explorar equivalências entre as formulações do modelo via MEE e TRI no campo multidimensional, propondo assim metodologias inferenciais de interesse que podem permitir modelagens mais flexíveis e considerar sua implementação computacional.

1.4 Organização do trabalho

Este trabalho está organizado em sete capítulos, a saber:

- No Capítulo 1 é feita uma introdução aos conceitos de TRI e MEE, são abordados os aspectos gerais, justificativa e importância do trabalho no âmbito dos objetivos gerais da tese.
- No Capítulo 2 é realizada a descrição conceitual do que são variáveis latentes, suas definições, propriedades e usos.
- No Capítulo 3 é realizada a descrição conceitual do método da TRI, apresentando o processo de estimação de cada parâmetro de item e predição do traço latente.
- No Capítulo 4 é realizada a descrição conceitual do método de MEE, mostrando seu contexto histórico e conceitos básicos.
- No Capítulo 5 é realizada a apresentação de resultados e análises de estudos de simulação.
- No Capítulo 6 é realizada a aplicação dos modelos propostos em dados reais coletados em avaliações educacionais no município de Cariacica, Espírito Santo, em 2019.

- No Capítulo 7 estão as considerações finais e sugestões para pesquisas futuras.

Capítulo 2 Variáveis latentes

It is impossible to date the first use of latent variables. The idea that observable phenomena are influenced by underlying and unobserved causes is at least as old as religion, where unseen forces affect real world events.

Kenneth A. Bollen

2.1 Introdução

A ideia contida na citação da página anterior [38] de que o intangível pode reger o tangível é tão admissível que, naturalmente, abstraímos as complexidades associadas à mensuração de conceitos familiares como quão “bem” acordei hoje de manhã, quão “perigoso” é um indivíduo condenado pela justiça ou quão “propenso” um cliente está de fechar uma compra. Esses exemplos ilustram que já faz parte do ferramental cognitivo humano utilizar medidas não diretamente observáveis para tomarmos decisões em nosso dia a dia. Variáveis não mensuráveis, fatores, variáveis não observáveis, construtos, escores verdadeiros são apenas alguns exemplos sobre como pesquisadores se referem a variáveis latentes, e essa diversidade de denominações talvez seja a principal causa do não reconhecimento da onipresença de variáveis latentes em disciplinas tão diversas quanto economia, psicologia, marketing e biologia.

Neste capítulo vamos discutir as definições formais e não formais de variáveis latentes, listar suas propriedades e comentar sobre a aplicabilidade dessas definições nos diferentes modelos estatísticos em que variáveis latentes são utilizadas.

2.2 Definições

É raro nas ciências exatas nos depararmos com um conceito que tenha tantas definições, formais e não formais, quanto se tem para variáveis latentes. Nesta seção revisaremos essas definições.

2.2.1 Variáveis observáveis ou indicadores

O conceito de variáveis latentes é indissociável do conceito de variáveis observáveis pois estas são utilizadas na definição ou predição das primeiras. As variáveis observáveis são aquelas cujos valores são obtidos ou observados por um mecanismo de mensuração objetivo, cujas medidas são aceitas como uma boa representação da realidade por seus usuários [39, p. 12]. São exemplos de variáveis observáveis em um contexto acadêmico: o número de questões que um aluno acertou em uma prova, sua frequência às aulas presenciais e sua participação em eventos estudantis; em um contexto médico: a pressão sanguínea, a porcentagem de gordura abdominal e o nível de colesterol de um paciente. No estudo de modelos latentes, as variáveis observáveis são tidas como manifestações (causas ou consequências) de conceitos não diretamente observáveis cuja mensuração se representa via variáveis latentes. Por exemplo, as medidas coletadas no exemplo acadêmico acima podem ser utilizadas na definição de uma medida de engajamento dos alunos no seu curso de graduação.

2.2.2 Definições não formais de variáveis latentes

Bollen [38] lista três diferentes definições informais para variáveis latentes.

Variáveis latentes podem ser vistas como artifícios de mensuração hipotéticos de características ou traços não observáveis, mas reais. São construções hipotéticas, construtos, que só existem dentro da imaginação de pesquisadores. Ninguém definiu esta visão melhor que Loewinger [40] quando disse que “Traços (características) existem nas pessoas; construtos existem nas mentes e nas revistas de psicólogos.”² Ou seja, assume-se que o traço latente existe, mas que variáveis latentes e seus construtos associados, não.

Em uma outra definição, variáveis latentes são vistas como não diretamente mensuráveis ou observáveis. O problema com esta definição é que a capacidade de se obter uma medida depende, em muitos casos, da tecnologia disponível no momento e, por isso, avanços tecnológicos podem tornar algo que antes era não observável em observável. Por exemplo, não muito tempo atrás o diagnóstico de morte encefálica de um paciente dependia da ponderação de sintomas observáveis do mesmo, hoje esse diagnóstico pode se dar de maneira bem mais objetiva através de máquinas de eletroencefalogramas que detectam não somente a intensidade da atividade elétrica no cérebro como também onde ela ocorre com mais ou menos frequência.

Ainda, variáveis latentes podem ser vistas como um mecanismo de redução de dimensionalidade de dados, uma maneira conveniente de resumir em poucas dimensões o que foi observado em uma coletânea de variáveis observáveis. Sob esta definição, podemos interpretar o construto *capacidade cognitiva de um aluno*, como uma representação agregada que nos permite interpretar de maneira resumida o padrão de respostas de um aluno a um teste.

Essas definições informais são úteis no entendimento conceitual de como variáveis latentes se relacionam com variáveis observáveis, mas não detalham com rigor necessário suas propriedades e suposições técnicas que permitam que elas sejam utilizadas em modelos estatísticos. A próxima seção aborda definições formais de variáveis latentes.

² “Traits exist in people; constructs exist in the minds and magazines of psychologists.”

2.2.3 Definições formais de variáveis latentes

2.2.3.1 Definição da independência local

A ideia principal por trás dessa definição é que uma variável latente é responsável pela associação que existe entre as variáveis observáveis [9][41], e disto decorre que se a variável latente é controlada e fixada a valor fixo, as variáveis observáveis, condicionadas nesse valor, são independentes, ou seja³:

$$P(Y_{1t} \cap Y_{2t} \cap \dots Y_{(I-1)t} \cap Y_{It} | \eta) = P(Y_{1t}|\eta)P(Y_{2t}|\eta) \dots P(Y_{(I-1)t}|\eta)P(Y_{It}|\eta) = \prod_{i=1}^I P(Y_{it}|\eta). \quad (2.1)$$

em que, Y_{it} é variável aleatória associada a uma amostra contendo observações de y_{iht} definida como a resposta do indivíduo h no tempo t ao item i com h variando de 1 a H , e η é a variável latente comum ou fator comum. Em outras palavras, a dependência que existe entre as variáveis aleatórias observáveis Y_{it} é explicada unicamente pela associação dessas com a variável latente comum η . A definição acima pode ser expandida para o caso em que variáveis observáveis dependem de mais de uma variável latente.

Essa definição assume que (a) os erros de mensuração são independentes ou não correlacionados, (b) não existem efeitos diretos ou indiretos entre as variáveis observáveis, (c) pelo menos duas variáveis observáveis são necessárias, (d) a variável latente deve ter efeitos diretos em mais de uma variável observável, e (e) as variáveis observáveis não afetam diretamente a variável latente [38].

2.2.3.2 Definição do valor esperado

Variáveis latentes podem representar o valor esperado de uma variável observável medida com erro [42], ou seja, do modelo de mensuração da teoria clássica dos testes [43], temos que, y_{mh} , valor observado de uma característica y do indivíduo h em uma tentativa de mensuração ou experimento m é igual à soma de seu escore verdadeiro v_h com o erro de mensuração e_{mh} , tal que:

$$y_{mh} = v_h + e_{mh} \quad (2.2)$$

Assumindo y_{mh} e e_{mh} são valores observados das variáveis aleatórias Y_h e E_h , conclui-se que, a partir da definição em (2.2), o escore verdadeiro é calculado como a média de M medições da

³ Dois eventos A e B são independentes se $P(A \cap B) = P(A)P(B)$. Ver DeGroot. [67]

valores observados y_{mh} , $m = 1, \dots, M$, assumindo-se que (a) a média dos erros de mensuração é zero, (b) que os erros de mensuração e o escore verdadeiro não são correlacionados e (c) que os experimentos são independentes, ou seja,

$$\begin{aligned} E(Y_h) &= v_h + E(E_h) \\ E(Y_h) &= v_h \end{aligned} \tag{2.3}$$

Além das propriedades anteriormente citadas, outras consequências dessa definição são: (a) a escala da variável latente é função da escala das variáveis observadas, (b) o escore verdadeiro tem efeito direto nas variáveis observáveis enquanto o oposto não acontece, e (c) entre duas variáveis observáveis, uma não exerce qualquer efeito, direto ou indireto, na outra [38].

2.2.3.3 Definição como uma função não determinística de variáveis observáveis

Bentler [44] define que uma variável latente não pode ser determinada unicamente a partir das variáveis observáveis relacionadas. A maior restrição desta definição é que sua aplicação se limita a sistemas lineares de equações estruturais e todos os modelos de variáveis latentes que tratam com variáveis observáveis categóricas são não lineares.

2.2.3.4 Definição via realização amostral⁴

Segundo esta definição, proposta por Bollen em [38], uma variável é latente se, na realização de uma amostra dessa variável, não há dados para algumas observações dessa amostra. Esta definição está em linha com a definição informal de que variáveis latentes são aquelas para as quais não se pode levantar dados. E ainda, como a definição está atrelada a uma amostra, uma variável pode ser latente em uma amostra e observável em outra. Desta maneira, todas as variáveis são latentes até que valores amostrais sejam obtidos para as mesmas. Esta definição permite modelos com erros de mensuração correlacionados e variáveis observáveis que influenciam umas às outras de maneira direta ou indireta. Bartholomew, Knott e Moustaki também sugeriram que variáveis latentes são variáveis sobre as quais todas as observações não existem [41, p. 245].

2.3 Propriedades de variáveis latentes

Nesta seção discutiremos algumas propriedades das variáveis latentes que influenciam o tipo de modelo a ser utilizado e restrições que impomos a eles.

⁴ Sample realization definition.

2.3.1 Tipos de variáveis latentes

Variáveis latentes podem ser contínuas ou categóricas e esta definição não está relacionada ao tipo de variáveis observáveis associadas, pois pode-se ter variáveis latentes contínuas associadas a variáveis observáveis categóricas e vice-versa. É apropriado sempre levar em consideração que tipo de gradação de valores a teoria assume como aceitável para uma variável latente. Não há resposta certa para essa pergunta se devemos considerar, por exemplo, a variável latente estresse contínua, categórica com 5 categorias ou dicotômica, mas esta escolha tem implicações severas nos tipos de modelos disponíveis para análise. Os métodos da TRI e dos MEE assumem variáveis latentes contínuas, enquanto modelos de classes latentes assumem variáveis latentes categóricas [29].

2.3.2 Identificabilidade

O problema da identificabilidade tenta responder se é possível encontrar soluções únicas para os parâmetros envolvidos em um modelo. A não identificabilidade põe em xeque a validade dos valores dos parâmetros mesmo para amostras expressivas da população. Determinar a identificabilidade de um modelo não é uma pergunta empírica e pode ser provada analiticamente, mas é desafio quase nunca trivial [39]. Na prática, a grande maioria dos pesquisadores segue um manual de heurísticas ou regras que geralmente envolvem configurar um número mínimo de variáveis observáveis por variável latente ou impor restrições nas variâncias envolvidas. Ver Kline [39] e Bollen [45] para detalhes.

2.3.3 Indeterminação fatorial

Segundo Mc Donald [46], “indeterminação fatorial é a inability de se determinar unicamente os fatores específicos e comuns do modelo de fator comum a partir de variáveis observáveis determinadas unicamente, porque o número de variáveis observáveis é menor do que o número de fatores comuns e específicos.” Segundo Bollen [38], resolver essa indeterminação é teoricamente possível sob algumas condições raramente plausíveis na prática: (a) tamanho da amostra infinito, (b) número de variáveis observáveis infinito e (c) quando o quadrado da correlação múltipla ⁵ da variável latente tende para 1 e os preditores são as variáveis latentes. A Figura 2.1 ilustra o item (b) desta propriedade ao mostrar a correlação entre os valores dos traços latentes simulados e preditos como função do número de variáveis observáveis ou itens. É evidente que a correlação tende para

⁵ Squared multiple correlation

1 somente para números grandes de itens. A figura evidencia também que o mesmo comportamento é observado independentemente do tipo de modelagem utilizado: TRI ou MEE. O Apêndice B contém detalhes sobre a condução deste experimento.

A Figura 2.2 compara a distribuição dos escores simulados (SIM) contra os preditos obtidos em diferentes condições de modelagem: (i) modelagem TRI com estimação dos fatores utilizando o método *Expected a Posteriori* – EAP_TRI, (ii) modelagem MEE com estimação *Empirical Bayes Modal* – EBM_MEE e (iii) modelagem MEE com *Maximum Likelihood* – ML_MEE). Claramente, os métodos são mais precisos na estimação da média do que na estimação da variância apesar da extremamente alta correlação, acima de 0,99, entre os valores preditos sob diferentes condições.

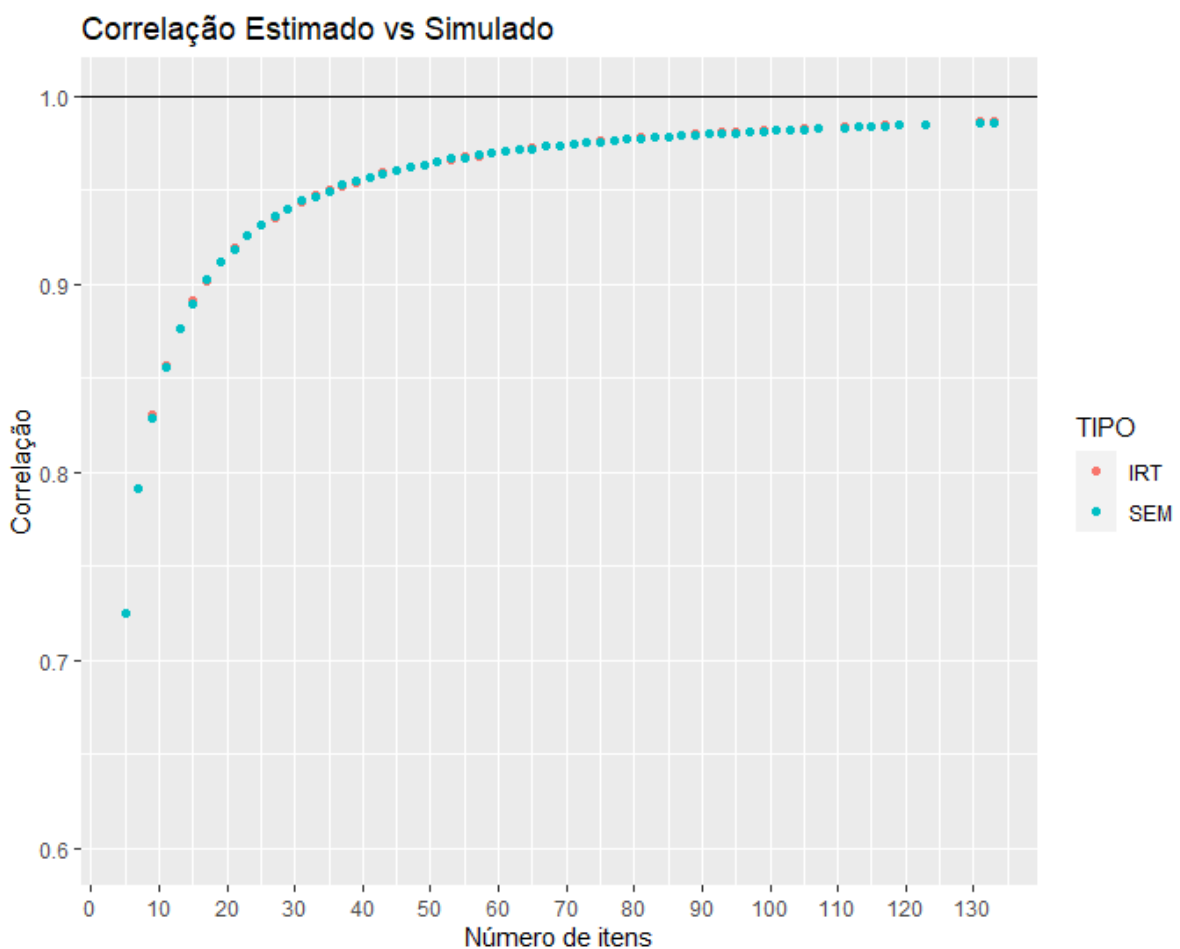


Figura 2.1 Exemplo de indeterminação fatorial

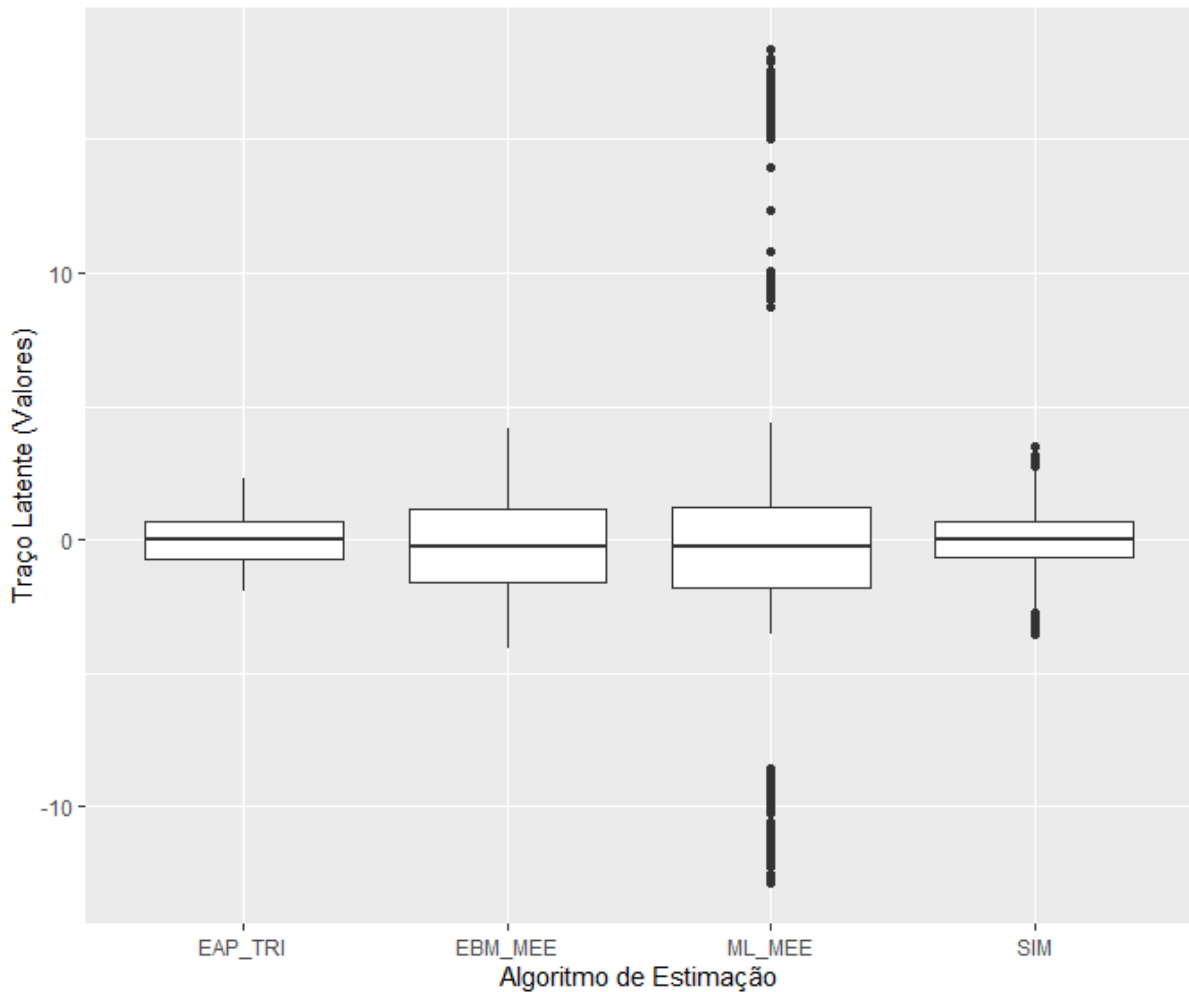


Figura 2.2 Distribuição dos escores fatoriais preditos por diferentes métodos. SIM = valores simulados

2.3.4 Variáveis observáveis como indicadores de causa ou efeito

Variáveis observáveis podem ter uma relação de causa ou efeito com as variáveis latentes associadas. Indicadores de causa, ou formativos, afetam diretamente os valores da variável latente, enquanto que indicadores de efeito, ou reflexivos, têm seus valores alterados a partir de alterações nos valores das variáveis latentes. Sanches [47] demonstrou que identificar incorretamente o tipo de um indicador pode levar a estimações errôneas dos parâmetros e a conclusões equivocadas a partir de um modelo mal especificado.

2.4 Notação

Nesta seção é apresentada a notação utilizada na especificação dos modelos, TRI ou MEE, tratados neste trabalho.

A resposta a um item pode ser interpretada como uma manifestação de traços latentes estimulados pelo item em questão. Seja Y_{iht} uma variável aleatória binária que representa a resposta do indivíduo h ao item i no momento t , tal que,

$$Y_{iht} = \begin{cases} 1, \text{resposta correta,} \\ 0, \text{resposta incorreta.} \end{cases} \quad (2.4)$$

Seja também, η_{hkt} , valor observado do k -ésimo traço latente do indivíduo h no momento t , e ζ_{it} , o vetor de parâmetros que definem a função resposta do item i no momento t , tal que $(Y_{iht} | \eta_{hkt}, \zeta_{it})$ é variável aleatória de Bernoulli com parâmetro P_{ihkt} , $i = 1, 2, \dots, I_t$, $h = 1, 2, \dots, H$, $k = 1, 2, \dots, K$, $t = 1, 2, \dots, T$. Temos que,

$$P_{ihkt} = P(Y_{iht} = 1 | \eta_{hkt}, \zeta_{it}), \quad (2.5)$$

em que, P_{ihkt} é a probabilidade condicional de Y_{iht} ser 1 dado η_{hkt} e ζ_{it} . Nesse contexto, os vetores aleatórios relacionados as respostas se organizam da seguinte maneira: $\mathbf{Y}_{.ht} = (Y_{1ht}, Y_{2ht}, \dots, Y_{I_t ht})'$ é um vetor aleatório de comprimento I_t do indivíduo h no momento t , $\mathbf{Y}_{..t} = (\mathbf{Y}'_{.1t}, \mathbf{Y}'_{.2t}, \dots, \mathbf{Y}'_{.Ht})'$ é o vetor aleatório de comprimento HI_t de todos os indivíduos no momento t , e $\mathbf{Y} = \mathbf{Y}_{...} = (\mathbf{Y}'_{..1}, \mathbf{Y}'_{..2}, \dots, \mathbf{Y}'_{..T})'$ é vetor aleatório total com comprimento HTI , sendo $I = \sum_{t=1}^T I_t$. De forma similar, a representação das respectivas observações se dá por $\mathbf{y}_{.ht}$, $\mathbf{y}_{..t}$, e $\mathbf{y}_{...}$. O vetor $\zeta_{.t} = (\zeta'_{1t}, \zeta'_{2t}, \dots, \zeta'_{I_t t})'$ contém os parâmetros de todos os itens no momento t , e o vetor $\zeta = \zeta_{..} = (\zeta'_{.1}, \zeta'_{.2}, \dots, \zeta'_{.T})'$ contém os parâmetros de todos os I itens aplicados nos T momentos, $\boldsymbol{\eta}_{.kt} = (\eta_{1kt}, \eta_{2kt}, \dots, \eta_{Hkt})'$ é um vetor de dimensão H contendo valores do k -ésimo traço latente de H indivíduos no momento t associados à variável latente η_{kt} com densidade $f(\eta_{kt} | \boldsymbol{\pi}_{kt})$ definida pelo vetor $\boldsymbol{\pi}_{kt}$ de parâmetros. O vetor $\boldsymbol{\eta}_{.t} = (\eta_{1k}, \eta_{2k}, \dots, \eta_{Kt})'$ contém K variáveis latentes no momento t , e $\boldsymbol{\eta}_{KT} = \boldsymbol{\eta}_{..} = (\boldsymbol{\eta}'_{.1}, \boldsymbol{\eta}'_{.2}, \dots, \boldsymbol{\eta}'_{.T})'$ contém todas as variáveis associadas aos traços latentes nos T momentos, as quais se relacionam segundo uma normal multivariada KT -dimensional.

Em modelos de análise de fatores, se a resposta ao item é contínua, assume-se uma relação linear entre Y_{it} e η_{kt} sendo definida pela seguinte equação,

$$Y_{it} = \nu_{it} + \lambda_{it}\eta_{kt} + \varepsilon_{it}. \quad (2.6)$$

em que ν_{it} e λ_{it} são, respectivamente, o intercepto e a carga fatorial do item i no momento t . A variável aleatória ε_{it} define a distribuição dos erros de mensuração e possui densidade $f(\varepsilon_{it} | \boldsymbol{\pi}_{\varepsilon_{it}})$ definida pelo vetor $\boldsymbol{\pi}_{\varepsilon_{it}}$ de parâmetros. Note que o índice h foi abreviado da representação da variável Y_{iht} , pois o foco de interesse é representar a distribuição das respostas dos H indivíduos ao item it .

Se a resposta ao item for ordinal categórica, seus valores dependem dos valores de uma variável latente subliminar Y_{it}^* que se relaciona linearmente com η_{kt} como a seguir:

$$Y_{it}^* = \nu_{it} + \lambda_{it}\eta_{kt} + \varepsilon_{it}, \quad (2.7)$$

sendo que a relação entre Y_{iht} e Y_{it}^* no caso dicotômico se dá pelo modelo do valor limiar abaixo definido.

$$Y_{iht} = \begin{cases} 0 & \text{se } -\infty < Y_{it}^* \leq \tau_{it} \\ 1 & \text{se } \tau_{it} < Y_{it}^* < +\infty \end{cases}, \quad (2.8)$$

em que τ_{it} é o valor limiar ou de corte que divide a escala de Y_{it}^* em intervalos, cada qual associado a um valor categórico específico.

Em seguida, define-se a matriz de covariância populacional $\boldsymbol{\Sigma}(\mathbf{Y})$ e a matriz de covariância amostral $\mathbf{S}(\mathbf{Y})$ com dimensões $I \times I$ para um vetor \mathbf{Y} de variáveis aleatórias de comprimento I . Os respectivos elementos dessas matrizes $\sigma_{m,n}^2$ e $s_{m,n}^2$ localizados na linha m e na coluna n são definidos por,

$$\sigma_{m,n}^2 = \text{COV}(Y_m, Y_n), \text{ com } m, n = 1, \dots, I, \quad (2.9)$$

e

$$s_{m,n}^2 = \text{cov}(Y_m, Y_n), \text{ com } m, n = 1, \dots, I, \quad (2.10)$$

em que, a função $\text{COV}(X, Y)$, em letras maiúsculas, retorna a covariância populacional e a função $\text{cov}(X, Y)$, em letras minúsculas, retorna a covariância amostral entre duas variáveis aleatórias. Ambas as funções estão definidas no Apêndice C.

Com esta notação podemos representar modelos unidimensionais transversais quando $K = 1$ e $T = 1$ (Figura 2.3), modelos multidimensionais transversais quando $K > 1$ e $T = 1$ (Figura 2.4), modelos unidimensionais longitudinais quando $K = 1$ e $T > 1$ (Figura 2.5) e modelos multidimensionais longitudinais quando $K > 1$ e $T > 1$ (Figura 2.6). Neste documento, índices podem ser abreviados quando seus valores são fixos na unidade, o índice t é abreviado na notação de modelos transversais, o índice g , na notação de modelos com um único grupo e o índice k , na notação de modelos unidimensionais.

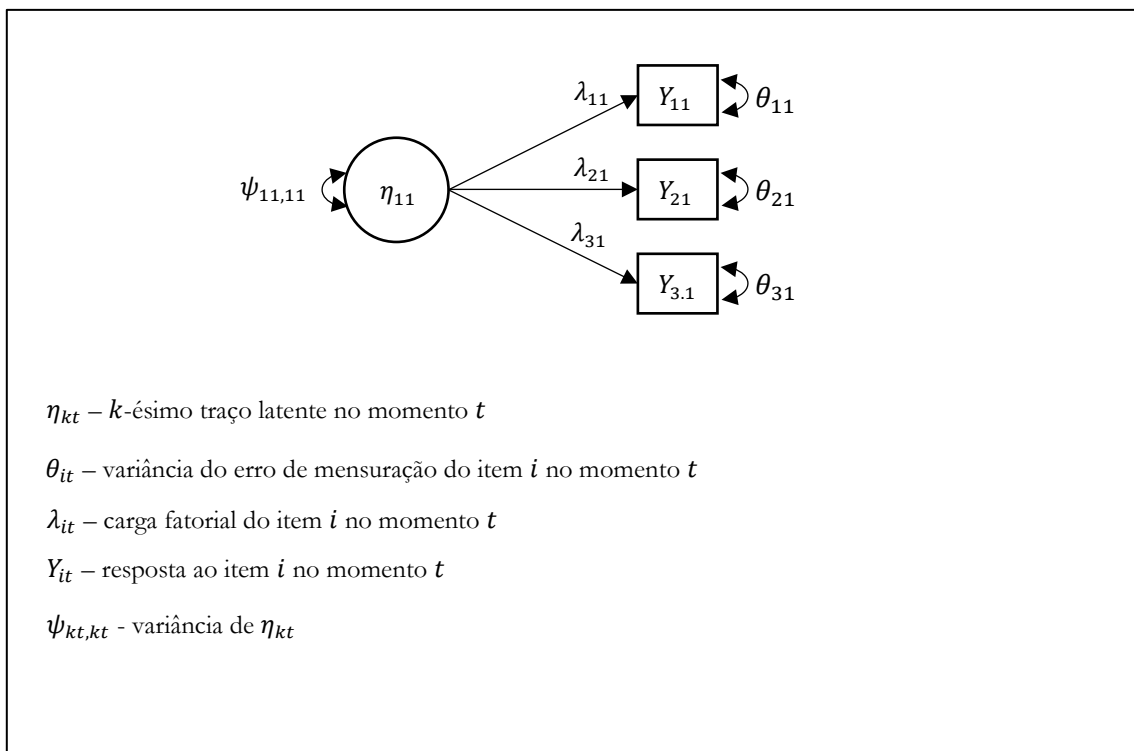


Figura 2.3 Modelo unidimensional transversal. $I=3, K=1, T=1$.

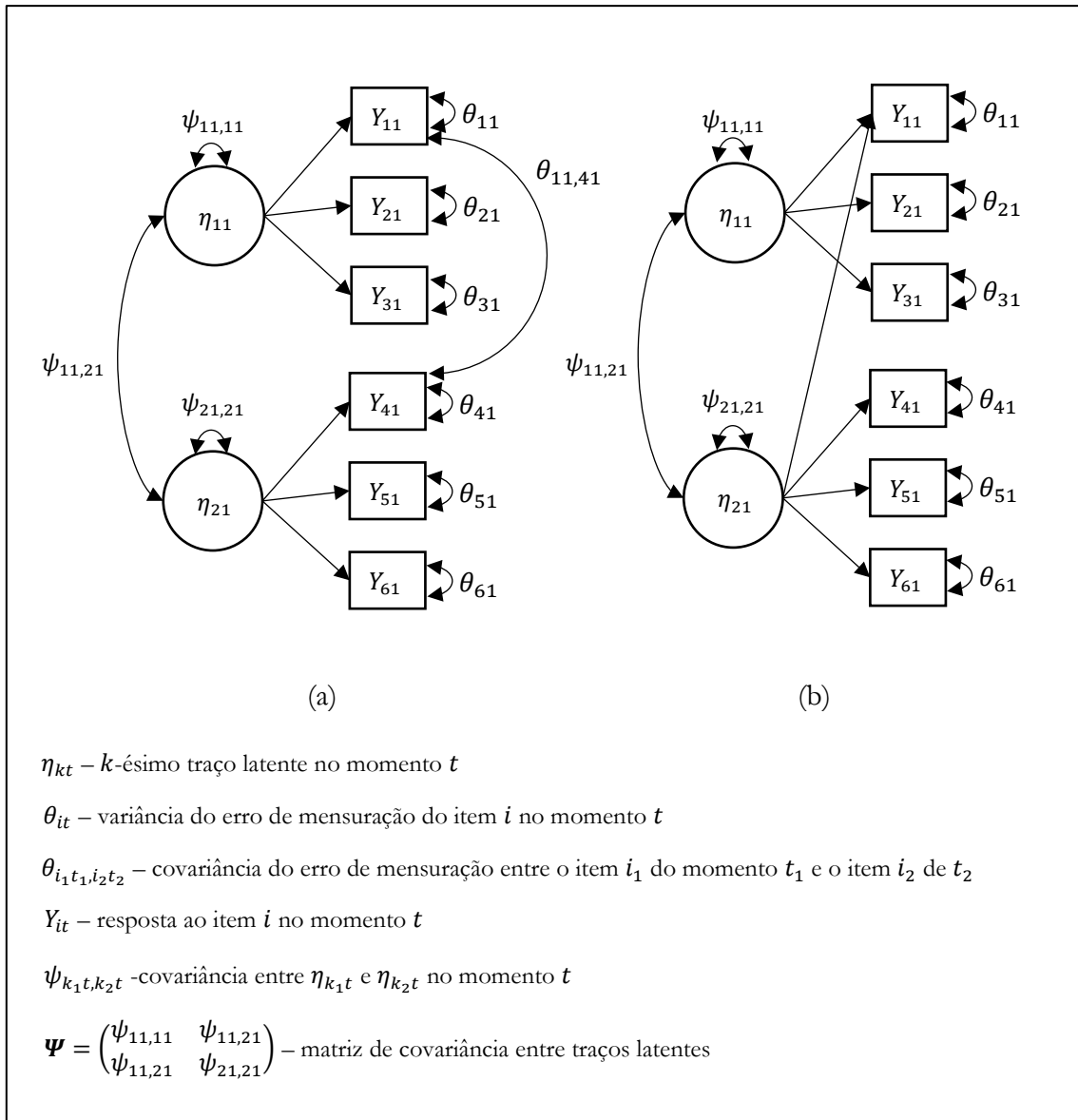


Figura 2.4 Modelo multidimensional transversal. I=6, K=2, T=1. Em (a) a resposta ao item 11 depende de um único traço latente enquanto que em (b) essa mesma resposta depende de múltiplos traços.

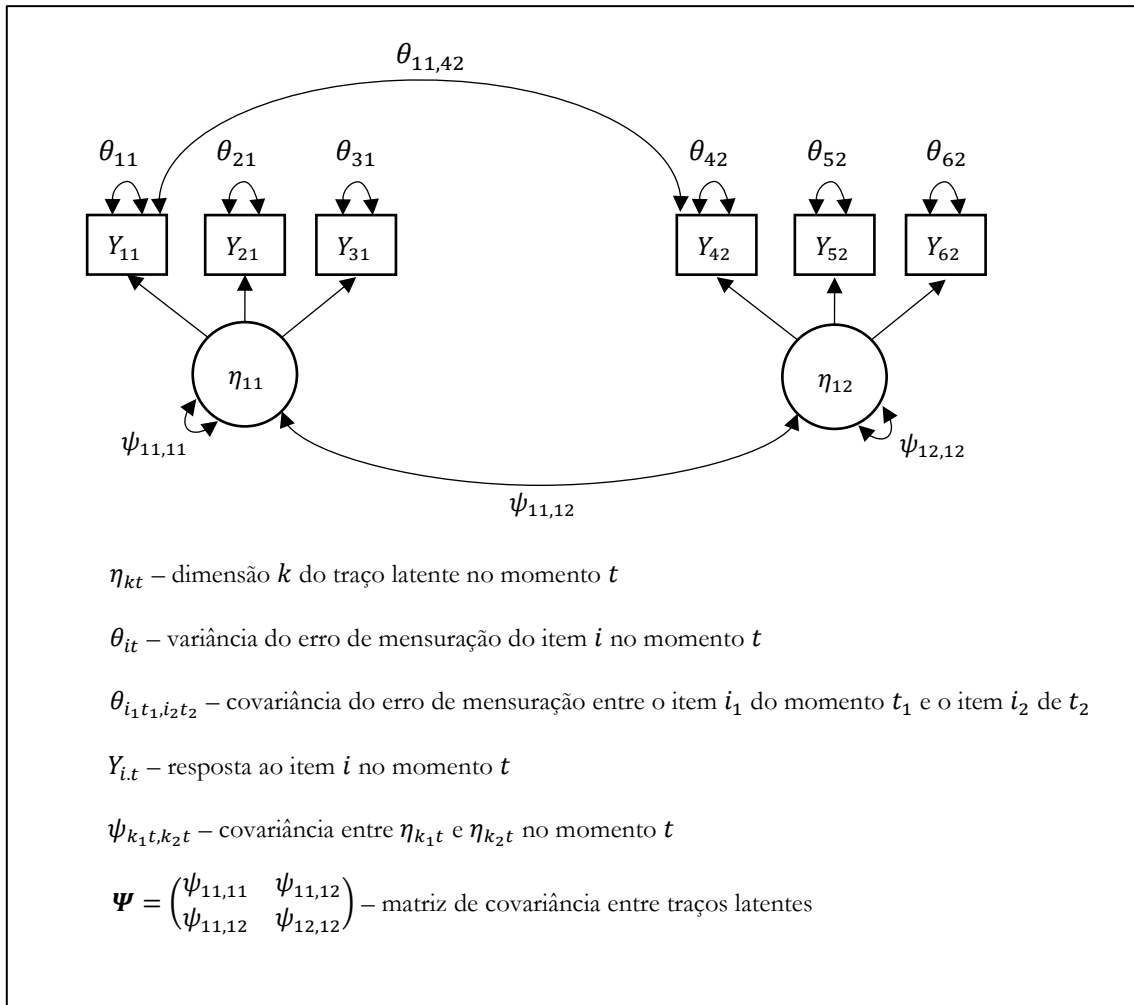


Figura 2.5 Modelo unidimensional longitudinal. I=6, K=1, T=2.

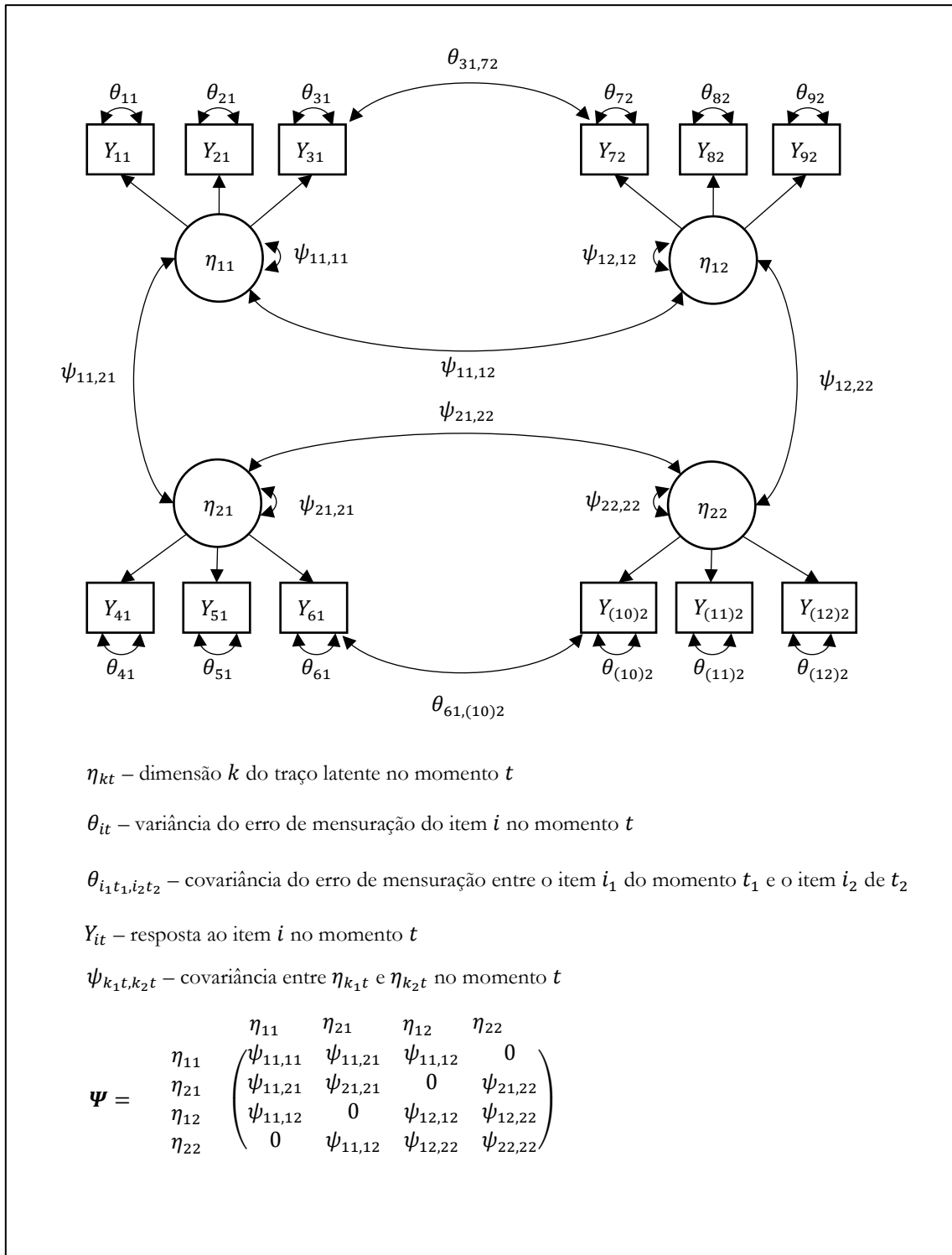


Figura 2.6 Modelo multidimensional longitudinal. I=12, K=2, T=2.

Capítulo 3 A teoria da resposta ao item

Traits exist in people; constructs (here usually about traits) exist in the minds and magazines of psychologists.

Jane Loevinger (1957)

3.1 Introdução

Este capítulo dedica-se a uma descrição conceitual dos modelos uni e multidimensionais da *Teoria da Resposta ao Item* (TRI). O foco está na caracterização do comportamento de um item quando este é apresentado a diversos indivíduos com diferentes habilidades considerando um único momento de avaliação dos alunos.

3.2 Curva característica do item

Segundo Pasquali [25], a TRI se desenvolveu com base em duas premissas fundamentais. A primeira se refere ao fato de que o nível da habilidade de um indivíduo ao responder a um item se explica por um ou mais traços latentes numa relação de causa e efeito na qual os traços latentes são a causa e os níveis de habilidades são efeitos. Nesse contexto, vale lembrar que o nível de habilidade é também uma variável latente. Já a segunda estabelece que essa relação de causa/efeito pode ser descrita numa equação monotônica crescente chamada de *Curva Característica do Item* (CCI) como mostra a Figura 3.1. Essa curva tem o formato de S cuja forma depende dos parâmetros do modelo adotado. Em geral as CCI possuem três intervalos distintos que merecem menção:

- Patamar de acerto de baixa habilidade: intervalo de habilidades no qual o item não consegue diferenciar indivíduos de baixa habilidade. Na Figura 3.1, este intervalo corresponde aos valores de habilidades menores que aproximadamente -3 .
- Intervalo de discriminação: intervalo de habilidade no qual o item consegue efetivamente diferenciar indivíduos. Na Figura 3.1, este intervalo corresponde aos valores de habilidades localizadas entre aproximadamente -3 e $+3$.
- Patamar de acerto de alta habilidade: intervalo de habilidades no qual o item não consegue diferenciar indivíduos de alta habilidade. Na Figura 3.1, este intervalo corresponde aos valores de habilidades maiores que aproximadamente $+3$.

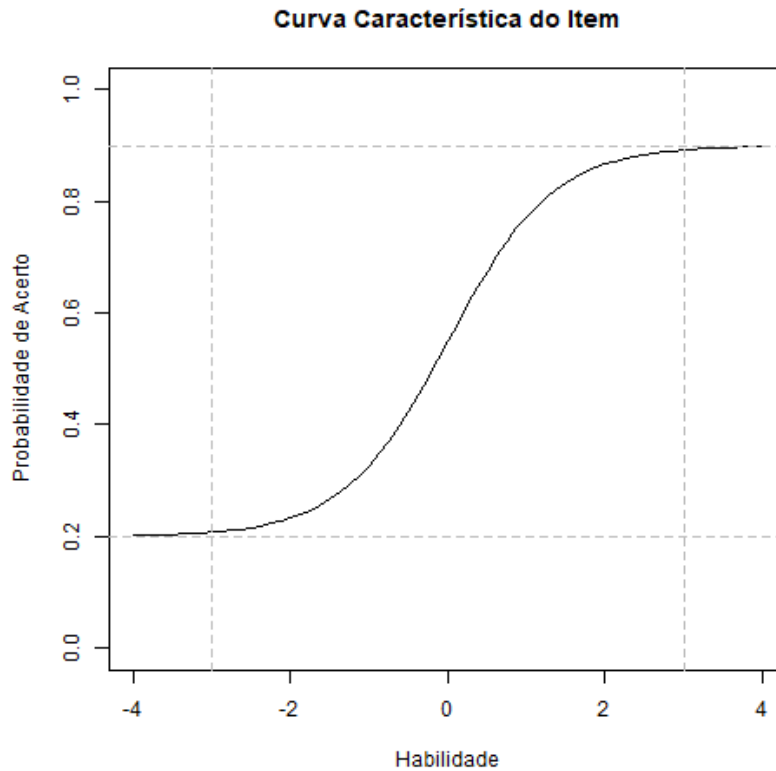


Figura 3.1 Exemplo de uma curva característica do item

3.3 Modelos unidimensionais transversais

Esta seção apresenta as formulações mais usuais de se modelar P_{ihkt} exposto na Equação (2.5), quando utilizado em modelos unidimensionais transversais ($K = 1$ e $T = 1$). A quantidade P_{ihkt} modela uma CCI unidimensional com características descritas na seção anterior e dentre os possíveis modelos que podem ser adotados, seis são de fundamental importância e serão detalhados em seguida.

3.3.1 Modelo ogiva normal

Dada a importância que a distribuição normal tem para a teoria estatística não é surpresa que a função distribuição acumulada normal seja utilizada na modelagem de CCI de dois parâmetros. Convenientemente, a CCI ou sigmoide de um item é dada pela equação da função distribuição acumulada de uma variável aleatória normal associada ao item i com média μ_i e desvio padrão σ_i .

A equação da função distribuição acumulada normal utilizada na definição da CCI do item i é dada por Baker e Kim [48] como:

$$P_{ih} = P(Y_{ih} = 1 | \eta_h, \mu_i, \sigma_i) = \Phi(Z_i) = \int_{-\infty}^{Z_i = \frac{\eta_h - \mu_i}{\sigma_i}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz, \quad (3.1)$$

ou, dada a simetria da função, por

$$P_{ih} = P(Y_{ih} = 1 | \eta_h, \mu_i, \sigma_i) = \Phi(Z_i) = \int_{-Z_i = -\frac{\eta_h - \mu_i}{\sigma_i}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz, \quad (3.2)$$

em que, para o item i ,

- μ_i representa a média da distribuição;
- σ_i representa o desvio padrão da distribuição.

A Figura 3.7 mostra o comportamento da CCI quando um dos parâmetros é fixo e o outro varia. A figura da esquerda (MON1) sugere uma relação entre μ_i e a dificuldade do item, β_i , enquanto que a figura da direita (MON2) sugere que σ_i tem uma relação inversa com a discriminação do item, α_i . De fato, pode-se escrever:

$$Z_i = \frac{\eta_h - \mu_i}{\sigma_i} = \alpha_i(\eta_h - \beta_i), \quad (3.3)$$

tal que,

$$\alpha_i = \frac{1}{\sigma_i}, \quad (3.4)$$

$$\beta_i = \mu_i.$$

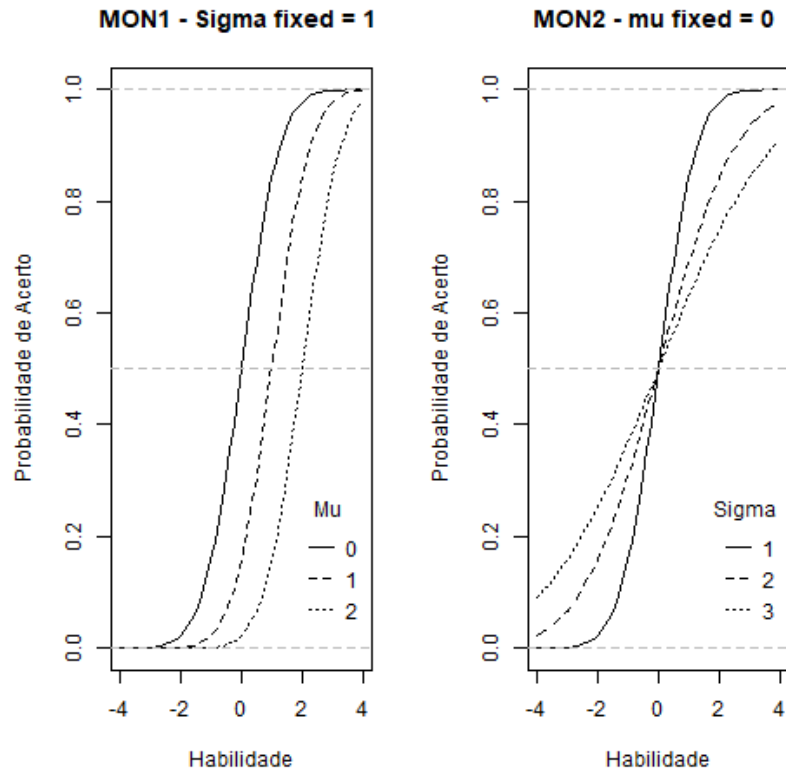


Figura 3.2 MON1 - Comportamento da função de distribuição normal acumulada quando $\sigma_i = 1.00$ e μ_i varia. MON2 - Comportamento da função de distribuição normal acumulada quando $\mu_i = 0.00$ e σ_i varia.

3.3.2 Modelo logístico de um parâmetro - ML1P

O modelo logístico de um parâmetro tem sua CCI definida pela seguinte equação:

$$P_{ihkt} = P(Y_{iht} = 1 | \eta_{hkt}, \zeta_i) = \frac{1}{1 + e^{-D(\eta_{hkt} - b_i)}} \quad (3.5)$$

Em modelos unidimensionais transversais os índices k e t do traço latente observado η_{hkt} e da probabilidade P_{iht} são usualmente abreviados para simplificar a notação pois esses índices têm seus valores fixos na unidade. Assim a Equação (3.1) se torna

$$P_{ih} = P(Y_{ih} = 1 | \eta_h, b_i) = \frac{1}{1 + e^{-D(\eta_h - b_i)}}, \quad (3.6)$$

em que,

- $i, i = 1, \dots, I$, representa o índice dos itens;
- $h, h = 1, \dots, H$, representa o índice dos indivíduos;

- y_{ih} é uma variável dicotômica que assume os valores 1 quando o indivíduo h responde corretamente ao item i , e 0 quando o indivíduo o responde erroneamente;
- b_i é o parâmetro de dificuldade do i -ésimo item, medido na mesma escala das habilidades;
- η_h representa o traço latente do h -ésimo indivíduo;
- D é um fator de escala. Utiliza-se o valor 1.702 de forma que a função logística forneça resultados semelhantes aos da função ogiva normal. Ver Seção 3.3.7 para detalhes;
- ζ_i é o vetor de parâmetros do item i , neste caso, $\zeta_i = (b_i)$

A Equação (3.2) fornece a probabilidade do indivíduo h com traço latente η_h , acertar o item i , caracterizado unicamente pelo parâmetro b_i de dificuldade. A diferença $(\eta_h - b_i)$ em (3.2) indica que o parâmetro b_i é medido na mesma escala do traço latente, e representa a habilidade necessária para uma probabilidade de acerto igual a $1/2$. Quanto maior o valor do parâmetro b_i , mais difícil é o item, e vice-versa. Valores positivos da diferença $(\eta_h - b_i)$ aumentam a probabilidade de acerto do item, enquanto valores negativos a diminuem. A Figura 3.2 mostra como diferentes valores de b_i influenciam no formato da CCI. Itens com maiores valores de b_i são mais difíceis e têm sua curva deslocada para a direita.

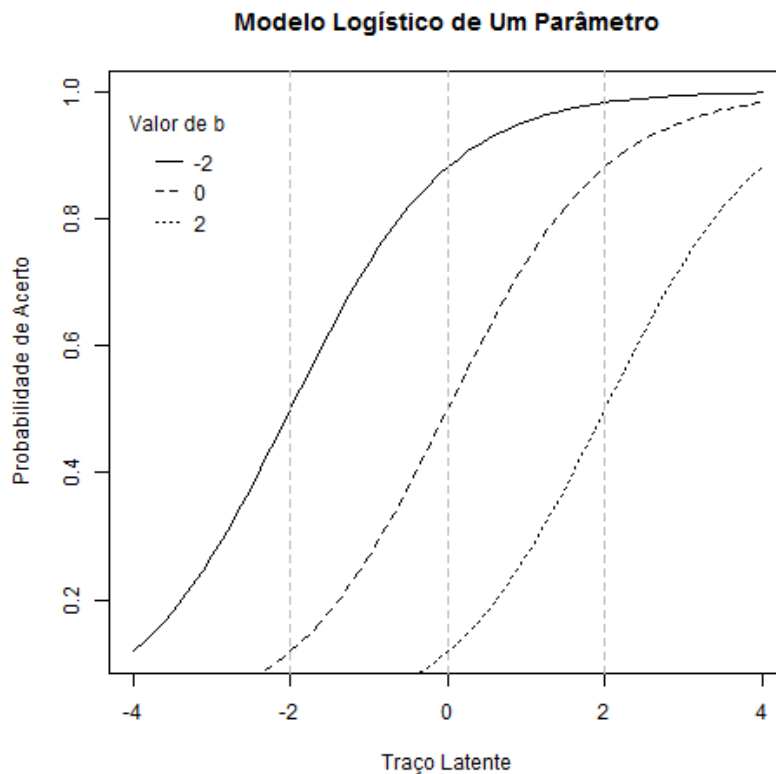


Figura 3.3 Exemplo de como diferentes valores de b_i influenciam o formato da CCI do modelo ML1P ($D = 1.00$)

3.3.3 Modelo logístico de dois parâmetros - ML2P

O modelo logístico de dois parâmetros tem sua CCI definida pela seguinte equação:

$$P_{ih} = P(Y_{ih} = 1 \mid \eta_h, a_i, b_i) = \frac{1}{1 + e^{-Da_i(\eta_h - b_i)}} \quad , \quad (3.7)$$

em que

- a_i é o parâmetro de discriminação do item i , com valor proporcional à inclinação da Curva Característica do Item (CCI) no ponto b_i ;
- ζ_i é o vetor de parâmetros do item i , neste caso, $\zeta_i = (a_i, b_i)'$. Os demais termos estão definidos como no modelo ML1P.

A adição de a_i ao modelo flexibiliza o gradiente de subida da CCI entre os patamares de alta e baixa habilidade. A Figura 3.3 mostra como valores de a_i influenciam o formato da CCI do modelo ML2P quando $b_i = 0.00$ e $D = 1.00$. Baker [49, p. 34] sugere algumas referências sobre como interpretar a discriminação de um item baseado no valor de a_i – Ver Tabela 3.1.

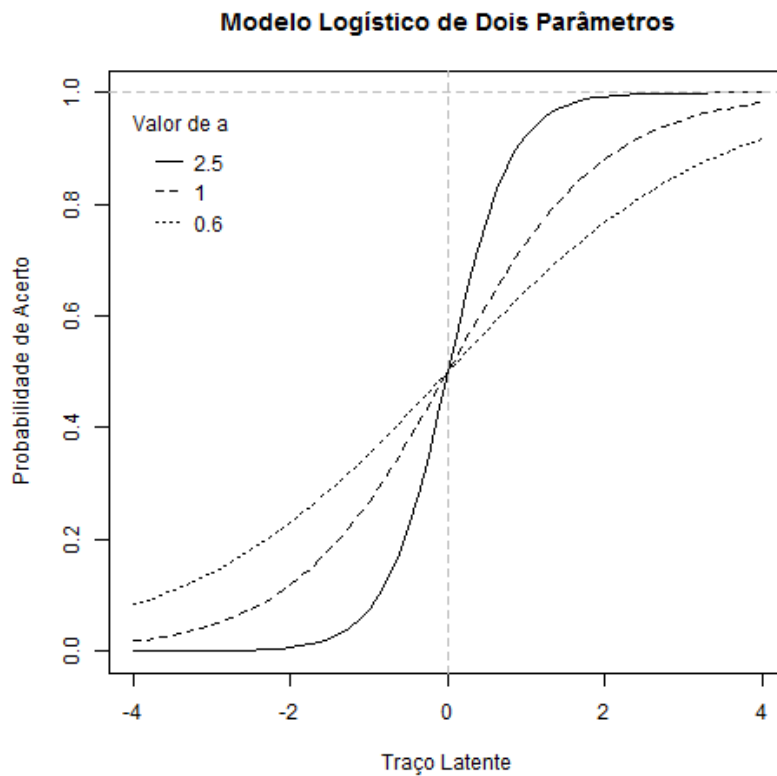


Figura 3.4 Exemplo de como diferentes valores de a_i influenciam o formato da CCI do modelo ML2P quando $b_i = 0.00$ e $D = 1.0$.

Tabela 3.1 Valores de referência para o parâmetro de discriminação a_i segundo Baker [49, p. 34]

Valor de a	Discriminação
0	Nenhuma
0.01 a 0.34	Baixa
0.35 a 0.64	Muito baixa
0.65 a 1.34	Moderada
1.35 a 1.69	Alta
1.70 ou mais	Muito alta

3.3.4 Modelo de três parâmetros - M3P

A equação do modelo de três parâmetros foi proposto por Birnbaum [50] e investigado por Marco [51], e adiciona o parâmetro c_i à equação do modelo de dois parâmetros para levar em consideração o fato de que a probabilidade de indivíduos de baixa habilidade acertarem um item é sempre maior que zero em itens de múltipla escolha com m alternativas ($m > 0$), já que mesmo sem entender o item, tem a probabilidade $\frac{1}{m}$ de escolher aleatoriamente a alternativa correta. A equação utilizada neste modelo é dada por:

$$P_{ih} = P(Y_{ih} = 1 | \eta_h, a_i, b_i, c_i) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\eta_h - b_i)}}, \quad (3.8)$$

em que:

- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i ;
- ζ_i é o vetor de parâmetros do item i , neste caso, $\zeta_i = (a_i, b_i, c_i)'$. Os demais termos estão definidos como no modelo ML2P.

A Figura 3.4 mostra como diferentes valores de c_i influenciam no formato da curva CCI, quando $a_i = 1.00$, $b_i = 0.00$ e $D = 1.00$.

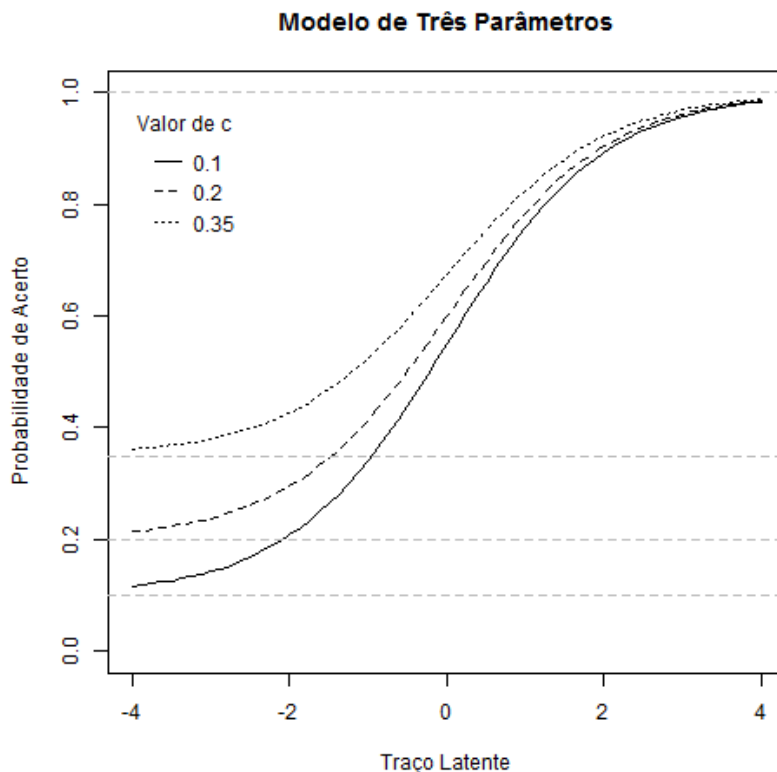


Figura 3.5 Exemplo de como diferentes valores de c_i influenciam o formato da CCI do modelo M3P quando $a_i = 1.00$, $b_i = 0.00$ e $D = 1.00$.

3.3.5 Modelo de quatro parâmetros - M4P

A equação do modelo de quatro parâmetros, proposto por Barton e Lord [52], adiciona o parâmetro γ_i , parâmetro de descuido, à equação do modelo de três parâmetros para levar em consideração o fato de a probabilidade de indivíduos de alta habilidade poderem eventualmente errar um item fácil por motivos não-relacionados à sua habilidade. Eles podem não terem tido tempo suficiente para analisar o item em uma prova longa, ou podem cometer erros na marcação de seus cartões-resposta. Esses casos têm probabilidade nula de acontecer em modelos como o ML3P, levando o processo de estimação de habilidades a subestimar a habilidade do indivíduo. Veja Loken e Rulison [53], para uma discussão sobre a dificuldade de se estimar o parâmetro γ_j e Tavares *et al.* [54] para uma aplicação deste modelo em Genética.

A equação utilizada neste modelo é a seguinte:

$$P_{ih} = P(Y_{ih} = 1 | \eta_h, a_i, b_i, c_i, \gamma_i) = c_i + \frac{\gamma_i - c_i}{1 + e^{-Da_i(\eta_h - b_i)}}, \quad (3.9)$$

em que,

- γ_i é o parâmetro de descuido do item. Representa a probabilidade de indivíduos com alta habilidade responderem incorretamente o item i ;
- ζ_i é o vetor de parâmetros do item i , neste caso, $\zeta_i = (a_i, b_i, c_i, \gamma_i)'$. Os demais termos estão definidos como no modelo ML3P.

A Figura 3.5 mostra como diferentes valores de γ_i influenciam no formato da curva CCI, quando $a_i = 1.00$, $b_i = 0.00$, $c_i = 0.20$ e $D = 1.00$.

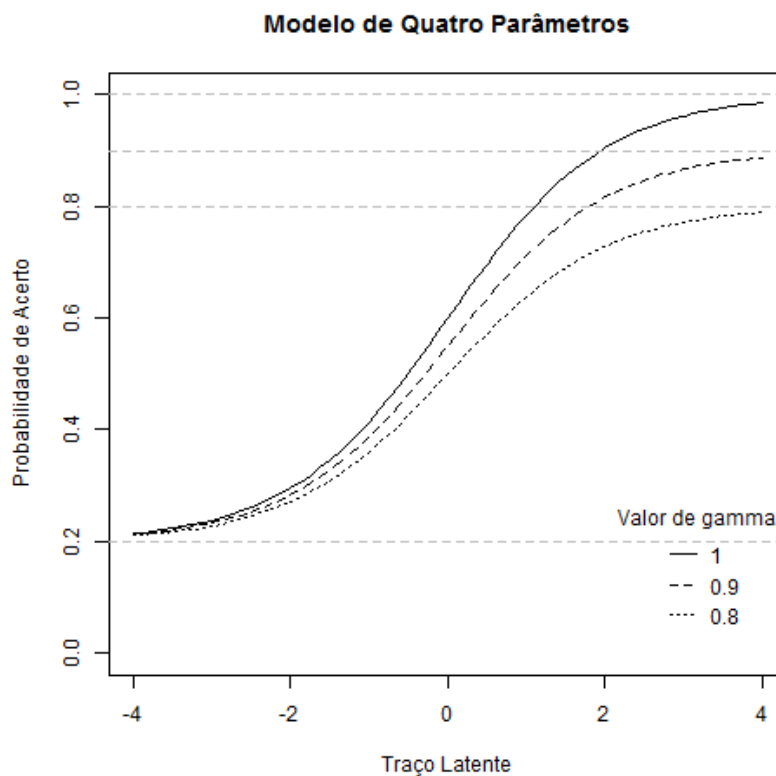


Figura 3.6 Exemplo de como diferentes valores de γ_j influenciam o formato da CCI do modelo M4P quando $a_i = 1.00$, $b_i = 0$, $c_i = 0.20$ e $D = 1.00$.

3.3.6 Modelo de cinco parâmetros - M5P

A equação do modelo de cinco parâmetros adiciona o parâmetro δ_i , parâmetro de simetria, à equação do modelo de quatro parâmetros para controlar a assimetria da CCI nos arredores do parâmetro b_i . Valores menores que 1 fazem com que o intervalo de discriminação da CCI se mova para a esquerda de b_i , enquanto que valores maiores que 1 o movem para a direita de b_i , como mostra a Figura 3.6. Veja [55] para uma discussão sobre a estimação do parâmetro δ_i .

A equação utilizada neste modelo é a seguinte:

$$P_{ih} = P(Y_{ih} = 1 | \eta_h, a_i, b_i, c_i, \gamma_i, \delta_i) = c_i + \frac{\gamma_i - c_i}{(1 + e^{-D a_i (\eta_h - b_i)})^{\delta_i}}, \quad (3.10)$$

em que,

- δ_i é o parâmetro de assimetria do item i ;
- ζ_i é o conjunto de parâmetros do item i , neste caso, $\zeta_i = (a_i, b_i, c_i, \gamma_i, \delta_i)'$. Os demais termos estão definidos como no modelo ML4P.

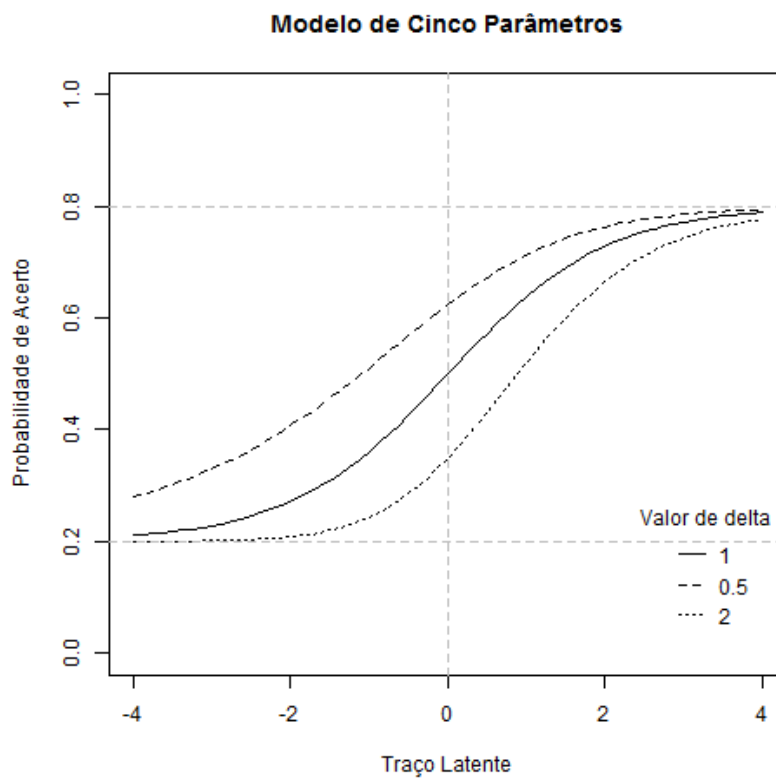


Figura 3.7 Exemplo de como diferentes valores de δ_i influenciam o formato da CCI do modelo M4P quando $a_i = 1.00$, $b_i = 0.00$, $c_i = 0.20$, $\gamma_i = 0.80$ e $D = 1.00$.

3.3.7 Relação entre os modelos logístico e ogiva normal

Haley [56] demonstrou que quando D assume o valor de 1.702 as Equações (3.3) e (3.8) são praticamente idênticas, como mostram as Figura 3.8 e Figura 3.9, ou seja,

$$\Phi(\alpha_i(\eta_h - \beta_i)) \approx \frac{1}{1 + \exp(-1.702\alpha_i(\eta_h - b_i))} \quad (3.11)$$

estabelecendo que os parâmetros de discriminação obtidos a partir de modelos logísticos da TRI de dois parâmetros são aproximadamente 1.702 vezes os parâmetros correspondentes dos modelos de ogiva normais, ou seja, da Equação (3.9) temos que:

$$\begin{aligned} a_i &\approx 1.702 \alpha_i, \\ b_i &\approx \beta_i . \end{aligned} \quad (3.12)$$

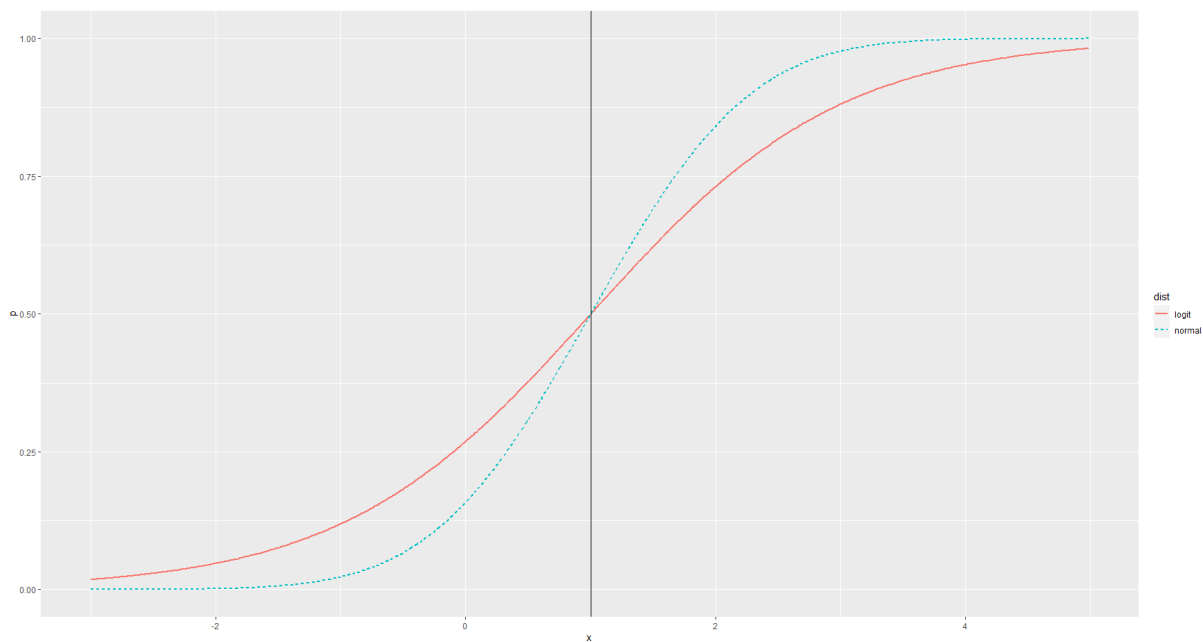


Figura 3.8 Funções de distribuição acumulada das distribuições normal e logística ($D=1.000$)

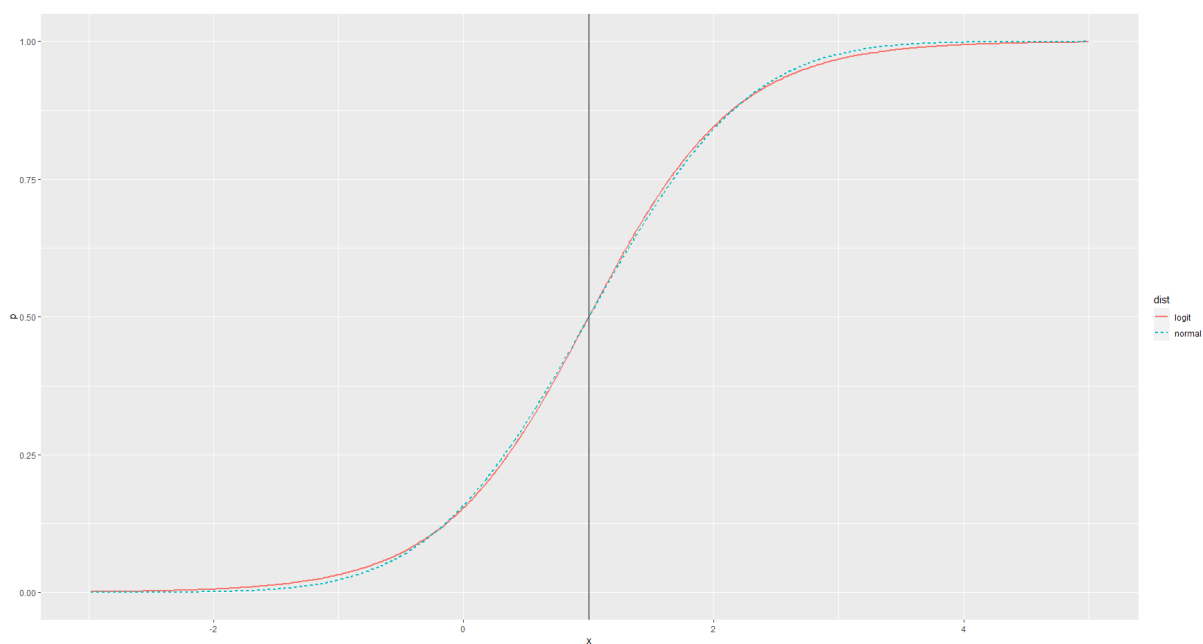


Figura 3.9 Funções de distribuição acumulada das distribuições normal e logística ($D=1.702$)

3.4 Modelos multidimensionais

Em muitas situações, as manifestações da resposta de um indivíduo a um item dependem de múltiplas variáveis latentes. Nesta seção apresentaremos esta extensão dos modelos da TRI.

3.4.1 Modelo multidimensional logístico

A multidimensionalidade não é rara no mundo real e autores, como Reckase [22, p. 53], chegam a sugerir que modelos unidimensionais são incapazes de expressar as interações existentes entre um indivíduo e os itens que o avaliam, pois por mais simples que um item seja ele, muito provavelmente, exige o exercício de múltiplas habilidades para se chegar a resposta correta. Um exemplo muito citado e estudado na literatura é sobre a dimensionalidade de um problema contextualizado de matemática no qual se testa a capacidade de um indivíduo de aplicar um conceito matemático em uma situação ou contexto real. Claramente, seu desempenho no problema depende não somente de sua habilidade lógico matemática, usada para resolver o problema quando este já se encontra em linguagem matemática, como também de sua habilidade verbal-linguística, usada para entender e transpor uma situação expressa em linguagem natural para a linguagem matemática [27].

Modelos multidimensionais são classificados em modelos compensatórios e parcialmente compensatórios. Modelos compensatórios são usados em situações em que um baixo desempenho em uma dimensão pode ser compensado por um alto desempenho em outra. O caso do parágrafo anterior é exemplo deste tipo de modelo e as dimensões são relacionadas de maneira linear. Os modelos parcialmente compensatórios são usados em situações em que o alto desempenho em uma dimensão em nada ou em pouco⁶ são compensados para alguns intervalos de desempenho em outra dimensão [22]. As dimensões em modelos parcialmente compensatórios têm uma relação não-linear e são bem menos estudados que os compensatórios devido a dificuldades relacionadas à estimação de parâmetros [26] [57].

Em modelos unidimensionais de dois parâmetros, a probabilidade de um indivíduo acertar um item é função, f' , da diferença entre a localização do indivíduo na escala η_h e a localização, na mesma escala, do item b_i multiplicada pelo fator de discriminação do item a_i , ou seja, $P(\eta_h, a_i, b_i) = f'(a_i(\eta_h - b_i))$. Ora, nada impede desse mesmo conceito ser expandido para um cenário multidimensional, ou seja, a contribuição da dimensão 1 na probabilidade de um indivíduo

⁶ Esses modelos também são chamados de *parcialmente compensatórios*.

acertar um item é função, f' , da diferença entre a localização do indivíduo η_{hk} na dimensão k e a localização do item b_{ik} naquela dimensão multiplicada pelo fator de discriminação do item a_{ik} , ou seja, $f'(a_{i1}(\eta_{h1} - b_{i1}))$. Da mesma maneira, a contribuição da dimensão 2 na probabilidade de acerto seria $f'(a_{i2}(\eta_{h2} - b_{i2}))$. Assim, é natural pensar que a contribuição conjunta destas duas dimensões poderia ser escrita na forma aditiva $f'(a_{i1}(\eta_{h1} - b_{i1}) + a_{i2}(\eta_{h2} - b_{i2}))$, com os parâmetros a_{i1} e a_{i2} atribuindo pesos e definindo a contribuição da dimensão k no cálculo da probabilidade final. De fato, veremos que a carga de um item em uma dimensão ou fator reflete o grau de relação deste item com a dimensão e que a carga está intrinsecamente ligada ao parâmetro de discriminação do item naquela dimensão. Assim, podemos escrever o modelo logístico multidimensional com duas dimensões como a seguir:

$$P(Y_{ih} = 1 | \eta_{h1}, \eta_{h2}, a_{i1}, a_{i2}, b_{i1}, b_{i2}) = \frac{1}{1 + e^{-a_{i1}(\eta_{h1} - b_{i1}) - a_{i2}(\eta_{h2} - b_{i2})}}, \quad (3.13)$$

ou

$$P(Y_{ih} = 1 | \eta_{h1}, \eta_{h2}, a_{i1}, a_{i2}, d_i) = \frac{1}{1 + e^{-(a_{i2}\eta_{h2} + a_{i1}\eta_{h2} + d_i)}}, \quad (3.14)$$

em que

$$d_i = -a_{i1}b_{i1} - a_{i2}b_{i2} = -\sum_{k=1}^2 a_{ik}b_{ik}. \quad (3.15)$$

Generalizando para $k = 1, \dots, K$, temos que a probabilidade de acerto de um item i segundo o modelo K -dimensional de dois parâmetros é dada por

$$P(Y_{ih} = 1 | \boldsymbol{\eta}_h, \mathbf{a}_i, d_i) = \frac{1}{1 + e^{-(\sum_{k=1}^K a_{ik}\eta_{hk} + d_i)}}. \quad (3.16)$$

Tal qual foi feito com o modelo unidimensional de dois parâmetros, pode-se adicionar à Equação (3.16) um parâmetro c_i que representa a probabilidade de acerto ao acaso do item, ou seja, a probabilidade de acerto de um item i segundo o modelo K -dimensional de três parâmetros é dada por:

$$P(Y_{ih} = 1 | \boldsymbol{\eta}_h, \mathbf{a}_i, c_i, d_i) = c_i + \frac{1 - c_i}{1 + e^{-(\sum_{k=1}^K a_{ik}\eta_{hk} + d_i)}}. \quad (3.17)$$

A Figura 3.10 mostra um exemplo da *superfície característica do item* e a Figura 3.11 mostra a relação compensatória existente entre η_{h1} e η_{h2} , que leva a um mesmo valor de probabilidade. Por exemplo, no caso mais extremo, nota-se que existem infinitas combinações de η_{h1} e η_{h2} que resultam numa probabilidade de 95% de acerto do item.

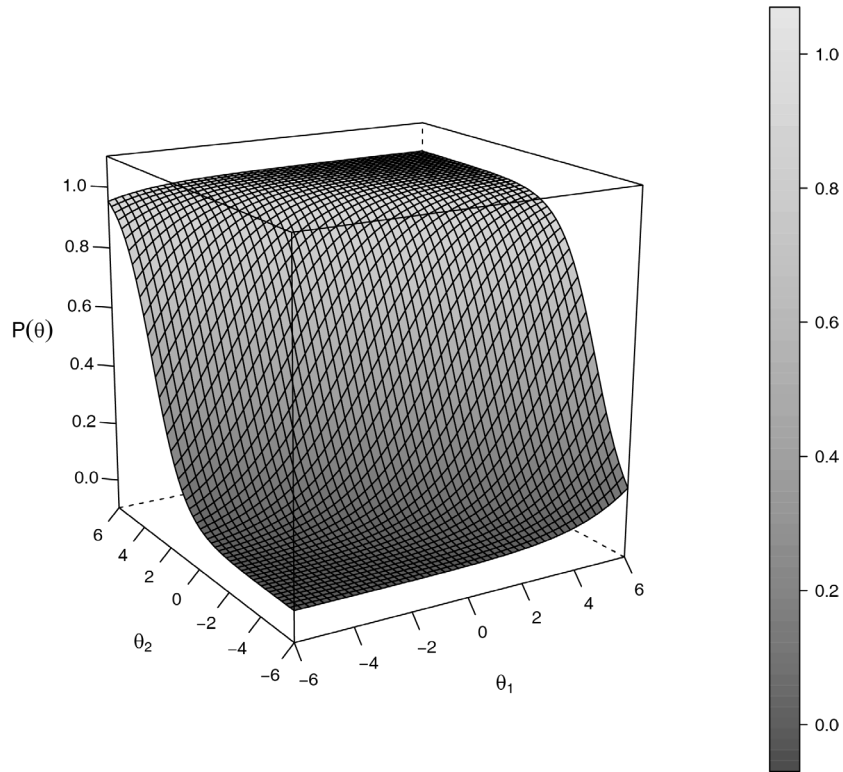


Figura 3.10 Exemplo da superfície característica do item quando $a_{i1} = 1.288$, $a_{i2} = 0.567$, $d_i = -0.047$, $c_i = 0.253$.

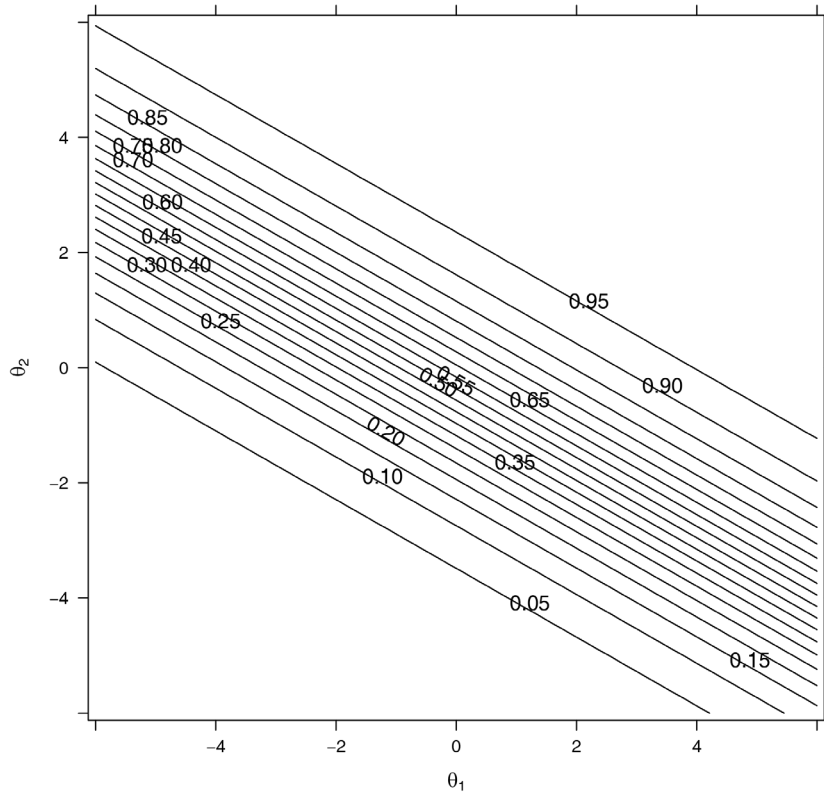


Figura 3.11 Curvas de nível da superfície característica do item quando $a_{i1} = 1.288$, $a_{i2} = 0.567$, $d_i = -0.047$, $c_i = 0.253$.

Wang e Chen [58] distinguem em Entre-Item e Intra-Item as relações existentes entre os itens e as variáveis latentes associadas, como mostra a Figura 3.12. Babcock [59] menciona que a distinção entre esses dois tipos de dimensionalidades é vital para a identificabilidade do modelo devido à indeterminação rotacional⁷ intrínseca de modelos multidimensionais (MIRT), principalmente em modelos com dimensionalidade intra-itens acentuada.

⁷ Rotational Indeterminacy

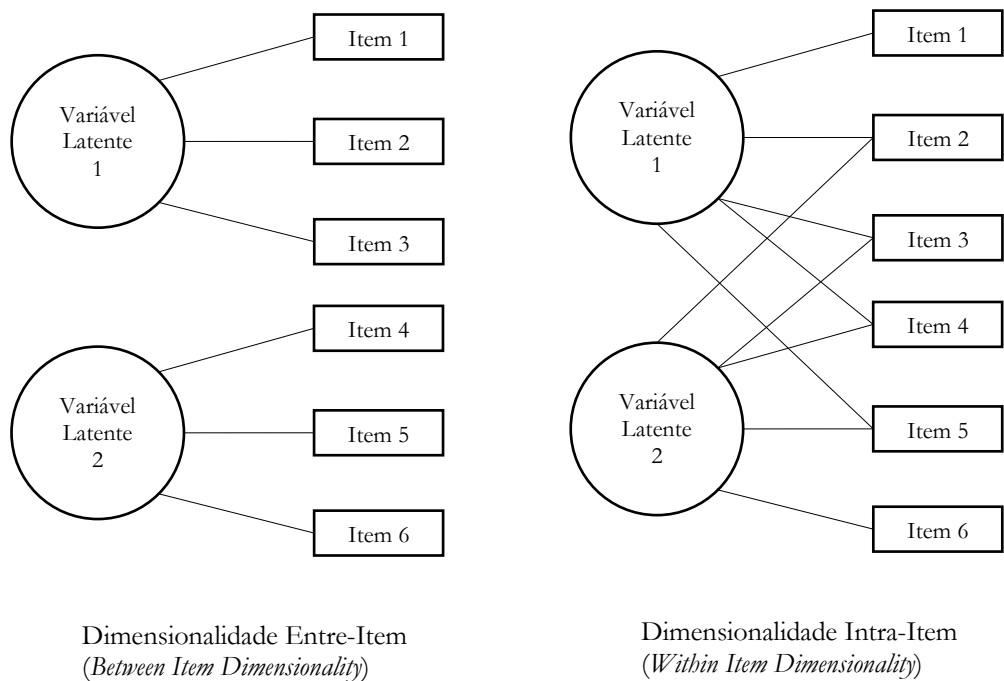


Figura 3.12 Dimensionalidades intra-item e entre-item.

3.5 Estimação dos parâmetros e habilidades na TRI

A estimação dos parâmetros dos itens e das habilidades dos respondentes é uma das etapas mais importantes do método da TRI. A probabilidade de um respondente acertar um determinado item depende somente do traço latente desse indivíduo e dos parâmetros que caracterizam os itens. Nas seções seguintes discutiremos métodos para estimar esses parâmetros a partir das respostas dos indivíduos aos itens do teste.

Segundo Andrade, Tavares e Valle [60, p. 27], a estimação dos parâmetros dos itens e/ou das habilidades dos respondentes pode ser dividida em três situações: a primeira situação ocorre quando os parâmetros dos itens são conhecidos, tendo apenas que estimar as habilidades. A segunda situação ocorre quando as habilidades dos respondentes são conhecidas, tendo que estimar os parâmetros dos itens. E uma terceira situação ocorre quando os parâmetros dos itens e as habilidades dos respondentes são desconhecidas sendo necessário realizar a estimação de ambas simultaneamente.

Seja η_h o valor observado do traço latente de indivíduo h , η com distribuição definida pelo vetor de parâmetros $\boldsymbol{\pi}$, $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_I)'$, o conjunto de parâmetros de todos os itens do teste, y_{ih}

a resposta do indivíduo h ao item i , $\mathbf{Y}_{.h} = (Y_{1h}, Y_{2h}, \dots, Y_{Ih})'$ o vetor de todas as respostas do indivíduo h aos I itens, $\mathbf{Y}_{..} = (\mathbf{Y}'_{.1}, \mathbf{Y}'_{.2}, \dots, \mathbf{Y}'_{.H})'$ a respostas de todos os indivíduos aos I itens, usando a independência entre as respostas de diferentes indivíduos, podemos escrever a probabilidade associada à matriz de respostas $\mathbf{y}_{..}$ de um modelo unidimensional transversal como:

$$L(\boldsymbol{\zeta}, \boldsymbol{\eta} | \mathbf{Y}_{..}) = P(\mathbf{Y}_{..} | \boldsymbol{\zeta}, \boldsymbol{\eta}) = \prod_{h=1}^H \prod_{i=1}^I P_{ih} = \prod_{h=1}^H \prod_{i=1}^I P(Y_{ih} | \eta_h, \zeta_i). \quad (3.18)$$

A verossimilhança descrita na Equação (3.18) é função dos parâmetros dos itens e dos parâmetros populacionais associados ao traço latente, e servirá de base para as abordagens descritas em seguida.

Como a granularidade da Equação (3.18) é a probabilidade de acerto a um item, o processamento computacional dessa verossimilhança pode ser desafiadora quando H for grande. Esse processamento pode ser otimizado se levarmos em conta que o cálculo da verossimilhança de um padrão de resposta pode ser reaproveitado no processamento da verossimilhança de todos os indivíduos que praticaram o mesmo padrão. Isso é viabilizado reorganizando a matriz de dimensões $H \times I$ contendo todas as respostas de indivíduos de um grupo em uma estrutura que leva em conta o número S de diferentes padrões de respostas contidos na matriz e o número r_u de ocorrências de cada padrão u . Como com I itens é possível construir $S = 2^I$ diferentes padrões de resposta binária, mas nem todos se realizam, segue que

$$s = \min(H, S), \quad (3.19)$$

e que

$$\sum_{u=1}^s r_u = H. \quad (3.20)$$

Com essas definições, a Equação (3.18) pode ser reescrita em função de padrões de resposta, ou seja,

$$P(\mathbf{Y}_{..} | \boldsymbol{\zeta}, \boldsymbol{\eta}) = M \prod_u^s (P(Y_{.u} | \boldsymbol{\eta}, \boldsymbol{\zeta}))^{r_u}, \quad (3.21)$$

em que M é o coeficiente multinomial definido por:

$$M = \frac{H!}{\sum_{u=1}^s r_u!} \quad (3.22)$$

3.5.1 Estimação do traço latente

3.5.1.1 Estimação por máxima verossimilhança

Quando os parâmetros dos itens são conhecidos, ou seja, a escala de mensuração já está definida, a Equação (3.18) torna-se função unicamente do traço latente e os valores estimados dos parâmetros são os valores que a maximizam, ou seja, é a solução da seguinte equação:

$$\frac{\partial L(\boldsymbol{\eta})}{\partial \eta_h} = 0. \quad (3.23)$$

É interessante notar que quando tratamos da estimação do traço latente com os parâmetros dos itens conhecidos, estes são estimados na métrica dos parâmetros dos itens e que os parâmetros da distribuição do traço latente não são estimados. Ver Andrade, Tavares e Valle [61, p. 44].

3.5.1.2 Estimação pelo método *Expected a Posteriori* (EAP) de Bayes

A estimação das habilidades quando os parâmetros dos itens são conhecidos é feita pelo método EAP de Bayes em que a esperança de η_{hkt} dadas as respostas do indivíduo $\mathbf{Y}_{.ht}$ e os parâmetros dos itens $\boldsymbol{\zeta}_i$ é dada por (Mislevy e Bock [62]):

$$E(\eta_{hkt} | \mathbf{Y}_{.ht}, \boldsymbol{\zeta}_i) = \frac{\int \eta_{hkt} f(\eta_{kt} | \boldsymbol{\pi}_{kt}) \prod_{i=1}^I P_{iht}^{y_{iht}}(\eta_{hkt}) Q_{iht}^{1-y_{iht}}(\eta_{hkt}) d\eta}{\int f(\eta_{kt} | \boldsymbol{\pi}_{kt}) \prod_{i=1}^I P_{iht}^{y_{iht}}(\eta_{hkt}) Q_{iht}^{1-y_{iht}}(\eta_{hkt}) d\eta}, \quad (3.24)$$

que estima a média da seguinte posteriori;

$$f(\eta_{hkt} | \mathbf{Y}_{.ht}, \boldsymbol{\zeta}_i, \boldsymbol{\pi}) = \frac{P(\mathbf{Y}_{.ht} | \eta_{kt}, \boldsymbol{\zeta}_i) f(\eta_{kt} | \boldsymbol{\pi}_{kt})}{P(\mathbf{Y}_{.ht} | \boldsymbol{\pi}_{kt}, \boldsymbol{\zeta}_i)}. \quad (3.25)$$

3.5.2 Estimação dos parâmetros dos itens

Quando os parâmetros populacionais do traço latente são conhecidos, a Equação (3.18) torna-se função unicamente dos parâmetros dos itens e os valores estimados destes são os valores que a maximizam, ou seja, são as soluções da seguinte equação [61]:

$$\frac{\partial L(\boldsymbol{\zeta})}{\partial \zeta_i} = 0 \quad (3.26)$$

Da mesma maneira, é interessante notar que quando tratamos da estimação dos parâmetros dos itens com valores de traço latente conhecidos, esses parâmetros são estimados na métrica do traço latente.

3.5.3 Estimação conjunta do traço latente e dos parâmetros dos itens

Neste caso, como tanto os parâmetros dos itens quanto os parâmetros populacionais do traço latente são desconhecidos, uma métrica deve ser atribuída para possibilitar a identificação do modelo. Isso é necessário porque existem mais que um conjunto de valores para esses parâmetros que resultam numa mesma probabilidade calculada pela Equação (2.5). Isso é demonstrado posteriormente na discussão da Equação (3.37) na Seção 3.6.3. Em geral, a distribuição do traço latente é normalizada com média $\mathbf{0}$ e desvio padrão $\mathbf{1}$, mas isso é feito por pura conveniência pois o modelo não impõe nenhuma restrição quanto ao tipo de distribuição do traço latente. Uma vez definida a métrica, a estimação é feita em um processo iterativo de dois estágios. Primeiramente, os parâmetros dos itens são estimados utilizando os valores iniciais de $\boldsymbol{\eta}_h$, e depois, utilizando os valores dos parâmetros estimados no estágio 1, estima-se os traços latentes. Esse processo repete-se até que a convergência seja alcançada [61].

3.5.4 Estimação por máxima verossimilhança marginal

Uma das críticas mais frequentes na literatura sobre o método de estimação conjunta descrita na Seção 3.5.3 é a estimação concomitante que Andersen [63] chamou de parâmetros estruturais (parâmetros dos itens) com parâmetros incidentes (traço latente dos indivíduos). Enquanto que o número de parâmetros estruturais para um dado teste não aumenta com o tamanho da amostra, o número de parâmetros incidentes aumenta proporcionalmente. De fato, Neyman [64] mostrou que

as estimações de parâmetros estruturais não são consistentes com o aumento do tamanho da amostra e De Gruijter [65], que eles são enviesados.

O método da Máxima Verossimilhança Marginal (MVM) como descrito por Bock e Aitkin [66] resolve essa questão estimando os parâmetros estruturais pelo uso da distribuição dos traços latentes para eliminar a dependência do processo de estimação dos valores individuais desses traços. Para isso calcula-se a probabilidade marginal $P(Y_h|\boldsymbol{\pi}, \boldsymbol{\zeta})$ de um padrão de respostas ser obtido a partir de um conjunto de I itens com $\boldsymbol{\zeta}$ parâmetros aplicados a indivíduos cujo traço latente η tem função densidade de distribuição $f(\eta|\boldsymbol{\pi})$ em $\boldsymbol{\pi}$ é o vetor de parâmetros da distribuição.

Considere $P(\eta_h)$ a probabilidade de um indivíduo h ter traço latente de valor η_h , e $P(y_h|\eta_h, \boldsymbol{\pi}, \boldsymbol{\zeta})$ a probabilidade do indivíduo h apresentar o padrão de resposta y_h dado seu traço latente η_h , e conjuntos $\boldsymbol{\pi}$ e $\boldsymbol{\zeta}$ de parâmetros, então pelo Teorema de Bayes⁸ [67] podemos escrever:

$$P(\eta_h|Y_h, \boldsymbol{\pi}, \boldsymbol{\zeta}) = \frac{P(\eta_h)P(Y_h|\eta_h, \boldsymbol{\pi}, \boldsymbol{\zeta})}{P(Y_h|\boldsymbol{\pi}, \boldsymbol{\zeta})} = \frac{P(\eta_h)P(Y_h|\eta_h, \boldsymbol{\pi}, \boldsymbol{\zeta})}{\sum_{h=1}^H P(\eta_h)P(Y_h|\eta_h, \boldsymbol{\pi}, \boldsymbol{\zeta})}. \quad (3.27)$$

Se $P(\eta_h)$ for calculada a partir do produto da função densidade de probabilidade do traço latente e de um infinitesimal $\Delta\eta$ centrado em η_h , temos que,

$$P(\eta_h) = f(\eta_h|\boldsymbol{\pi})\Delta\eta. \quad (3.28)$$

A Equação (3.27) então fica,

$$P(\eta_h|Y_h, \boldsymbol{\zeta}) = \frac{f(\eta_h|\boldsymbol{\pi})\Delta\eta P(Y_h|\eta_h, \boldsymbol{\pi}, \boldsymbol{\zeta})}{\sum_{h=1}^H P(Y_h|\eta_h, \boldsymbol{\pi}, \boldsymbol{\zeta}) f(\eta_h|\boldsymbol{\pi})\Delta\eta}, \quad (3.29)$$

ou seja, a equação da probabilidade marginal $P(Y_h|\boldsymbol{\pi}, \boldsymbol{\zeta})$ é o denominador da Equação (3.29) e é dada por,

$$P(Y_h|\boldsymbol{\pi}, \boldsymbol{\zeta}) = \sum_{h=1}^H P(Y_h|\eta_h, \boldsymbol{\pi}, \boldsymbol{\zeta}) f(\eta_h|\boldsymbol{\pi})\Delta\eta. \quad (3.30)$$

Finalmente, estimamos os valores dos parâmetros $\boldsymbol{\zeta}$ e $\boldsymbol{\pi}$ ao maximizar a verossimilhança dada a seguir.

⁸ Teorema de Bayes: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} = \frac{P(A)P(B|A)}{\sum_{i=1}^{\infty} P(A_i)P(B|A_i)}$. Ver Degroot.e Schervish [67].

$$L(\boldsymbol{\zeta}) = \prod_{h=1}^H P(Y_{.h} | \boldsymbol{\pi}, \boldsymbol{\zeta}). \quad (3.31)$$

Andrade, Tavares e Valle [61] discutem os detalhes da implementação numérica desses cálculos.

3.6 Modelos de grupos múltiplos

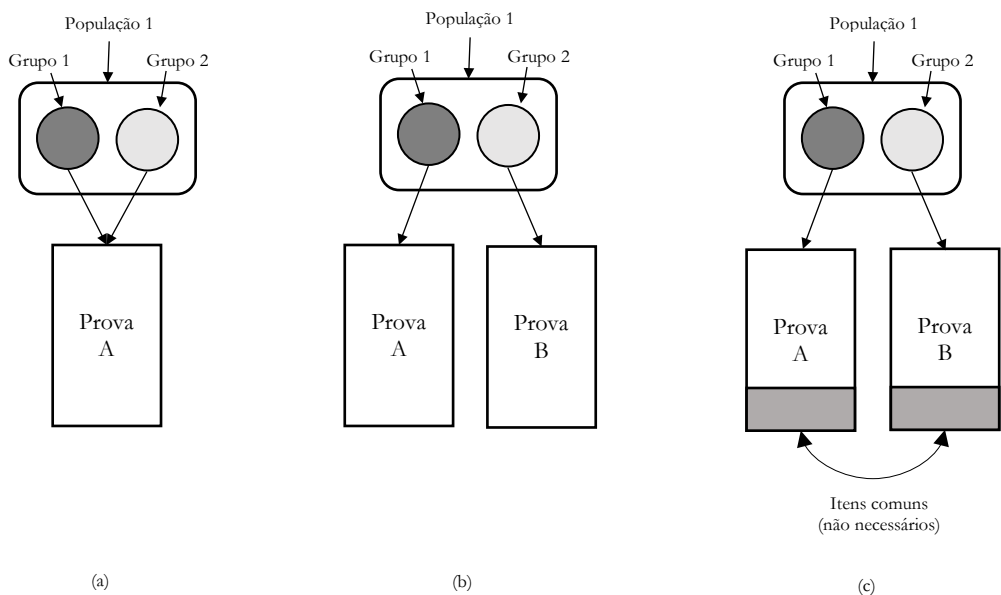
O objetivo de uma análise de grupos múltiplos é permitir a estimação conjunta dos parâmetros dos itens e do fator latente de maneira a permitir a análise diferencial de itens (análise de DIF), análise de desvio de parâmetros dos itens (análise de DRIFT), e a equalização de itens e do traço latente. A análise de DIF permite a identificação de itens que se comportam de maneira diferente em grupos não equivalentes: grupos com diferentes características demográficas, grupos divididos por sexo, grupos com diferentes características socioeconômicas. A análise de DRIFT permite identificar como o comportamento de um item varia com o tempo. Finalmente, a equalização nos permite manter a comparabilidade de traços latentes observados em diferentes momentos e/ou diferentes testes. Uma característica comum no planejamento destes experimentos é a presença de um conjunto de itens comuns que é respondido pelos elementos dos grupos envolvidos [35].

Em [36], Bock e Zimowski descrevem como estimar os parâmetros de itens e a distribuição do traço latente em cenários de grupos múltiplos utilizando o método da máxima verossimilhança marginal proposto em [66] para cenários com um grupo.

3.6.1 Equalização em modelos de grupos múltiplos

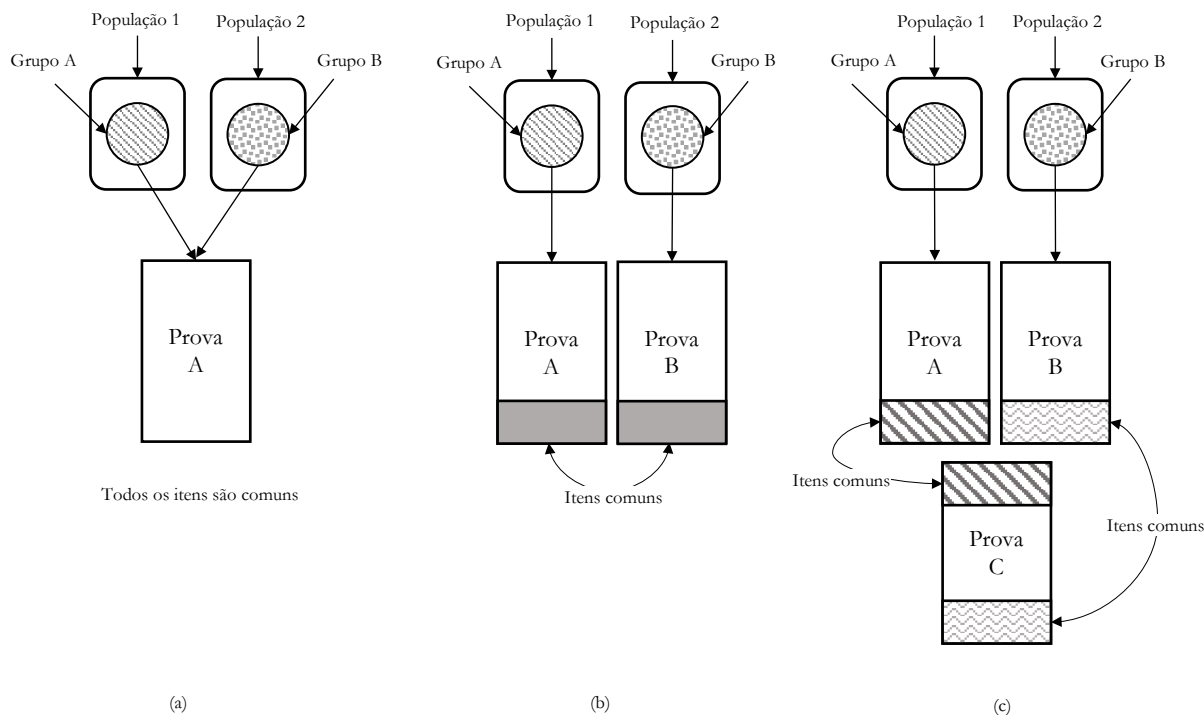
Equalizar significa tornar comparáveis os parâmetros de itens e traços latentes obtidos a partir de instrumentos aplicados a diferentes grupos. A abordagem de equalização varia dependendo se o grupo ou amostra de indivíduos veio de uma mesma população ou não. Quando os grupos são amostras aleatórias de uma mesma população, os parâmetros e traços latentes obtidos após o processo de estimação estão na mesma escala e são, portanto, comparáveis. Esta equalização é também chamada de *equalização via população* e a Figura 3.13 mostra alguns casos de sua aplicação. Note que a presença de itens comuns entre as provas aplicadas aos grupos é opcional e que a equalização é automática dado que todos os respondentes são oriundos de uma mesma população. Por sua vez, quando os grupos são oriundos de populações diferentes, a equalização via itens comuns é aplicada. A Figura 3.14 mostra exemplos de como o instrumento de medida deve ser desenhado para conter itens comuns entre as aplicações. A Figura 3.14 (b) mostra o caso mais óbvio em que dois grupos são equalizados através do compartilhamento de um mesmo grupo de

itens, e a Figura 3.14 (c) mostra o caso em que dois grupos, que respondem a itens completamente diferentes, são equalizados através de uma prova intermediária que contém itens comuns com a prova aplicada aos dois grupos. A equalização pode ser feita através da estimação simultânea dos parâmetros como descrito por Bock e Zimowski em [36] ou *a posteriori*.



Grupos 1 e 2 são equivalentes, são formados a partir de amostras aleatórias de uma mesma população.

Figura 3.13 Equalização de grupos equivalentes ou equalização via população



Grupo A e Grupo B não são equivalentes, são formados a partir de amostras aleatórias de populações diferentes.

Figura 3.14 Equalização de grupos não equivalentes via itens comuns.

3.6.2 Equalização via estimação simultânea

Nesta seção é apresentado o método de estimação de grupos múltiplos proposta por Bock e Zimowski [36], cuja grande vantagem é que a equalização dos resultados obtidos em grupos não equivalentes se dá durante o processo de estimação conjunta dos parâmetros dos itens e dos parâmetros da distribuição dos traços latentes dos diferentes grupos.

Em modelos com um grupo, a escala de medida do traço latente ou métrica é definida fixando-se valores para sua origem e intervalo, feito, muito comumente, fixando-se os parâmetros da distribuição atribuída ao traço latente. Por exemplo, com bastante frequência assume-se que o traço latente é uma normal de média $\mu = 0$ e desvio padrão $\sigma = 1$. Já em modelos de grupos múltiplos com G grupos, os parâmetros populacionais da distribuição do grupo de referência são fixos, restando $G - 1$ conjuntos de parâmetros que precisam ser estimados na escala do grupo de referência. Neste desenho, é importante ressaltar que tanto os parâmetros da distribuição latente quanto os parâmetros dos itens ficam na métrica da escala determinada para o grupo de referência.

Seja ζ_i o conjunto de parâmetros do item i , $i = 1, \dots, I$, e I sendo o número total de itens aplicados aos G grupos, define-se $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_I)$ o conjunto de todos os parâmetros dos itens aplicados.

Segue ainda que se I_g é o número de itens aplicados ao grupo g , temos que $I \leq \sum_{g=1}^G I_g$ já que alguns itens serão comuns entre os grupos. A Figura 3.15 ilustra o desenho de um experimento com dois grupos não equivalentes ligados através de itens comuns. Em geral, nem todos os itens do experimento são apresentados a cada grupo, ou seja, $I_g \leq I$, e I_g são os itens comuns entre os grupos que permitem que os parâmetros estimados estejam na métrica do grupo de referência.

Seja y_{ihg} a resposta do indivíduo h pertencente ao grupo g ao item i , define-se $\mathbf{Y}_{.hg} = (y_{1hg}, y_{2hg}, \dots, y_{I_g hg})$ como o vetor de respostas do indivíduo h pertencente ao grupo g a I_g itens, e $\mathbf{Y}_{..g} = (\mathbf{Y}'_{.1g}, \mathbf{Y}'_{.2g}, \dots, \mathbf{Y}'_{.H_g g})$ como o conjunto de respostas de todos os H_g indivíduos do grupo g que responderam aos I_g itens, e $\mathbf{Y}_{...} = (\mathbf{Y}'_{..1}, \mathbf{Y}'_{..2}, \dots, \mathbf{Y}'_{..G})$ como o conjunto total de respostas no G grupos. Seja ainda η_{hkg} o valor observado do k -ésimo traço latente do indivíduo h do grupo g de uma população com distribuição η_{kg} definida pelo vetor de parâmetros $\boldsymbol{\pi}_{kg}$, define-se $\boldsymbol{\eta}_k = (\eta_{k1}, \eta_{k2}, \dots, \eta_{kG})$ o vetor de distribuições latentes associadas aos G grupos na dimensão k , então usando a independência entre as respostas de diferentes indivíduos, podemos escrever a probabilidade associada ao conjunto de respostas $\mathbf{Y}_{...}$ como:

$$L(\boldsymbol{\zeta}, \boldsymbol{\eta}_k | \mathbf{Y}) = P(\mathbf{Y}_{...} | \boldsymbol{\zeta}, \boldsymbol{\eta}_k) = \prod_{g=1}^G \prod_{h=1}^{H_g} \prod_{i=1}^{I_g} P_{ihg} = \prod_{g=1}^G \prod_{h=1}^{H_g} \prod_{i=1}^{I_g} P(Y_{ihg} | \eta_{hkg}, \zeta_i) \quad (3.32)$$

A verossimilhança descrita na Equação (3.32) é função dos parâmetros dos itens e dos parâmetros latentes associados a cada grupo, e os valores estimados desses parâmetros são os valores que a maximizam. Da mesma maneira como foi feito para o caso com um grupo, essa verossimilhança pode ser escrita em função de padrões de resposta, e assim, seja s_g o número de diferentes padrões de respostas contidos no grupo g , r_{ug} o número de ocorrências de cada padrão u , e $S_g = 2^{I_g}$ o número máximo possível de diferentes padrões que podem ser observados com I_g itens, segue que

$$s_g = \min(H_g, S_g), \quad (3.33)$$

e que

$$\sum_{u=1}^{s_g} r_{ug} = H_g. \quad (3.34)$$

Com essas definições, a Equação (3.32) pode ser reescrita em função de padrões de respostas, ou seja,

$$L(\boldsymbol{\zeta}, \boldsymbol{\eta}_k | \mathbf{Y}_{\dots}) = P(\mathbf{Y}_{\dots} | \boldsymbol{\zeta}, \boldsymbol{\eta}_{\dots}) = \prod_{g=1}^G \left[M \prod_u^{s_g} (P(Y_{.ug} | \boldsymbol{\eta}_{hkg}, \boldsymbol{\zeta}_i))^{r_{ug}} \right], \quad (3.35)$$

em que M é o coeficiente multinomial definido por:

$$M = \frac{H_g!}{\sum_{u=1}^{s_g} r_{ug}!}. \quad (3.36)$$

A verossimilhança descrita na Equação (3.35), equivalente à descrita na Equação (3.32), é também função dos parâmetros dos itens e dos parâmetros latentes associados a cada grupo, e os valores estimados desses parâmetros são os valores que a maximizam.

O modelo de grupos múltiplos pode ser facilmente adaptado para modelar o comportamento longitudinal de um mesmo grupo de indivíduos. Para isso, assume-se que o grupo inicial g_1 no momento t_1 se transformou em outro grupo, g_2 , com parâmetros latentes distintos no momento t_2 , e que o grupo g_2 se transformou em g_3 no momento t_3 , e assim por diante. Ou seja, a mesma modelagem que culminou na Equação (3.32), pode ser utilizada simplesmente substituindo o índice g , $g = 1, \dots, G$, pelo índice t , $t = 1, \dots, T$, com $G = T$. Nestas condições é importante salientar que o modelo não impõe nenhuma relação entre os parâmetros latentes dos diferentes grupos, mesmo que seja razoável esperar que exista relação entre os parâmetros latentes de um mesmo grupo avaliado em momentos distintos. Ainda assim, como os mesmos indivíduos têm seus traços latentes avaliados nos diferentes momentos - seus dados são pareados - há covariância entre os conjuntos de traços latentes individuais estimados.

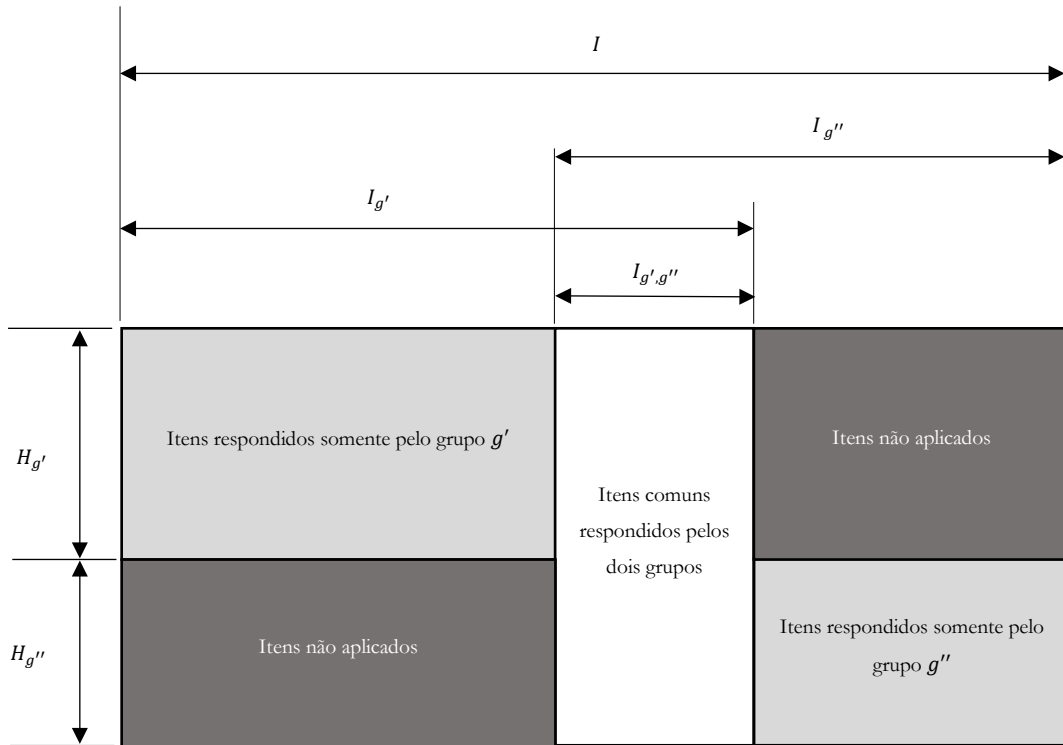


Figura 3.15 Equalização de grupos não equivalentes via itens comuns – Equalização simultânea

3.6.3 Equalização a posteriori

Quando um mesmo grupo de itens é submetido a indivíduos advindos de diferentes populações, como os itens comuns hachurados do caso (b) da Figura 3.14, seus parâmetros são estimados em escalas diferentes. Seja η_{hkt}^E , o traço latente do indivíduo h na dimensão k no momento t na escala E , ζ_i^E , o conjunto de parâmetros do item i na escala E , segundo Kolen e Brennan [17], o processo de equalização de escalas da TRI é feito de tal maneira que vale a seguinte relação para P_{ihkt} da Equação (2.5):

$$P(Y_{it} = 1 | \eta_{hkt}^{E1}, \zeta_i^{E1}) = P(Y_{it} = 1 | \eta_{hkt}^{E2}, \zeta_i^{E2}). \quad (3.37)$$

A Equação (3.37) é válida para escalas construídas com os modelos ML1P, ML2P, M3P, e M4P desde que existam as seguintes transformações lineares entre os traços latentes e os parâmetros dos itens nas diferentes escalas:

$$\eta_{hkt}^{E1} = A\eta_{hkt}^{E2} + B \quad (3.38)$$

$$a_i^{E1} = \frac{a_i^{E2}}{A} \quad (3.39)$$

$$b_i^{E1} = Ab_i^{E2} + B \quad (3.40)$$

$$c_i^{E1} = c_i^{E2} \quad (3.41)$$

$$\gamma_i^{E1} = \gamma_i^{E2} . \quad (3.42)$$

Ao substituírmos as cinco equações acima no lado direito da Equação (3.37), como abaixo, facilmente se demonstra essa igualdade.

$$P(Y_{it} = 1 | \eta_{hkt}^{E1}, \zeta_i^{E1}) = c_i^{E1} + (\gamma_i^{E1} - c_i^{E1}) \frac{e^{Da_i^{E1}(\eta_{hkt}^{E1} - b_i^{E1})}}{1 + e^{Da_i^{E1}(\eta_{hkt}^{E1} - b_i^{E1})}} \quad (3.43)$$

$$P(Y_{it} = 1 | \eta_{hkt}^{E1}, \zeta_i^{E1}) = c_i^{E2} + (\gamma_i^{E2} - c_i^{E2}) \frac{e^{\frac{Da_i^{E2}}{A}(A\eta_{hkt}^{E2} + B - Ab_i^{E2} - B)}}{1 + e^{\frac{Da_i^{E2}}{A}(A\eta_{hkt}^{E2} + B - Ab_i^{E2} - B)}} \quad (3.44)$$

$$c_i^{E2} + (\gamma_i^{E2} - c_i^{E2}) \frac{e^{Da_i^{E2}(\eta_{hkt}^{E2} - b_i^{E2})}}{1 + e^{Da_i^{E2}(\eta_{hkt}^{E2} - b_i^{E2})}} = P(y_{iht} = 1 | \eta_{hkt}^{E2}, \zeta_i^{E2}) \quad (3.45)$$

Assim, o conhecimento das constantes A e B definem o conjunto de equações necessário para transpor parâmetros e traços latentes expressos na escala $E1$ para a escala $E2$. Para quaisquer dois itens i e i' , e quaisquer dois indivíduos h e h' , a constante A pode ser calculada por, respectivamente, a partir das Equações (3.38), (3.39) e (3.40), como:

$$A = \frac{a_i^{E2}}{a_i^{E1}} , \quad (3.46)$$

$$A = \frac{b_i^{E1} - b_{i'}^{E1}}{b_i^{E2} - b_{i'}^{E2}} , \quad (3.47)$$

$$A = \frac{\eta_{hkt}^{E1} - \eta_{h'kt}^{E1}}{\eta_{hkt}^{E2} - \eta_{h'kt}^{E2}} . \quad (3.48)$$

Aplicando a Equação (3.46) para todos os N itens comuns, somando todas as N equações resultantes e dividindo por N , chega-se à fórmula das médias para A ,

$$A = \frac{\mu(a_i^{E2})}{\mu(a_i^{E1})}. \quad (3.49)$$

Convenientemente, fazendo b_i^1 igual à média de todos os parâmetros b_i^{E1} dos itens comuns na escala $E1$, $\mu(b_i^{E1})$, e, similarmente, substituindo b_i^{E2} por $\mu(b_i^{E2})$, pode-se reescrever a Equação (3.47):

$$A = \frac{b_i^{E1} - \mu(b_i^{E1})}{b_i^{E2} - \mu(b_i^{E2})}. \quad (3.50)$$

Em seguida, elevando a Equação (3.50) ao quadrado e aplicando para todos os N itens comuns, obtém-se N equações que, se somadas e divididas por N , chega-se à fórmula dos desvios padrões para A ,

$$A = \frac{\sigma(b_i^{E1})}{\sigma(b_i^{E2})}. \quad (3.51)$$

Transformações similares aplicadas à Equação (3.48) derivam a fórmula dos desvios padrão para A utilizando os traços latentes, ou seja,

$$A = \frac{\sigma(\eta_{hkt}^{E1})}{\sigma(\eta_{hkt}^{E2})}. \quad (3.52)$$

Já a constante B é calculada pela seguinte equação, a partir das Equações (3.38) e (3.40),

$$B = b_i^{E1} - Ab_i^{E2}, \quad (3.53)$$

$$B = \eta_{hkt}^{E1} - A\eta_{hkt}^{E2}. \quad (3.54)$$

Aplicando a Equação(3.53) para todos os N itens comuns, somando todas as N equações e dividindo por N , chega-se à fórmula das médias para B ,

$$B = \mu(b_i^{E1}) + A\mu(b_i^{E2}). \quad (3.55)$$

Similarmente, a partir da Equação (3.54), chega-se à fórmula dos desvios padrões para B utilizando os traços latentes, ou seja,

$$= \mu(\eta_{hkt}^{E1}) + A\mu(\eta_{hkt}^{E2}). \quad (3.56)$$

O método *mean-sigma* proposto por Marco [29] utiliza as Equações (3.51) e (3.55), já o método *mean mean* proposto por Loyd e Hoover [38] utiliza as Equações (3.49) e (3.57), e, por fim, a variante do método *mean mean* utilizando traços latentes discutido por Kolen e Brennan [17] utiliza as Equações (3.52) e (3.56). A Tabela 3.2 sumariza esses métodos e equações utilizadas.

Tabela 3.2 - Métodos de equalização a posteriori

Método	A	B
Mean/Mean (Loyd & Hoover, 1980)	$\frac{\mu(a_i^{E2})}{\mu(a_i^{E1})}$	$\mu(b_i^{E1}) + A\mu(b_i^{E2})$
Mean/Sigma (Marco, 1977)	$\frac{\sigma(b_i^{E1})}{\sigma(b_i^{E2})}$	$\mu(b_i^{E1}) + A\mu(b_i^{E2})$
Mean/Sigma (Kolen & Brennan, 2014)	$\frac{\sigma(\eta_{hkt}^{E1})}{\sigma(\eta_{hkt}^{E2})}$	$\mu(\eta_{hkt}^{E1}) + A\mu(\eta_{hkt}^{E2})$

Capítulo 4 Modelos de equações estruturais

No statistical model is real [...], nor is it intended to be. Rather, it aims to capture part of reality to a sufficient degree to give genuine insight and to provide a tool for changing things. Latent variable models do just that.

Bartholomew, Knott, Moustaki (2011)

Neste capítulo será abordada a teoria de Modelos de Equações Estruturais (MEE), já comparando sua estrutura com a TRI para alguns casos particulares, a fim de, posteriormente, realizar empiricamente a comparação dos resultados dos dois métodos. É objetivo deste trabalho estender tais relações entre a TRI e MEE para situações mais gerais, como no caso de estudos longitudinais em que testes são aplicados aos mesmos estudantes em diferentes momentos.

4.1 Introdução

A MEE corresponde a um conjunto de procedimentos estatísticos, os quais podem ser aplicados a dados quantitativos bem como qualitativos, e que permite ao pesquisador testar modelos teóricos das relações (a) entre indicadores e variáveis latentes, (b) entre conjuntos de variáveis latentes e (c) entre conjuntos de indicadores (*path analysis*). A simplicidade de suas premissas e a disponibilidade de *softwares* bastante amigáveis justificam a crescente adoção dessas técnicas em áreas tão variadas quanto a Psicologia Organizacional e do Trabalho, a Psicologia Social, Marketing, Gestão de Pessoas, Economia, entre outras [39], [70].

Especificar um modelo de equações estruturais se resume a relacionar os indicadores com suas variáveis latentes no que se denomina equações de mensuração e a relacionar as diferentes variáveis latentes do sistema em equações estruturais. Os MEE têm como objetivo analítico a aproximação da matriz de covariância populacional por meio das covariâncias amostrais associadas às equações de mensuração, cujos parâmetros expressam uma maior ou menor correlação entre as variáveis observáveis e as latentes. Se a matriz de covariância implicada a partir do conjunto de parâmetros do modelo reproduz de maneira satisfatória a matriz de covariância observada, o modelo especificado é dito plausível. Modelos plausíveis indicam que a especificação adotada gerou resíduos reduzidos e que os dados corroboram a teoria. Do contrário, o modelo é dito não plausível e uma revisão teórica da estrutura proposta é necessária - Pilati e Laros [40]. Os MEE também podem ser expandidos para modelar as médias das variáveis latentes o que permite avaliar a diferença entre as médias de diferentes grupos ou as médias de um mesmo grupo em estudos longitudinais [29].

Detalhes das equações que definem um modelo MEE são discutidos na Seção 4.2 deste capítulo. A Seção 4.3 apresenta algoritmos de estimações dos modelos, a Seção 4.4 discute medidas de ajuste e a última seção finaliza com uma discussão sobre invariância de mensuração.

4.2 Modelagem MEE

4.2.1 Cenário unidimensional transversal

A Figura 2.3 mostra os tipos de modelos que serão abordados nesta seção. Nos modelos unidimensionais transversais todos os indicadores estão envolvidos na mensuração e análise de uma única variável latente. A Equação (2.6), reproduzida na expressão (4.1), é a base do modelo do fator comum e expressa a relação linear existente entre a variável observável Y_{it} e a variável latente η_{kt} ,

$$Y_{it} = \nu_{it} + \lambda_{it}\eta_{kt} + \varepsilon_{it}, \quad (4.1)$$

em que ν_{it} e λ_{it} são, respectivamente, o intercepto e a carga fatorial do item i no momento t . A variável aleatória ε_{it} define a distribuição dos erros de mensuração e possui densidade $f(\varepsilon_{it} | \boldsymbol{\pi}_{\varepsilon_{it}})$ definida pelo vetor $\boldsymbol{\pi}_{\varepsilon_{it}}$ de parâmetros.

Se as variáveis observáveis forem ordinais categóricas, seus valores dependem dos valores de uma variável latente subliminar Y_{it}^* que se relaciona linearmente com η_{kt} como a seguir:

$$Y_{it}^* = \nu_{it} + \lambda_{it}\eta_{kt} + \varepsilon_{it}, \quad (4.2)$$

sendo que a relação entre Y_{iht} e Y_{it}^* se dá pelo modelo do valor limiar:

$$Y_{iht} = \begin{cases} 0 & \text{se } -\infty < Y_{it}^* \leq \tau_{it} \\ 1 & \text{se } \tau_{it} < Y_{it}^* < +\infty \end{cases}, \quad (4.3)$$

em que τ_{it} é um valor de corte estabelecido.

A Figura 4.1 mostra graficamente a relação entre Y_{it}^* e o valor de corte, τ_{it} , adotado, tal que,

$$\begin{aligned} P(Y_{iht} = 0) &= P(Y_{iht} = 0) = P(Y_{it}^* \leq \tau_{it}) \\ P(Y_{iht} = 1) &= P(Y_{iht} = 1) = P(Y_{it}^* > \tau_{it}) \end{aligned} \quad (4.4)$$

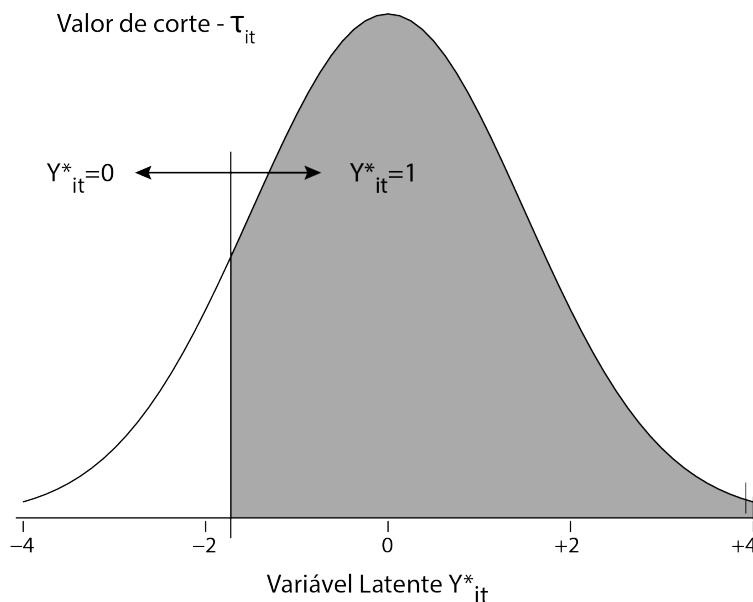


Figura 4.1 Relação entre τ_{it} e Y_{it}^* segundo o modelo de valor limiar / valor de corte

A partir da Equação (4.1) pode-se facilmente deduzir a equação da média de Y_{it} , $E(Y_{it}) = \mu_{it}$, em função da média do fator comum, $E(\eta_{kt}) = \alpha_{kt}$, e a média de ε_{it} , $E(\varepsilon_{it})$.

$$\begin{aligned} E(Y_{it}) &= E(v_{it}) + E(\lambda_{it}\eta_{kt}) + E(\varepsilon_{it}), \\ &= v_{it} + \lambda_{it}E(\eta_{kt}) + E(\varepsilon_{it}), \\ \mu_{it} &= v_{it} + \lambda_{it}\alpha_{kt} + E(\varepsilon_{it}). \end{aligned} \quad (4.5)$$

Como é conveniente assumir que a distribuição dos erros de mensuração tem média zero, a Equação (4.5) é frequentemente encontrada na seguinte forma:

$$\mu_{it} = v_{it} + \lambda_{it}\alpha_{kt} \quad (4.6)$$

Similarmente, a partir da Equação (4.1) pode-se facilmente deduzir a equação da variância de Y_{it} , $VAR(Y_{it}) = \sigma_{it}^2$, em função da variância do fator comum, $VAR(\eta_{kt}) = \psi_{kt}$, e a variância de ε_{it} , $VAR(\varepsilon_{it}) = \theta_i$. Utilizando as igualdades (C.14) e (C.16) do Apêndice C, obtemos:

$$\begin{aligned}
 VAR(Y_{it}) &= VAR(v_{it} + \lambda_{it}\eta_{kt} + \varepsilon_{it}) \\
 &= VAR(v_{it}) + VAR(\lambda_{it}\eta_{kt}) + VAR(\varepsilon_{it}) \\
 &= 0 + \lambda_{it}^2 VAR(\eta_{kt}) + VAR(\varepsilon_{it}) \\
 \sigma_{it}^2 &= \lambda_{it}^2 \psi + \theta_{it}.
 \end{aligned} \tag{4.7}$$

Já o cálculo da covariância entre variáveis latentes Y_{it} e Y_{jt} associadas às respostas aos indicadores distintos i e j se dá a partir da combinação da Equação (C.10) com a Equação (4.1) e fazendo $v_{it} = v_{jt} = 0$, temos:

$$\begin{aligned}
 COV(Y_{it}, Y_{jt}) &= E(Y_{it}Y_{jt}) - E(Y_{it})E(Y_{jt}) \\
 &= E\left((\lambda_{it}\eta_{kt} + \varepsilon_{it})(\lambda_{jt}\eta_{kt} + \varepsilon_{jt})\right) \\
 &\quad - E(\lambda_{it}\eta_{kt} + \varepsilon_{it})E(\lambda_{jt}\eta_{kt} + \varepsilon_{jt}) \\
 &= E(\lambda_{it}\lambda_{jt}\eta_{kt}^2 + \lambda_{it}\eta_{kt}\varepsilon_{jt} + \lambda_{jt}\eta_{kt}\varepsilon_{it} + \varepsilon_{it}\varepsilon_{jt}) \\
 &\quad - E(\lambda_{it}\eta_{kt} + \varepsilon_{it})E(\lambda_{jt}\eta_{kt} + \varepsilon_{jt})
 \end{aligned} \tag{4.8}$$

Assumindo que a média dos erros é comumente fixada em zero e que η_{kt} , ε_{it} e ε_{jt} são independentes entre si, a partir da Equação (C.4), temos que:

$$\begin{aligned}
 E(\eta_{kt}\varepsilon_{mt}) &= E(\eta_{kt})E(\varepsilon_{mt}) = E(\eta_{kt})0 = 0, \text{ para } m = i, j. \\
 E(\varepsilon_{it}\varepsilon_{jt}) &= E(\varepsilon_{it})E(\varepsilon_{jt}) = 0
 \end{aligned} \tag{4.9}$$

Combinando as Equações (4.8) e (4.9)

$$\begin{aligned}
 COV(Y_{it}, Y_{jt}) &= E(\lambda_{it}\lambda_{jt}\eta_{kt}^2) - E(\lambda_{it}\eta_{kt})E(\lambda_{jt}\eta_{kt}) \\
 &= \lambda_{it}\lambda_{jt}E(\eta_{kt}^2) - \lambda_{it}\lambda_{jt}E(\eta_{kt})^2 \\
 &= \lambda_{it}\lambda_{jt}(E(\eta_{kt}^2) - E(\eta_{kt})^2)
 \end{aligned} \tag{4.10}$$

Finalmente, a partir da Equação (C.11), temos que:

$$\begin{aligned}
 COV(Y_{it}, Y_{jt}) &= \lambda_{it}\lambda_{jt}(E(\eta_{kt}^2) - E(\eta_{kt})^2) \\
 &= \lambda_{it}\lambda_{jt}VAR(\eta_{kt})
 \end{aligned} \tag{4.11}$$

$$= \lambda_{it}\lambda_{jt}\psi_{kt}.$$

Dado que o vetor $\mathbf{Y}_{.t} = (Y_{1t}, Y_{2t}, \dots, Y_{It})^t$ contém todas as variáveis aleatórias observadas no momento t , e $\mathbf{Y} = \mathbf{Y}_{..} = (\mathbf{Y}_{1t_1}^t, \mathbf{Y}_{.t_2}^t, \dots, \mathbf{Y}_{It}^t)^t$ contém todas as variáveis aleatórias observadas nos T momentos, e que a função $COV(X, Y)$, em letras maiúsculas, retorna a covariância populacional entre duas variáveis aleatórias, pode-se definir a matriz de covariância populacional $\mathbf{\Sigma}(\mathbf{Y})$ para um vetor \mathbf{Y} de variáveis aleatórias em função de seus elementos $\sigma_{m,n}^2$, localizados na linha m e na coluna n , tal que,

$$\sigma_{m,n}^2 = COV(Y_m, Y_n), \text{ com } m, n = 1, \dots, I \quad (4.12)$$

Ou seja, lembrando que $\mathbf{\Sigma}(\mathbf{Y})$ é simétrica, temos

$$\mathbf{\Sigma}(\mathbf{Y}) = \begin{pmatrix} COV(Y_{1t}, Y_{1t}) & COV(Y_{1t}, Y_{2t}) & \dots & COV(Y_{1t}, Y_{(I-1)t}) & COV(Y_{1t}, Y_{It}) \\ COV(Y_{2t}, Y_{1t}) & COV(Y_{2t}, Y_{2t}) & \dots & COV(Y_{2t}, Y_{(I-1)t}) & COV(Y_{2t}, Y_{It}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ COV(Y_{(I-1)t}, Y_{1t}) & COV(Y_{(I-1)t}, Y_{2t}) & \dots & COV(Y_{(I-1)t}, Y_{(I-1)t}) & COV(Y_{(I-1)t}, Y_{It}) \\ COV(Y_{It}, Y_{1t}) & COV(Y_{It}, Y_{2t}) & \dots & COV(Y_{It}, Y_{(I-1)t}) & COV(Y_{It}, Y_{It}) \end{pmatrix}. \quad (4.13)$$

Podemos reescrever (4.13) como função da variância do fator comum, das variâncias dos erros aleatórios e das cargas fatoriais dos itens utilizando as igualdades (4.7) e (4.11), ou seja,

$$\mathbf{\Sigma}(\mathbf{Y}) = \begin{pmatrix} \lambda_{1t}^2\psi_{kt} + \theta_{1t} & \lambda_{1t}\lambda_{2t}\psi_{kt} & \dots & \lambda_{1t}\lambda_{(I-1)t}\psi_{kt} & \lambda_{1t}\lambda_{It}\psi_{kt} \\ \lambda_{2t}\lambda_{1t}\psi_{kt} & \lambda_{2t}^2\psi_{kt} + \theta_{2t} & \dots & \lambda_{2t}\lambda_{(I-1)t}\psi_{kt} & \lambda_{2t}\lambda_{It}\psi_{kt} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{(I-1)t}\lambda_{1t}\psi_{kt} & \lambda_{(I-1)t}\lambda_{2t}\psi_{kt} & \dots & \lambda_{(I-1)t}^2\psi_{kt} + \theta_{(I-1)t} & \lambda_{It}\lambda_{(I-1)t}\psi_{kt} \\ \lambda_{It}\lambda_{1t}\psi_{kt} & \lambda_{It}\lambda_{2t}\psi_{kt} & \dots & \lambda_{It}\lambda_{(I-1)t}\psi_{kt} & \lambda_{It}^2\psi_{kt} + \theta_{It} \end{pmatrix}. \quad (4.14)$$

Assim, o modelo unidimensional transversal é definido pelos seguintes elementos:

- A equação de mensuração, Equação (4.1);
- A equação das médias, Equação (4.6);
- A matriz de covariância definida em função das equações das covariâncias, Equação (4.14);
- As restrições e parametrizações que garantem a identificabilidade do conjunto de equações descritas na Seção 4.4.

Nas seções seguintes discutiremos como este modelo deve ser estendido para contemplar modelos multidimensionais transversais.

4.2.2 Cenário multidimensional transversal

A Figura 2.4 mostra os tipos de modelos que serão abordados nesta seção. Esses modelos contrastam com os modelos da seção anterior na existência de mais de uma variável latente e, conseqüentemente, no fato de que os indicadores disponíveis se dividem na mensuração dessas diferentes variáveis. Isso implica na existência de uma equação de mensuração associada à cada variável latente. Se assumirmos que os indicadores Y_{it} estão relacionados à variável latente η_{k_1t} e que os indicadores X_{jt} , à variável latente η_{k_2t} , suas equações de mensuração são dadas por:

$$\begin{aligned} Y_{it} &= v_{it} + \lambda_{it}\eta_{k_1t} + \varepsilon_{it}, \quad i = 1, \dots, I, t = 1, \\ X_{jt} &= v_{jt} + \lambda_{jt}\eta_{k_2t} + \varepsilon_{jt}, \quad j = 1, \dots, J, t = 1. \end{aligned} \quad (4.15)$$

As equações das médias se derivam tal qual foi feito para a Equação(4.6)

$$\begin{aligned} \mu_{it}^y &= v_{it} + \lambda_{it}\alpha_{k_1t} \\ \mu_{jt}^x &= v_{jt} + \lambda_{jt}\alpha_{k_2t}. \end{aligned} \quad (4.16)$$

A matriz de covariância das variáveis observáveis é uma matriz simétrica de dimensões $I \times J$ que pode ser desmembrada em quatro sub-matrizes como abaixo:

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{xx} \end{pmatrix}. \quad (4.17)$$

Σ_{yy} e Σ_{xx} são matrizes quadradas com dimensões $I \times I$ e $J \times J$, respectivamente, e modeladas segundo a Equação (4.14), tal que;

$$\begin{aligned} COV(Y_{it}, Y_{jt}) &= E(Y_{it}X_{jt}) - E(Y_{it})E(X_{jt}), \\ &= E\left((\lambda_{it}\eta_{k_1t} + \varepsilon_{it})(\lambda_{jt}\eta_{k_2t} + \varepsilon_{jt})\right) \\ &\quad - E(\lambda_{it}\eta_{k_1t} + \varepsilon_{it})E(\lambda_{jt}\eta_{k_2t} + \varepsilon_{jt}), \\ &= E(\lambda_{it}\lambda_{jt}\eta_{k_1t}\eta_{k_2t} + \lambda_{it}\eta_{k_1t}\varepsilon_{jt} + \lambda_{jt}\eta_{k_2t}\varepsilon_{it} + \varepsilon_{it}\varepsilon_{jt}) \\ &\quad - E(\lambda_{it}\eta_{k_1t} + \varepsilon_{it})E(\lambda_{jt}\eta_{k_2t} + \varepsilon_{jt}). \end{aligned} \quad (4.18)$$

Assumindo que a média dos erros é comumente fixada em zero, que η_{k_1t} e η_{k_2t} são, respectivamente, independentes de ε_{it} e ε_{jt} e que os erros não são correlacionados temos que,

$$E(\eta_{k_nt}\varepsilon_{mt}) = E(\eta_{k_nt})E(\varepsilon_{mt}) = E(\eta_{k_nt}) = 0, \text{ para } n = j, i, m = i, j. \quad (4.19)$$

Substituindo (4.9) em (4.8), temos que:

$$\begin{aligned} COV(Y_{it}, X_{jt}) &= E(\lambda_{it}\lambda_{jt}\eta_{k_1t}\eta_{k_2t}) - E(\lambda_{it}\eta_{k_1t})E(\lambda_{jt}\eta_{k_2t}) + \\ &\quad E(\varepsilon_{it}\varepsilon_{jt}) - E(\varepsilon_{it})E(\varepsilon_{jt}), \\ &= \lambda_{it}\lambda_{jt}[E(\eta_{k_1t}\eta_{k_2t}) - E(\eta_{k_1t})E(\eta_{k_2t})] + \\ &\quad E(\varepsilon_{it}\varepsilon_{jt}) - E(\varepsilon_{it})E(\varepsilon_{jt}), \\ &= \lambda_{it}\lambda_{jt}\psi_{k_1t,k_2t} + \theta_{it,jt}, \end{aligned} \quad (4.20)$$

na qual, ψ_{k_1t,k_2t} denota a covariância entre as variáveis latentes η_{k_1t} e η_{k_2t} , e $\theta_{it,jt}$, a covariância entre os erros de mensuração dos indicadores Y_{it} e X_{jt} , caso ela exista. A Equação (4.21) exemplifica como a Equação (4.20) é utilizada na construção da matriz Σ_{yx} .

Assim, o modelo multidimensional transversal é definido pelos seguintes elementos:

- As equações de mensuração para cada variável latente, Equação (4.15);
- A equação das médias para cada variável latente, Equação (4.16);
- As sub-matrizes de covariância Σ_{xx} e Σ_{yy} escritas em função das equações das covariâncias, Equação (4.14);
- A sub-matriz de covariância $\Sigma_{xy} = \Sigma'_{yx}$ escrita em função das equações das covariâncias, Equação (4.21);
- As restrições e parametrizações que garantem a identificabilidade do conjunto de equações, a serem consideradas na Seção 4.4.

$$\Sigma_{yx} = \begin{pmatrix} \lambda_{1t}^{k_1t} \lambda_{1t}^{k_2t} \psi_{k_1t,k_2t} + \theta_{1t,1t} & \lambda_{1t}^{k_1t} \lambda_{2t}^{k_2t} \psi_{k_1t,k_2t} & \dots & \lambda_{1t}^{k_1t} \lambda_{(j-1)t}^{k_2t} \psi_{k_1t,k_2t} & \lambda_{1t}^{k_1t} \lambda_{jt}^{k_2t} \psi_{k_1t,k_2t} \\ \lambda_{2t}^{k_1t} \lambda_{1t}^{k_2t} \psi_{k_1t,k_2t} & \lambda_{2t}^{k_1t} \lambda_{2t}^{k_2t} \psi_{k_1t,k_2t} + \theta_{2t,2t} & \dots & \lambda_{2t}^{k_1t} \lambda_{(j-1)t}^{k_2t} \psi_{k_1t,k_2t} & \lambda_{2t}^{k_1t} \lambda_{jt}^{k_2t} \psi_{k_1t,k_2t} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{(i-1)t}^{k_1t} \lambda_{1t}^{k_2t} \psi_{k_1t,k_2t} & \lambda_{(i-1)t}^{k_1t} \lambda_{2t}^{k_2t} \psi_{k_1t,k_2t} & \dots & \lambda_{(i-1)t}^{k_1t} \lambda_{(j-1)t}^{k_2t} \psi_{k_1t,k_2t} + \theta_{(i-1)t,(j-1)t} & \lambda_{(i-1)t}^{k_1t} \lambda_{jt}^{k_2t} \psi_{k_1t,k_2t} \\ \lambda_{it}^{k_1t} \lambda_{1t}^{k_2t} \psi_{k_1t,k_2t} & \lambda_{it}^{k_1t} \lambda_{2t}^{k_2t} \psi_{k_1t,k_2t} & \dots & \lambda_{it}^{k_1t} \lambda_{(j-1)t}^{k_2t} \psi_{k_1t,k_2t} & \lambda_{it}^{k_1t} \lambda_{jt}^{k_2t} \psi_{k_1t,k_2t} + \theta_{it,jt} \end{pmatrix}. \quad (4.21)$$

4.2.3 Cenário unidimensional longitudinal

O cenário multidimensional transversal da Figura 2.4 e o cenário unidimensional longitudinal da Figura 2.5 podem ser vistos como estruturalmente equivalentes no sentido de que as equações derivadas na seção anterior, caso multidimensional transversal, podem ser utilizadas na modelagem unidimensional longitudinal. Entretanto, na modelagem longitudinal deve-se atentar que as escalas das variáveis latentes nos diferentes momentos devem ser as mesmas apesar da possível diferença entre os parâmetros das distribuições latentes observadas entre esses momentos. Esse problema é resolvido utilizando itens comuns ou itens âncora entre os dois momentos e equalizando os parâmetros utilizando um dos métodos descritos na Seção 3.6. Neste caso, assume-se que os itens comuns não sofreram variância de mensuração, ou seja, seus parâmetros (carga fatorial, valor limiar⁹) e, conseqüentemente, suas propriedades psicométricas permaneceram inalteradas nas coletas de dados realizadas nos diferentes momentos. Ver Figura A.9 do Apêndice A para um exemplo sobre como itens comuns devem ser especificados em um modelo unidimensional longitudinal.

4.2.4 Cenário multidimensional longitudinal

A Figura 2.6 denota um cenário multidimensional longitudinal quando $K = 2$ e $T = 2$. Trata-se de uma simples extensão dos conceitos descritos no cenário unidimensional longitudinal em que todas as variáveis latentes η_{kt} são condensadas na variável KT -dimensional multivariada $\eta_{KT} = \eta_{..} = (\eta'_{.1}, \eta'_{.2}, \dots, \eta'_{.T})'$ no qual $\eta_{.t} = (\eta_{1k}, \eta_{2t}, \dots, \eta_{kt})$.

4.3 Estimação dos parâmetros dos modelos

Seja \mathbf{S} a matriz de covariância observada e $\mathbf{\Sigma}$ a matriz de covariância implicada pelo modelo no fim do processo de estimação, ambas de dimensão $I \times I$ para $h = 1, \dots, H$. Idealmente, almeja-se que $\mathbf{S} = \mathbf{\Sigma}$, mas, pragmaticamente, contenta-se com que a diferença entre \mathbf{S} e $\mathbf{\Sigma}$ seja a menor possível. O processo de estimação dos parâmetros nos MEE é guiado por funções de ajuste $F = F(\mathbf{\Sigma}, \mathbf{S})$ que retornam uma medida da diferença existente entre \mathbf{S} e $\mathbf{\Sigma}$ e, assim, proveem uma indicação de quanto os valores intermediários dos parâmetros do modelo, calculados através de algoritmos iterativos,

⁹ A palavra inglesa threshold tem duas possíveis traduções em português: valor limiar ou valor de corte.

estão próximos dos valores finais. Em 1984, Browne [72] mostrou que as funções de ajuste que têm as seguintes quatro propriedades geram estimadores consistentes de parâmetros:

1. $F(\boldsymbol{\Sigma}, \mathbf{S})$ é escalar;
2. $F(\boldsymbol{\Sigma}, \mathbf{S}) \geq 0$;
3. $F(\boldsymbol{\Sigma}, \mathbf{S}) = 0$ se e somente se $\mathbf{S} = \boldsymbol{\Sigma}$;
4. $F(\boldsymbol{\Sigma}, \mathbf{S})$ é contínua em \mathbf{S} e em $\boldsymbol{\Sigma}$.

As funções de ajuste para os métodos de máxima verossimilhança (ML), de mínimos quadrados não ponderados (ULS) e de mínimos quadrados generalizados (GLS) são listados a seguir. Elas são discutidas por Bollen e são apropriadas quando todas as variáveis observadas são quantitativas ou intervalares [45]. Considerando I indicadores quantitativos, tem-se:

$$F_{ML} = \log|\boldsymbol{\Sigma}| - \log|\mathbf{S}| + \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - I, \quad (4.22)$$

$$F_{GLS} = \frac{1}{2} \text{tr}\{[\mathbf{I} - \boldsymbol{\Sigma}\mathbf{S}^{-1}]^2\}, \quad (4.23)$$

$$F_{ULS} = \frac{1}{2} \text{tr}\{[\mathbf{S} - \boldsymbol{\Sigma}]^2\}. \quad (4.24)$$

Em 1984, Muthén [32] propôs o método dos *Mínimos Quadrados Ponderados Robustos* (Robust Weighted Least Squares - WLSMV) usado quando dentro das variáveis do modelo existe pelo menos uma variável categórica. Outra abordagem utilizada na estimação de modelos com variáveis observadas categóricas se refere ao método *Pairwise Likelihood*-baseado na teoria de verossimilhança composta¹⁰ [73], [14], [74], em que a verossimilhança é decomposta em produtos de verossimilhanças, cujos componentes são computacionalmente mais tratáveis. Esses métodos permitem a análise de modelos que envolvam variáveis latentes e covariâncias de erros não nulos nas equações estruturais.

¹⁰ *composite likelihood*

4.4 Identificabilidade

Esta seção explora as restrições que devem ser impostas ao modelo para garantir sua identificabilidade quando uma variável latente é observada com um, dois ou mais que três indicadores. Seja I o número de indicadores usados na mensuração de uma variável latente, o número de observações é dado por:

$$N_{obs} = \frac{I(I + 1)}{2}. \quad (4.25)$$

Um modelo é dito identificável quando N_{obs} é maior ou igual ao número de parâmetros do modelo que precisam ser estimados.

4.4.1 Identificação com um indicador por variável latente

Pode-se escrever a matriz de covariância observada \mathbf{S} a partir dos parâmetros do modelo conforme a Equação (4.14) quando somente um indicador é utilizado na mensuração de $\boldsymbol{\eta}$.

$$\begin{aligned} \mathbf{S} &= \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \\ (\hat{\sigma}_{it}^2) &= (\lambda_{it}\psi\lambda_{it} + \theta_{it}) \end{aligned} \quad (4.14)$$

É fácil observar que temos um valor observável ($\hat{\sigma}_{it}^2$) para estimar três valores desconhecidos (λ_{it} , ψ , θ_{it}). Neste caso, Little sugere uma das seguintes abordagens para identificar o modelo [29, p. 87].

1. Padroniza-se a variância da variável latente ($\psi = 1$) e assume-se que a variância do erro de mensuração é nula ($\theta_{it} = 0$).
2. Padroniza-se a variância da variável latente ($\psi = 1$) e assume-se um valor experimental para a variância do erro de mensuração a partir de valores conhecidos da fidedignidade de mensuração da variável latente na literatura¹¹ ($\theta_{it} = \sigma_{it}^2(1 - \text{fidedignidade})$).
3. Assume-se carga fatorial unitária ($\lambda_{it} = 1$) e a variância do erro de mensuração nula ($\theta_{it} = 0$).

¹¹ Fidedignidade = reliability. Fator

4. Assume-se carga fatorial unitária ($\lambda_{it} = 1$) e um valor experimental para a variância residual a partir de valores conhecidos da fidedignidade de mensuração da variável latente na literatura ($\theta_{it} = \sigma_{it}^2(1 - fidedignidade)$).

4.4.2 Identificação com dois indicadores por variável latente

Pode-se escrever a matriz de covariância observada \mathbf{S} , a partir dos parâmetros do modelo conforme a Equação (4.14) quando dois indicadores são utilizados na mensuração de η ;

$$\mathbf{S} = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}' + \mathbf{\Theta},$$

$$\begin{pmatrix} \hat{\sigma}_{it}^2 & \\ \hat{\sigma}_{it,jt}^2 & \hat{\sigma}_{jt}^2 \end{pmatrix} = \begin{pmatrix} \lambda_{it} \\ \lambda_{jt} \end{pmatrix} \psi \begin{pmatrix} \lambda_{it} & \lambda_{jt} \end{pmatrix} + \begin{pmatrix} \theta_{it} & \\ \theta_{it,jt} & \theta_{jt} \end{pmatrix}. \quad (4.14)$$

É fácil observar que temos três valores observáveis ($\hat{\sigma}_{it}^2, \hat{\sigma}_{jt}^2, \hat{\sigma}_{it,jt}^2$) no lado direito da equação, \mathbf{S} é simétrica e somente a triangular inferior é mostrada, para estimar seis valores desconhecidos ($\lambda_{it}, \lambda_{jt}, \psi, \theta_{it}, \theta_{jt}, \theta_{it,jt}$) no lado esquerdo. Neste caso, Little sugere uma das seguintes abordagens para identificar o modelo [29, p. 87].

1. Padroniza-se a variância da variável latente ($\psi = 1$), e assume-se $\theta_{it,jt} = 0$ e que $\lambda_{it} = \lambda_{jt}$.
2. Assume-se carga fatorial unitária ($\lambda_{it} = 1$), que $\theta_{it,jt} = 0$ e que $\lambda_{it} = \lambda_{jt}$.

4.4.3 Identificação com três ou mais indicadores por variável latente

Pode-se escrever a matriz de covariância observada \mathbf{S} a partir dos parâmetros do modelo conforme a Equação (4.14) quando três ou mais indicadores são utilizados na mensuração de η .

$$\mathbf{S} = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}' + \mathbf{\Theta}$$

$$\begin{bmatrix} \hat{\sigma}_{it}^2 & & \\ \hat{\sigma}_{jt,it}^2 & \hat{\sigma}_{jt}^2 & \\ \hat{\sigma}_{kt,it}^2 & \hat{\sigma}_{kt,jt}^2 & \hat{\sigma}_{kt}^2 \end{bmatrix} = \begin{bmatrix} \lambda_{it} \\ \lambda_{jt} \\ \lambda_{kt} \end{bmatrix} \psi \begin{bmatrix} \lambda_{it} & \lambda_{jt} & \lambda_{kt} \end{bmatrix} + \begin{bmatrix} \theta_{it} & & \\ \theta_{jt,it} & \theta_{jt} & \\ \theta_{kt,it} & \theta_{kt,jt} & \theta_{kt} \end{bmatrix} \quad (4.14)$$

É fácil observar que existem seis valores observáveis no lado direito da equação, \mathbf{S} é simétrica e somente a triangular inferior é mostrada, para estimar dez valores desconhecidos no lado esquerdo.

Neste caso, Little [29, p. 88] e Kline [39, p. 203] sugerem assumir que variâncias residuais não sejam correlacionadas e que a escala do fator comum seja determinada por uma das parametrizações sugeridas na Seção D.6.1 do Apêndice D: *factor parametrization*, *referencee indicator* e *effects coding*. Para $I \geq 4$ é fácil observar que o número de valores observáveis N_{obs} , dado pela Equação (4.25), é maior que o número de incógnitas no lado direito da Equação (4.14) quando é apropriado assumir que os erros de mensuração não se correlacionam.

4.5 Parametrização

No Apêndice D, Seção D.6, foi feita uma revisão da literatura e se resumiu 18 possíveis parametrizações que podem ser utilizados no equacionamento de modelos MEE. Nesta etapa do trabalho, primeiro, elucidou-se o jargão utilizado na MEE em que frequentemente diferentes termos se referem ao mesmo conceito. Isso pode gerar confusão ou ambiguidade na comunicação entre pesquisadores. Por exemplo, uma das alternativas para se fixar a escala do fator comum é fixar sua variância na unidade, essa parametrização é chamada de *Unit Variance Identification* (UVI) por Kline [39], de *Standardized Factor* por Kamata e Bauer [36] e de *fixed factor method* por Little [29]. Um iniciante no ramo certamente se beneficiará destes esclarecimentos. Segundo, generalizou-se a fórmula proposta Kamata e Bauer em 2008 [36] para aumentar o número de parametrizações suportadas por ela de 4 para 18. Como na MEE, a escolha da parametrização é bastante flexível, leva a diferentes estimativas de parâmetros e afeta a interpretabilidade dos resultados, um usuário não cuidadoso pode facilmente aplicar a fórmula num contexto não suportado. Este trabalho identificou e sistematizou 18 diferentes maneiras de se parametrizar um modelo unidimensional e generalizou a fórmula original para suportar as diferentes parametrizações identificadas.

4.6 Medidas de ajuste

Nesta seção, diferentes medidas de ajuste propostas na literatura são introduzidas e limitações e aplicabilidade são revistas e comentadas.

Little [29, p. 107] argumenta que essas medidas seguem dois tipos de racional: o estatístico e o de modelagem. A medida de ajuste que segue o racional estatístico é a estatística χ^2 usada para medir quanto a matriz de covariância observada \mathbf{S} diverge da matriz de covariância implicada pelo modelo $\mathbf{\Sigma}$. Todas as outras medidas seguem o racional de modelagem, que comparam o modelo a ser testado ou contra o modelo saturado, ou contra o modelo nulo, como mostra a Figura 4.2. No modelo saturado, todos os parâmetros são estimados, o número de graus de liberdade é zero e o

modelo se ajusta perfeitamente aos dados em um estado de *overfitting*. O modelo nulo está no outro extremo e provê uma medida do pior modelo possível em termos de ajuste. Neste contexto, o modelo a ser testado se encontra em algum ponto do contínuo que liga o modelo saturado ao nulo. Obviamente, almeja-se um modelo que consiga explicar os dados ou sua variância com o menor número possível de parâmetros. Os índices ou medidas de ajuste que comparam o modelo a ser testado contra o modelo saturado são denominados índices ou medidas de ajuste absoluto, enquanto as que comparam o modelo a ser testado contra o modelo nulo são denominadas índices ou medidas de ajuste relativo.

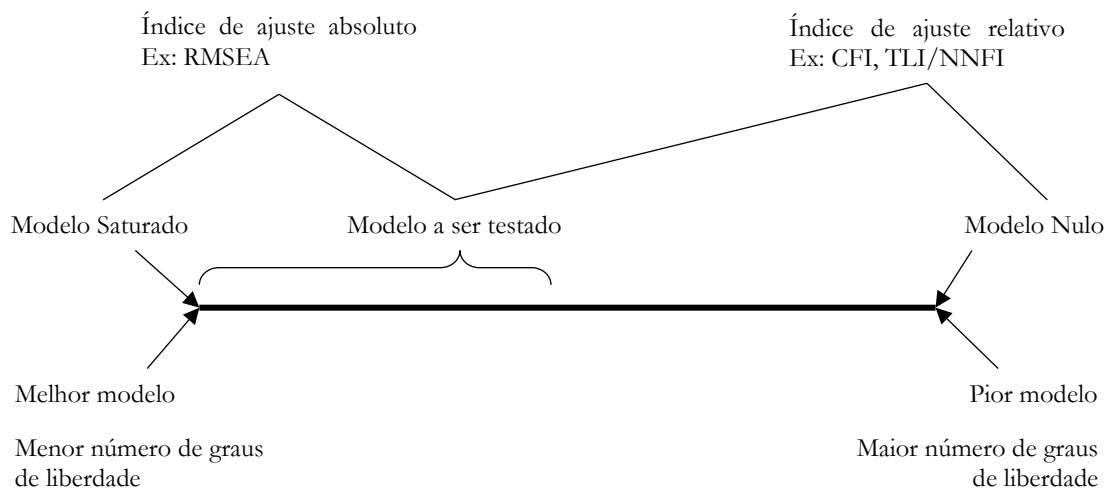


Figura 4.2 Como os modelos saturado e nulo servem de referência para cálculo de medidas de ajuste do modelo a ser testado

4.6.1 Estatística χ^2

A estatística χ^2 é usada para medir quanto a matriz de covariância observada \mathbf{S} diverge da matriz de covariância implicada pelo modelo $\mathbf{\Sigma}$, e rejeita-se o modelo se há diferenças estatisticamente significantes entre essas matrizes. A estatística χ^2 é calculada utilizando os valores retornados pelas funções de ajustes discutidas na Seção 4.3 conforme abaixo [39], [75]:

$$N * F(\boldsymbol{\Sigma}, \boldsymbol{S})$$

ou,

(4.26)

$$(N - 1) * F(\boldsymbol{\Sigma}, \boldsymbol{S}).$$

O valor de χ^2 pode ser afetado por diferentes fatores: (a) dados com distribuição multivariada não-normal podem aumentar ou diminuir o valor de χ^2 , (b) alta correlação entre variáveis observáveis aumenta o valor de χ^2 ao permitir maiores discrepâncias entre as correlações estimadas e observadas, (c) variáveis observáveis com variância específica alta indicam uma alta probabilidade de falta de ajuste local e, por fim, (d) tamanho da amostra como indicado pela Equação (4.26). Segundo Schumacker, é conhecida a tendência dessa estatística de reportar níveis significantes de probabilidades para amostras grandes (maiores que 200) e de reportar níveis não significantes para amostras pequenas (menores que 100) [75, p. 113].

4.6.2 Raiz do erro quadrático médio de aproximação (RMSEA)¹²

Proposto por Steiger e Lind [76], o RMSEA fornece uma medida de falta de ajuste do modelo por grau de liberdade. Na Equação (4.27), χ_t^2 , df_t e N indicam, respectivamente, o valor de χ^2 para o modelo a ser testado, o número de graus de liberdade do modelo a ser testado e o tamanho da amostra. Como o valor esperado de χ_t^2 é igual ao número de graus de liberdade (df_t), a falta de ajuste é igual ao valor de χ_t^2 abaixo ou acima desse valor esperado. Essa falta de ajuste, $\chi_t^2 - df_t$, denominado parâmetro de não-centralidade, é dividido pelo tamanho da amostra N menos 1, para eliminar o efeito do tamanho da amostra sobre o valor de χ_t^2

$$RMSEA = \sqrt{\frac{\frac{\chi_t^2 - df_t}{N - 1}}{df_t}} . \quad (4.27)$$

A Tabela 4.1 contém os valores de referência a serem usados na interpretação dos valores obtidos para o RMSEA segundo Little [29, p. 109].

¹² Root mean square error of approximation

Tabela 4.1 Valores de referência para interpretação do RMSEA

Ajuste	Valores de referência
Ajuste péssimo	> 0.10
Ajuste medíocre	$(0.08, 0.10]$
Ajuste aceitável	$(0.05, 0.08]$
Ajuste bom	$[0.02, 0.05]$
Ajuste perfeito	< 0.01

4.6.3 Comparative fit index (CFI)

O índice CFI é uma métrica que compara a falta de ajuste do modelo a ser testado ($\chi^2_t - df_t$) com a falta de ajuste relativamente ao modelo nulo ($\chi^2_0 - df_0$), conforme a Equação (4.28). Seu valor varia entre 0 e 1, sendo 1 seu melhor valor de ajuste. Tabela 4.2 contém os valores de referência a serem usados na interpretação dos valores obtidos para o CFI, segundo Little [29, p. 115]. O CFI é dado por:

$$CFI = 1 - \frac{\max[\chi^2_t - df_t, 0]}{\max[\chi^2_t - df_t, \chi^2_0 - df_0, 0]} \quad (4.28)$$

Tabela 4.2 Valores de referência para interpretação do CFI/TLI/NNFI

Ajuste	Valores de referência
Ajuste péssimo	< 0.85
Ajuste medíocre	[0.85, 0.90)
Ajuste aceitável	[0.90, 0.95)
Ajuste bom	[0.95, 0.99)
Ajuste perfeito	≥ 0.99

4.6.4 Tucker-Lewis index (TLI) e Non-normed fit index (NNFI)

O índice TLL/NNFI compara o valor da estatística χ^2 por grau de liberdade entre o modelo a ser testado e o modelo nulo [77], [78]. Esse índice tem dupla denominação porque foi desenvolvido por dois grupos diferentes de pesquisadores em diferentes momentos para cumprir propósitos ligeiramente distintos. Uma consequência histórica disso é que, dependendo do software adotado, essa medida é reportada de maneira distinta, `Mplus` e `lavaan` a reportam como TLI, enquanto que `LISREL` o reporta como NNFI. Seus valores são interpretados tal qual é feito para o CFI na Tabela 4.2. O índice TLI/NNFI pode exceder 1 quando χ^2_t for menor que df_t o que, segundo Little [29,

p. 116] deve ser tomado como um indício de que o modelo pode ser mais parcimonioso do que ele realmente o é. Esta métrica é dada por:

$$TLI/NNFI = \frac{\frac{\chi_0^2}{df_0} - \frac{\chi_t^2}{df_t}}{\frac{\chi_0^2}{df_0} - 1}. \quad (4.29)$$

4.7 Invariância fatorial ou de mensuração

É primordial garantir que os instrumentos de mensuração desempenhem da mesma maneira quando aplicados de maneiras diferentes em momentos diferentes para populações diferentes. Por exemplo, Whitaker e McKinney testaram se os instrumentos utilizados em pesquisas de satisfação com trabalho são invariantes sob diferentes métodos de aplicação: aplicação online via a Internet ou aplicação presencial utilizando papel e caneta [79]. Em cenários educacionais é comum que a dificuldade relativa dos itens diminua com o tempo devido, ou pela exposição frequente dos alunos a esses itens, ou porque o conteúdo avaliado por eles passou a ser melhor ensinado nas escolas após a cobrança desses conteúdos nos testes. Essa mudança no valor dos parâmetros dos itens ao longo do tempo se chama DRIFT [80], [81]. Por fim, atenção especial deve ser dispensada para que construtos não favoreçam um grupo sobre outro e que medições efetuadas por um construto sejam independentes do grupo ao qual o respondente pertence. Como a lógica para testar a invariância que existe entre diferentes métodos de aplicação, entre diferentes momentos e entre diferentes grupos é basicamente a mesma, a discussão a seguir se limitará à testagem de invariância entre diferentes grupos.

Existem quatro diferentes tipos de invariância fatorial: invariância configuracional, invariância fraca, invariância forte e invariância rígida. Essa tipagem se caracteriza por impor, nessa ordem, condições cada vez mais restritivas para atestar invariância [82], [83].

O primeiro tipo, o da invariância configuracional é o tipo mais básico e ela se mantém se o número de variáveis latentes (fatores) e a correspondência que existe entre as variáveis latentes e seus indicadores se mantém nas mensurações efetuadas em diferentes grupos. Dado que a invariância configuracional já foi atestada, a invariância fraca se mantém quando a carga fatorial dos itens de um construto se mantém nas mensurações efetuadas em diferentes grupos. Em seguida, da mesma maneira, dado que a invariância fraca foi atestada, a invariância forte se mantém quando os interceptos, no caso de variáveis numéricas, ou, valores limiares, no caso de variáveis categóricas, se mantêm nas mensurações efetuadas em diferentes grupos. Por último, dado que a invariância

forte foi atestada, a invariância rígida se mantém quando as variâncias residuais e as covariâncias se mantêm nas mensurações efetuadas em diferentes grupos. A Tabela 4.3 resume os tipos de invariância e lista as diferentes denominações encontradas na literatura.

Tabela 4.3 Invariância fatorial - Resumo

Tipo de invariância (inglês)	Característica invariante	Nomenclatura alternativa (inglês)
Invariância configural <i>(Configural invariance)</i>	Número de variáveis latentes Correspondência entre variáveis latentes e seus indicadores	Invariância de padrão <i>(Pattern invariance)</i>
Invariância fraca <i>(Weak invariance)</i>	Carga fatorial	Invariância métrica <i>(Metric invariance)</i> Invariância de carga <i>(Loading invariance)</i>
Invariância forte <i>(Strong invariance)</i>	Intercepto para variáveis observáveis numéricas Valor limiar para variáveis observáveis categóricas	Invariância escalar <i>(Scalar invariance)</i> Invariância de intercepto <i>(Intercepto invariance)</i>
Invariância rígida <i>(Strict invariance)</i>	Variância residual Covariância residual	Invariância residual <i>(Residual invariance)</i> Invariância de variância de erro <i>(Error variance invariance)</i>

Capítulo 5 Simulações

Learning is the measurement of
knowledge before and after.

Eben Pagan

5.1 Modelos transversais

5.1.1 Metodologia

5.1.1.1 O Modelo

Em simulações unidimensionais transversais ($K = 1, T = 1$), utilizando o modelo de 2 parâmetros e assumindo que o traço latente dos H indivíduos segue uma distribuição normal $\eta_{11} \sim (\mu_{11} = 0, \sigma_{11}^2 = 1)$, o conjunto de equações que definem o modelo é o seguinte.

$$Y_{i1} | (\eta_{11}, \zeta_i) \sim \text{Bernoulli}(P_{ih11}) \quad (5.1)$$

$$P_{ih11} = P(Y_{i1} = 1 | \eta_{h11}, \zeta_i) = \frac{1}{1 + e^{a_i D(\eta_{h11} - b_i)}} \quad (5.2)$$

$$\zeta_i = (a_i, b_i) \quad (5.3)$$

$$\eta_{11} | \boldsymbol{\pi}_{11} \sim N(\mu_{11}, \sigma_{11}^2) \quad (5.4)$$

5.1.1.2 Software utilizado

O pacote MIRT (versão 1.37.1) foi configurado para usar o algoritmo de esperança-maximização (EM) padrão com quadratura fixa para estimar os parâmetros dos itens utilizando o modelo logístico de dois parâmetros (ML2P). Os fatores são estimados pelo método da média da posteriori (Expected a Posteriori - EAP) também chamado de Bayes Empírico (Empirical Bayes – EB).

O pacote lavaan (versão 0.6-14) usou o algoritmo de mínimos quadrados diagonalmente ponderados (Diagonally Weighted Least Squares - DWLS) para estimar os parâmetros dos indicadores e o métodos EBM para estimar os fatores.

5.1.1.3 Geração de dados

Experimentos foram realizados simulando a mensuração de uma variável latente utilizando 13 itens. Para todos os itens foi suposto o mesmo parâmetro de discriminação, igual a 0.7, com parâmetros de dificuldade equidistantes de 0.5 uns dos outros ao longo da escala variando de -3 a +3 conforme a TABLE D.2. Essa abordagem permite mensurar toda a escala de forma homogênea

e avaliar as limitações dos métodos de estimação nos extremos da mesma. Os valores dicotômicos de entrada foram gerados seguindo os procedimentos listados abaixo.

- Gere H amostras de $\eta_{kt} \sim N(0,1)$
- Gera I_t pares de parâmetros $\zeta_{it} = (a_{it}, b_{it})$ conforme a TABLE D.2.
- Para cada indivíduo $h = 1, \dots, H$, repita
 - Para cada item $i = 1, \dots, I_t$, repita
 - Calcule a probabilidade do indivíduo h acertar o item i no momento t ,

$$P = P(\eta_{kht}, \zeta_{it})$$
 - Gere $U \sim U(0,1)$
 - $Y_{iht} = 1$ se $U \leq P$, $Y_{iht} = 0$, caso contrário.

5.1.1.4 Procedimentos

Os valores gerados \mathbf{Y}_{\dots} foram alimentados ao modelo TRI e os parâmetros de discriminação (a_{it}) e de dificuldade (b_{it}) foram estimados. \mathbf{Y}_{\dots} alimentou também os modelos MEE sob todas as parametrizações listadas na TABLE D.1 e os parâmetros dos modelos foram estimados. Os valores de a_{it} e de b_{it} correspondentes a cada conjunto de parâmetros MEE (v_{it} , τ_{it} , λ_{it} , α , ψ e θ_{it}) associado a cada parametrização foram calculados utilizando as equações (D.14) e (D.15). Finalmente, os valores estimados de a_{it} e b_{it} foram comparados aos valores simulados da TABLE D.2.

5.2 Discussão

Como mencionado anteriormente, a comparabilidade entre os modelos da TRI e dos MEE está bem estabelecida em modelos transversais (ou não-longitudinais). Takane e De Leeuw [31] provaram a equivalência entre modelos da TRI e MEE para variáveis dicotômicas baseando-se em trabalhos de Muthén [32], [33] e David John [34]. Glockner-Rist e Hoijtink [35] mostraram que ambas as teorias podem ser utilizadas na verificação da dimensionalidade de dados e em testes de invariância da medida. Por fim, Kamata e Bauer [36] resumiram as diferentes opções de parametrizações envolvidas na determinação das escalas das variáveis latentes e como essas parametrizações resultam em diferentes equações de equivalência entre a TRI e MEE.

Esta tese expande o trabalho de Kamata e Bauer no que tange a:

- (1) revisar a literatura, identificar e sistematizar 18 diferentes maneiras de se parametrizar um modelo MEE,
- (2) generalizar a fórmula proposta em [36] para aumentar o número de parametrizações suportadas por ela, de 4 para 18.

Os resultados desta pesquisa estão apresentados no Apêndice D na forma de um artigo. Estudos de simulação foram realizados, sob diferentes cenários, para avaliar a implementação computacional e o desempenho das equivalências estabelecidas.

5.3 Modelos longitudinais

5.3.1 Metodologia

5.3.1.1 O modelo

O conjunto de equações que definem o modelo utilizado em simulações multidimensionais longitudinais ($K = 2, T = 2$), ilustrado na Figura 5.4, utilizando o modelo logístico de 2 parâmetros (Equação (3.7)) e assumindo que a distribuição do vetor de traços latentes de cada indivíduo h , $h = 1, \dots, H$, da amostra, $(\eta_{h11}, \eta_{h12}, \eta_{h21}, \eta_{h22})$, segue uma distribuição normal bidimensional multivariada $\boldsymbol{\eta}_4$ é definido pelas equações abaixo listadas

$$Y_{iht} \mid (\eta_{kt}, \boldsymbol{\zeta}_i) \sim \text{Bernoulli}(P_{iht}), k = \{1,2\}, t = \{1,2\}. \quad (5.5)$$

$$P_{iht} = P(y_{iht} = 1 \mid \eta_{hkt}, \boldsymbol{\zeta}_i) = \frac{1}{1 + e^{a_i D(\eta_{hkt} - b_i)}}, k = \{1,2\}, t = \{1,2\}. \quad (5.6)$$

$$\boldsymbol{\zeta}_i = (a_i, b_i), i = \{1, \dots, I\} \quad (5.7)$$

$$\boldsymbol{\eta}_4 \sim N_4(\boldsymbol{\mu}, \boldsymbol{\Psi})$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{pmatrix} \text{ e } \boldsymbol{\Psi} = \begin{pmatrix} \sigma_{11,11}^2 & \sigma_{11,12}^2 & \sigma_{11,21}^2 & \sigma_{11,22}^2 \\ \sigma_{12,11}^2 & \sigma_{12,12}^2 & \sigma_{12,21}^2 & \sigma_{12,22}^2 \\ \sigma_{21,11}^2 & \sigma_{21,12}^2 & \sigma_{21,21}^2 & \sigma_{21,22}^2 \\ \sigma_{22,11}^2 & \sigma_{22,12}^2 & \sigma_{22,21}^2 & \sigma_{22,22}^2 \end{pmatrix} \quad (5.8)$$

sendo μ_{kt} o parâmetro de média da variável aleatória latente η_{kt} , e $\sigma_{k_1t_1, k_2t_2}^2$ a covariância entre as variáveis $\eta_{k_1t_1}$ e $\eta_{k_2t_2}$. Da normal multivariada $\boldsymbol{\eta}_4$ pode-se obter as correspondentes distribuições marginais (normais uni e bidimensionais) para as variáveis latentes unidimensionais, η_{kt} , bem como para as variáveis latentes bidimensionais, η_1 e η_2 .

Neste modelo assume-se que os parâmetros do conjunto de itens utilizados nos diferentes momentos estão calibrados na mesma escala do traço latente, ou, alternativamente, que existem itens comuns entre os dois momentos que permitam colocar os dois conjuntos em uma mesma escala através de métodos de equalização.

5.3.1.2 Software utilizado

O pacote MIRT (versão 1.37.1) foi configurado para usar o algoritmo de esperança-maximização (EM) padrão com quadratura fixa para estimar os parâmetros dos itens utilizando o modelo logístico de dois parâmetros (ML2P). Os fatores são estimados por meio de três métodos diferentes: máxima verossimilhança (Maximum Likelihood - ML), sob a notação `tri_ml`, pelo máximo da posteriori (maximum a posteriori - MAP), também chamado de Bayes Empírico Modal (Empirical Bayes Modal - EBM), sob a notação `tri_map_ebm`, e pela média da posteriori (Expected a Posteriori - EAP) também chamado de Bayes Empírico (Empirical Bayes – EB), sob a notação `tri_eap_eb` [84, p. 221].

O pacote lavaan (versão 0.6-14) usou o algoritmo de mínimos quadrados diagonalmente ponderados (Diagonally Weighted Least Squares - DWLS) para estimar os parâmetros dos indicadores e os métodos EBM e ML para estimar os fatores.

5.3.1.3 Geração de dados

Os valores dicotômicos de entrada foram gerados seguindo os procedimentos listados abaixo.

- Gera matriz $M_{H \times KT}$ com amostras de uma normal multivariada KT -dimensional $\boldsymbol{\eta} \sim N_{KT}(\boldsymbol{\mu}_{KT \times 1}, \boldsymbol{\Psi}_{KT \times KT})$
- Para cada coluna da matriz M repita
 - Gera parâmetros $\boldsymbol{\zeta}_{it} = (a_{it}, b_{it})$ para todos os itens utilizados na mensuração do k -ésimo traço latente no momento t
 - Para cada item repita

- Calcule a probabilidade do indivíduo h acertar o item i no momento t fixo.

$$P = P(\eta_{kht}, \zeta_{it})$$

- Gere $U \sim U(0,1)$

$Y_{iht} = 1$ se $U \leq P$, $Y_{iht} = 0$, caso contrário.

5.3.1.4 Valores dos parâmetros simulados

5.3.1.4.1 Número de itens simulados

Seja I o número total de itens aplicados, I_{kt} o número de itens associados à mensuração de η_{kt} e I_{kt_1,kt_2} o número de itens comuns associados à mensuração das variáveis η_{kt_1} e η_{kt_2} . Nas simulações realizadas, foi adotado,

$$\begin{aligned} I &= 100, \\ I_{11} &= I_{12} = I_{21} = I_{22} = 30, \\ I_{11,12} &= I_{21,22} = 10. \end{aligned} \tag{5.9}$$

5.3.1.4.2 Parâmetros das distribuições das variáveis latentes

Considerando a Equação (5.8) com $K = 2$ e $T = 2$, adotou-se,

$$\begin{aligned} \boldsymbol{\mu} &= \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{pmatrix} = \begin{pmatrix} 0.00 \\ 0.50 \\ 0.00 \\ 0.90 \end{pmatrix}, \\ \boldsymbol{\Psi} &= \begin{pmatrix} \sigma_{11,11}^2 & & & \\ \sigma_{12,11}^2 & \sigma_{12,12}^2 & & \\ \sigma_{21,11}^2 & \sigma_{21,12}^2 & \sigma_{21,21}^2 & \\ \sigma_{22,11}^2 & \sigma_{22,12}^2 & \sigma_{22,21}^2 & \sigma_{22,22}^2 \end{pmatrix} = \begin{pmatrix} 1 & & & \\ 0.70 & 1 & & \\ 0.35 & 0.25 & 1 & \\ 0.25 & 0.35 & 0.70 & 1 \end{pmatrix}. \end{aligned} \tag{5.10}$$

5.3.1.4.3 Parâmetros do experimento

Os valores estimados nas simulações são médias de 50 experimentos aleatórios realizados com cada teste sendo respondido por 10.000 indivíduos.

5.3.1.4.4 Parâmetros dos itens

Os valores simulados para todos os 100 itens envolvidos nas simulações estão listados na Tabela 5.1.

Tabela 5.1 Valores simulados para parâmetros dos itens

Itens	a_{it}	b_{it}	Itens	a_{it}	b_{it}
ETA11_Q001	2.717	-0.173	ETA21_Q051	2.251	0.204
ETA11_Q002	2.752	0.645	ETA21_Q052	2.935	-0.561
ETA11_Q003	2.916	1.534	ETA21_Q053	2.441	0.385
ETA11_Q004	2.222	-0.518	ETA21_Q054	2.424	1.345
ETA11_Q005	2.402	0.471	ETA21_Q055	2.004	0.95
ETA11_Q006	2.452	0.59	ETA21_Q056	2.9	1.671
ETA11_Q007	2.913	0.065	ETA21_Q057	2.784	0.666
ETA11_Q008	2.184	-0.088	ETA21_Q058	2.253	0.013
ETA11_Q009	1.964	-0.049	ETA21_Q059	2.035	1.239
ETA11_Q010	2.841	0.798	ETA21_Q060	2.836	-0.194
ETA11_Q011	2.826	0.194	ETA21_Q061	2.079	0.353
ETA11_Q012	2.01	1.094	ETA21_Q062	2.153	1.217
ETA11_Q013	2.885	0.489	ETA21_Q063	2.278	0.09
ETA11_Q014	2.007	0.105	ETA21_Q064	2.241	1.047
ETA11_Q015	2.036	0.896	ETA21_Q065	2.834	-0.128
ETA11_Q016	2.818	-0.27	ETA21_Q066	2.562	1.778
ETA11_Q017	2.937	-0.064	ETA21_Q067	2.201	1.028
ETA11_Q018	2.185	-0.461	ETA21_Q068	2.779	-0.808
ETA11_Q019	2.448	-0.355	ETA21_Q069	2.704	-1.116
ETA11_Q020	1.975	1.417	ETA21_Q070	2.254	1.768
ETA11_ETA12_Q021	2.649	-0.047	ETA21_ETA22_Q071	2.993	-0.927
ETA11_ETA12_Q022	2.8	-0.422	ETA21_ETA22_Q072	2.728	0.652
ETA11_ETA12_Q023	2.179	0.407	ETA21_ETA22_Q073	2.468	0.318
ETA11_ETA12_Q024	2.452	1.393	ETA21_ETA22_Q074	2.537	0.347
ETA11_ETA12_Q025	2.919	0.664	ETA21_ETA22_Q075	2.901	0.931
ETA11_ETA12_Q026	2.278	1.447	ETA21_ETA22_Q076	2.23	0.933
ETA11_ETA12_Q027	2.625	-1.67	ETA21_ETA22_Q077	2.977	0.09
ETA11_ETA12_Q028	2.115	0.533	ETA21_ETA22_Q078	2.579	0.919
ETA11_ETA12_Q029	2.867	0.855	ETA21_ETA22_Q079	1.933	-0.335
ETA11_ETA12_Q030	2.232	0.313	ETA21_ETA22_Q080	2.828	0.523
ETA12_Q031	2.797	0.01	ETA22_Q081	2.285	0.008

Itens	a_{it}	b_{it}	Itens	a_{it}	b_{it}
ETA12_Q032	2.754	0.716	ETA22_Q082	2.728	0.534
ETA12_Q033	2.624	-0.459	ETA22_Q083	2.345	-0.281
ETA12_Q034	2.475	0.094	ETA22_Q084	2.855	0.325
ETA12_Q035	2.403	0.106	ETA22_Q085	2.773	-0.173
ETA12_Q036	2.469	-0.099	ETA22_Q086	2.452	-0.023
ETA12_Q037	2.029	1.847	ETA22_Q087	2.739	0.079
ETA12_Q038	2.616	0.687	ETA22_Q088	2.293	0.177
ETA12_Q039	2.807	0.798	ETA22_Q089	2.206	1.924
ETA12_Q040	2.831	-0.178	ETA22_Q090	2.656	0.221
ETA12_Q041	2.625	0.354	ETA22_Q091	2.001	1.58
ETA12_Q042	2.825	-0.313	ETA22_Q092	2.605	-0.398
ETA12_Q043	2.752	1.35	ETA22_Q093	2.516	-0.988
ETA12_Q044	2.506	0.631	ETA22_Q094	2.591	-0.165
ETA12_Q045	2.263	0.855	ETA22_Q095	2.384	0.06
ETA12_Q046	2.311	-0.488	ETA22_Q096	2.698	0.913
ETA12_Q047	2.46	0.294	ETA22_Q097	2.686	-0.982
ETA12_Q048	2.719	-0.377	ETA22_Q098	2.648	-1.125
ETA12_Q049	2.043	-0.227	ETA22_Q099	2.182	0.934
ETA12_Q050	2.668	0.163	ETA22_Q100	2.989	-0.055

5.3.1.5 Procedimentos

A análise dos dados simulados com os métodos da TRI foi realizada sob a combinação de dois modelos bidimensionais longitudinais de dois momentos utilizando a abordagem de grupos múltiplos de equalização simultânea descrita na Seção 3.6.2. Um modelo unidimensional longitudinal mensura as variáveis η_{11} e η_{12} enquanto outro mensura η_{21} e η_{22} , e cada modelo unidimensional longitudinal é ajustado independentemente do outro. Ou seja, não se modelou nenhuma interação entre as variáveis latentes dos diferentes modelos. A Figura 3.15 exemplifica como os dados devem ser estruturados em uma modelagem de grupos múltiplos que utiliza itens comuns. O pacote MIRT da linguagem R foi utilizado no ajuste dos modelos, e os parâmetros estimados foram comparados com os fixados na simulação.

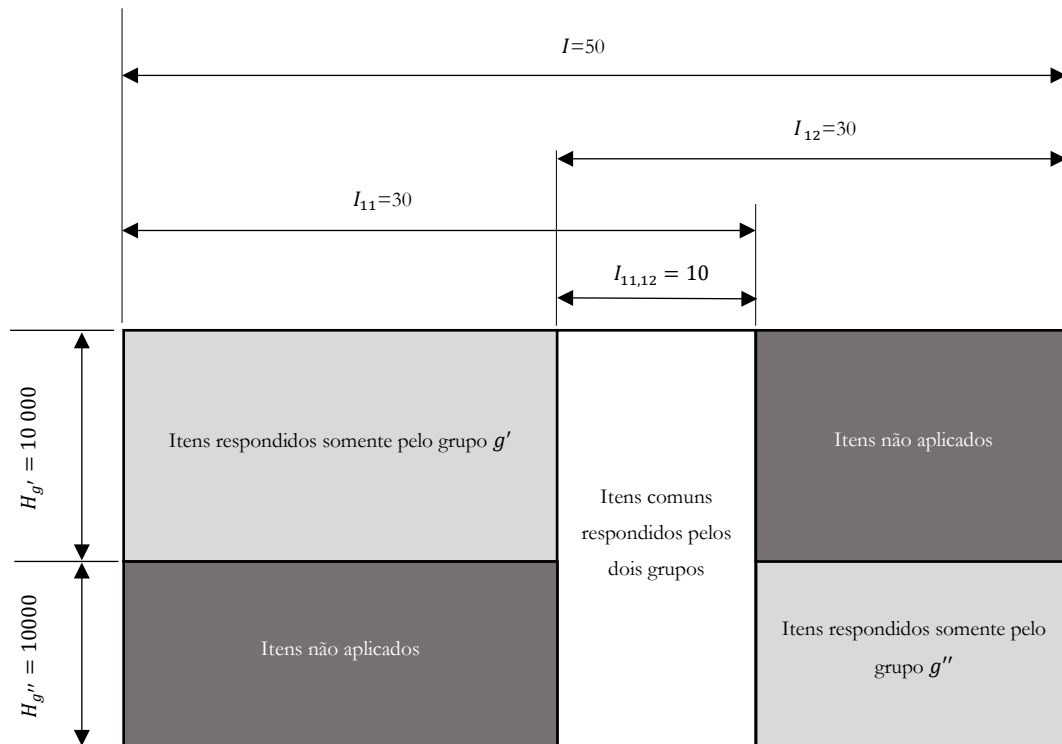


Figura 5.1 Equalização de grupos não equivalentes via itens comuns – Equalização Simultânea

Já a análise dos dados simulados com os métodos da MEE assume um modelo multidimensional longitudinal combinando dois modelos unidimensionais longitudinais ao permitir que as diferentes variáveis latentes interfiram entre si através de uma matriz de covariância não-estruturada como indicado na Figura 5.4. Neste experimento foi assumida que a invariância fatorial das cargas fatoriais e dos valores limiares dos itens comuns é válida e que existe covariância entre os itens comuns correspondentes de diferentes momentos.

A estrutura de dados passada ao pacote `lavaan` para análise dos mesmos dados simulados é substancialmente diferente da utilizada na seção anterior em que foi usado os recursos do pacote `MIRT` sob o modelo da TRI. Neste caso, por se tratar de uma modelagem longitudinal de um grupo, todos os indivíduos pertencem ao um mesmo grupo e responderam a todos os testes, ou seja, as respostas de todos os testes estão associados ao mesmo indivíduo na mesma linha do arquivo de input de dados, como ilustrado na Figura 5.5. A atual versão do pacote `lavaan` (0.6-14) não consegue processar dados faltantes sistemáticos e, por isso, esta simulação não inclui dados faltantes.

Os parâmetros do modelo foram estimados com parametrização RI_CP_FT (Reference Indicator, Conditional Parameterization, Free Threshold). De posse dos parâmetros MEE, utilizando as equações (D.14) e (D.15), os valores de a_{it} e de b_{it} foram calculados e comparados aos valores fixados na simulação.

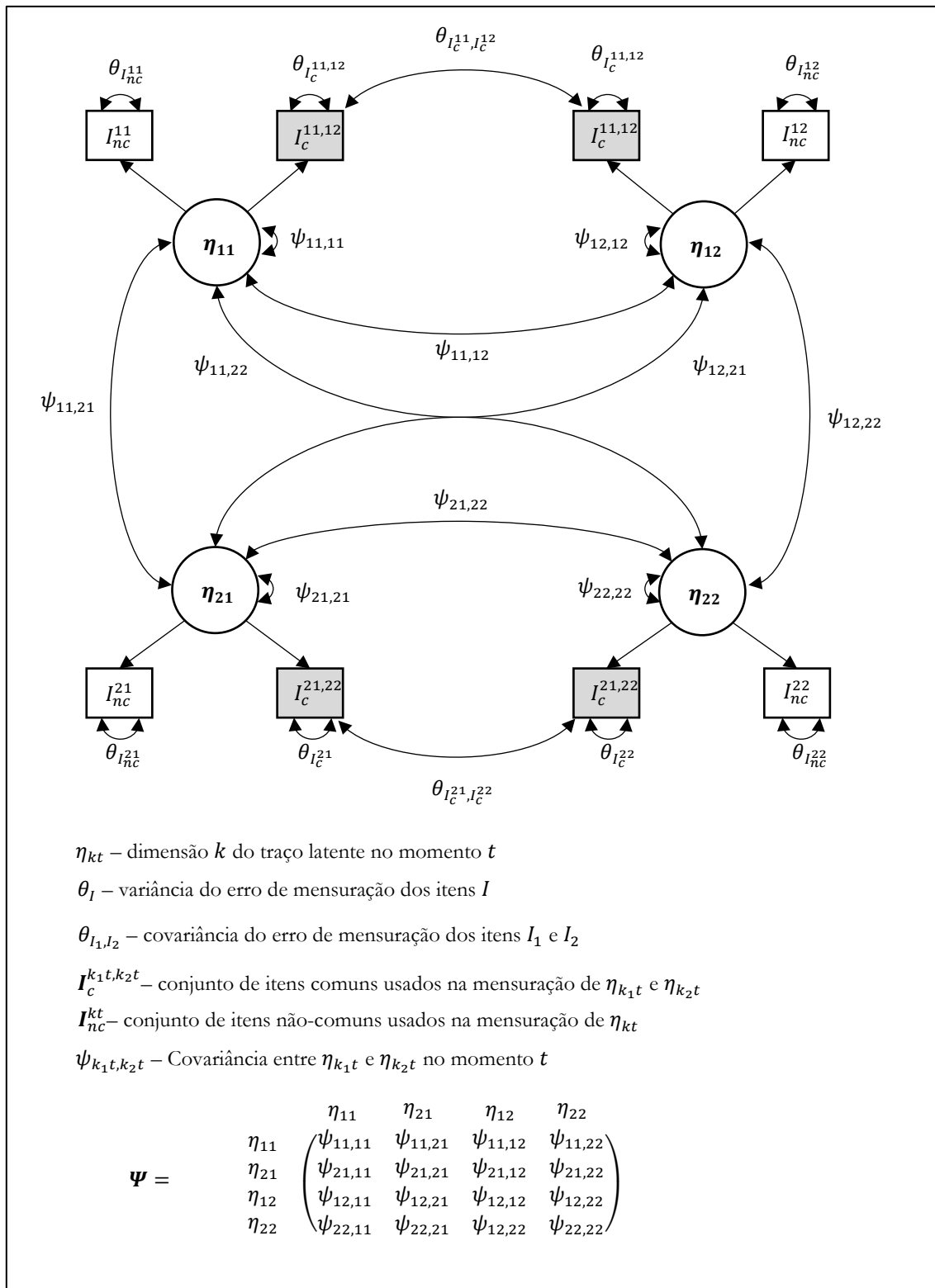


Figura 5.2 Modelo Multidimensional Longitudinal. K=2, T=2.

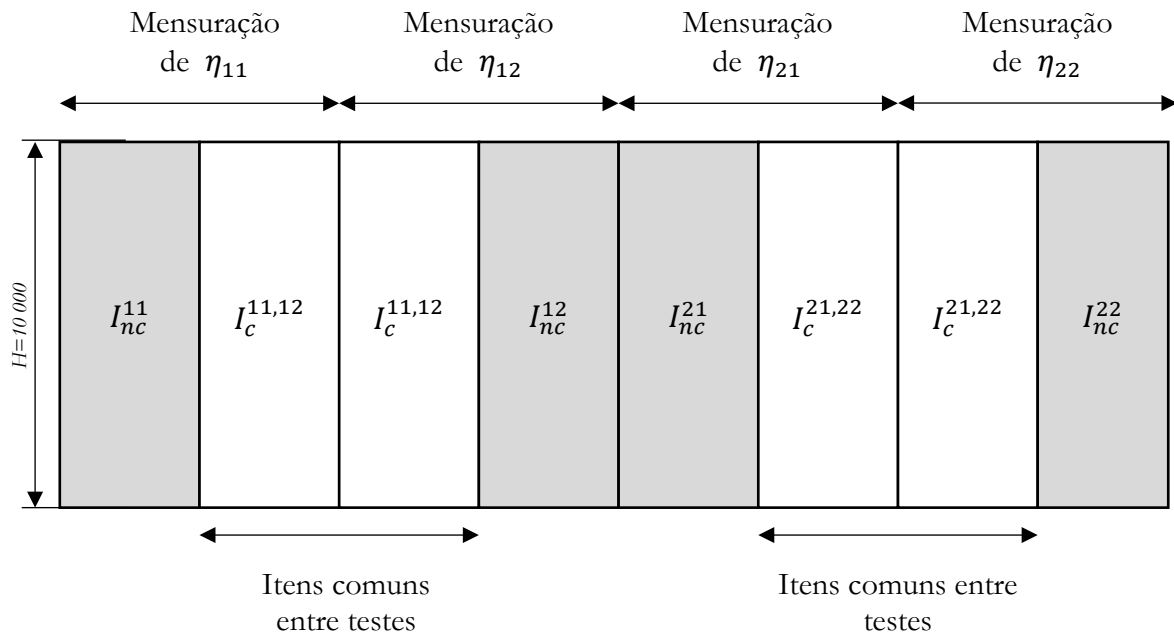


Figura 5.3 Estrutura de dados para equalização simultânea de quatro testes – modelo de um grupo.

5.3.2 Discussão

5.3.2.1 Análise utilizando a TRI

A Figura 5.2 e a Figura 5.3 apontam uma recuperação quase perfeita dos parâmetros simulados dos itens. A Tabela 5.2 mostra os parâmetros das distribuições das variáveis latentes estimados pelo modelo e note que os valores da covariância não estão disponíveis pois não fazem parte do modelo. Alternativamente, é possível estimar as escores latentes individuais e calcular a covariância e a correlação entre os vetores obtidos. Esse resultado é mostrado na Tabela 5.3 para escores latentes individuais estimados pelo método EAP.

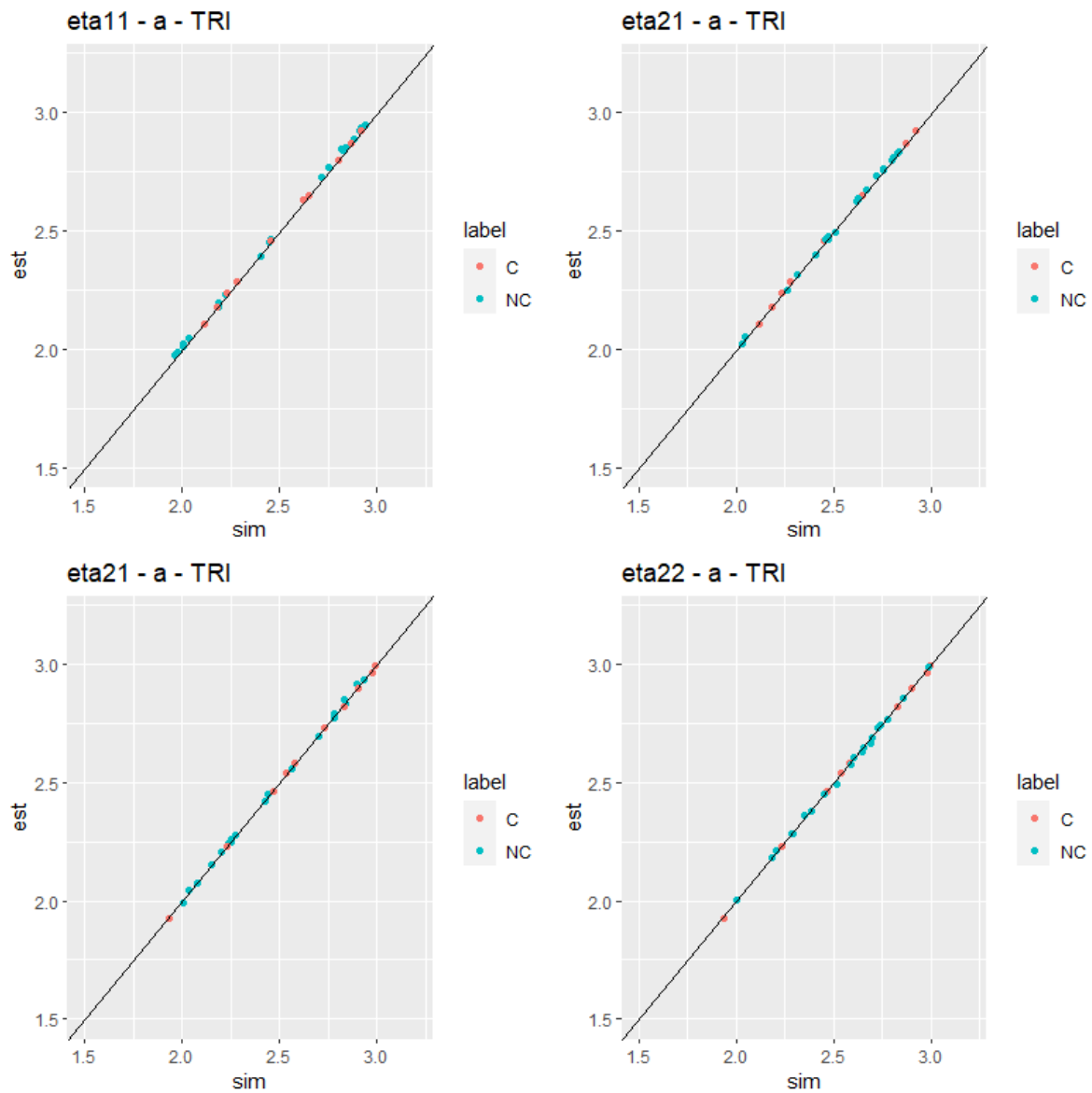


Figura 5.4 Recuperação do parâmetro de discriminação a_{it} utilizando a TRI – Modelo unidimensional longitudinal de dois momentos. C – Item comum. NC – Item não-comum. sim – valores simulados, est – valores estimados.

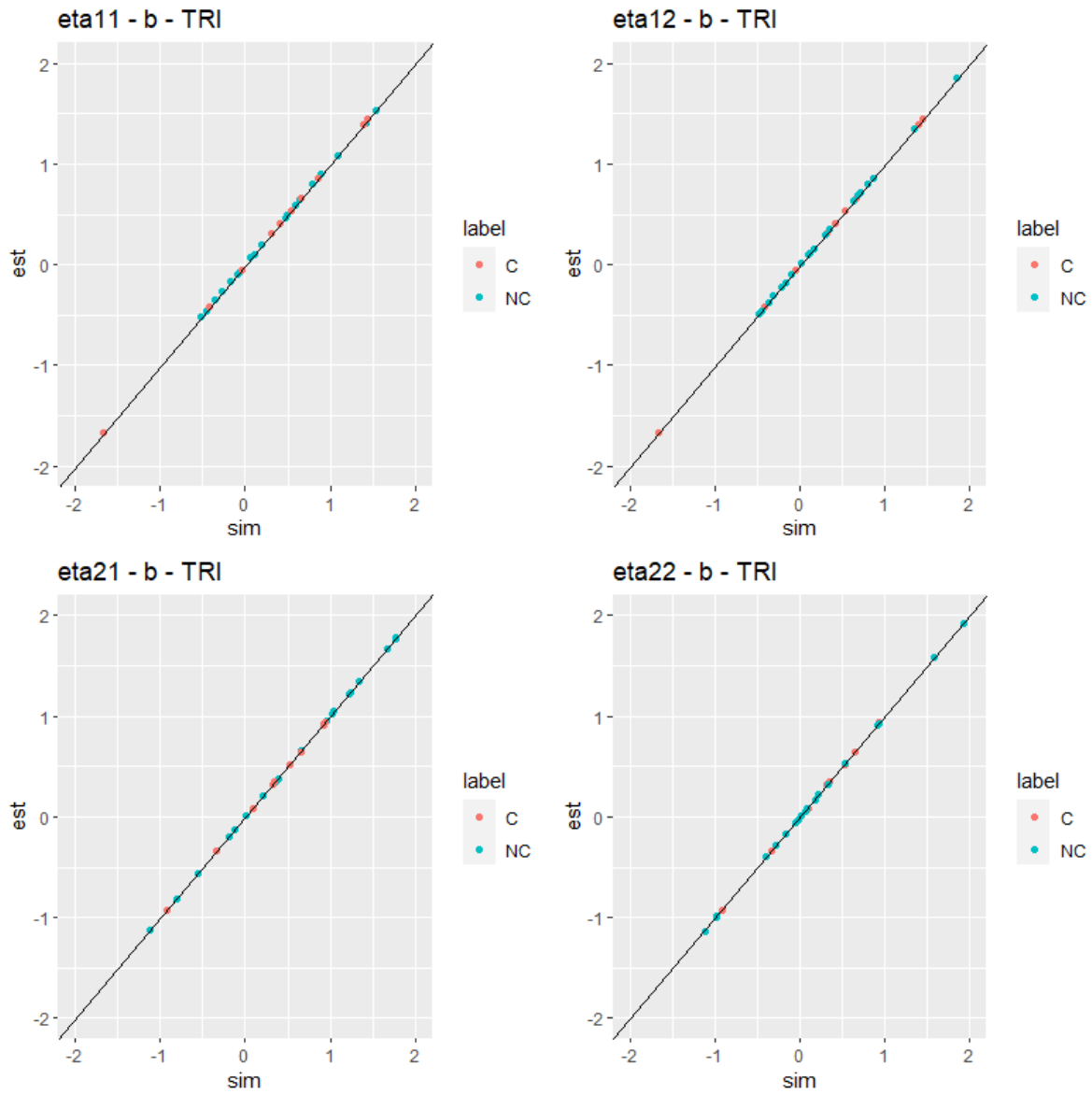


Figura 5.5 Recuperação do parâmetro de dificuldade b_{it} utilizando a TRI – Modelo unidimensional longitudinal de dois momentos. C – Item comum. NC – Item não-comum. sim – valores simulados, est – valores estimados.

Tabela 5.2 Recuperação dos parâmetros da distribuição das variáveis latentes – Modelo unidimensional ($K = 1$) longitudinal de dois momentos ($T = 2$). TRI.

Parâmetros Populacionais	Valores Simulados	Valores Estimados
Média μ	$\begin{pmatrix} 0.00 \\ 0.50 \\ 0.00 \\ 0.90 \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ (0.000) \\ 0.499 \\ (0.012) \\ 0.000 \\ (0.000) \\ 0.899 \\ (0.012) \end{pmatrix}$
Matriz de Covariância Σ	$\begin{pmatrix} 1 & & & \\ 0.70 & 1 & & \\ 0.35 & 0.25 & 1 & \\ 0.25 & 0.35 & 0.70 & 1 \end{pmatrix}$	$\begin{pmatrix} 1.000 \\ (0.000) & & & \\ NA & 1.003 \\ & (0.028) & & \\ NA & NA & 1.000 \\ NA & NA & (0.000) & 1.004 \\ & & NA & (0.030) \end{pmatrix}$

Tabela 5.3 Recuperação dos parâmetros da distribuição das variáveis latentes calculados a partir dos escores latentes individuais – Modelo unidimensional ($K = 1$) longitudinal de dois momentos ($T = 2$). TRI.

Parâmetros Populacionais	Valores Simulados	Valores Estimados
Média μ	$\begin{pmatrix} 0.00 \\ 0.50 \\ 0.00 \\ 0.90 \end{pmatrix}$	$\begin{pmatrix} -0.001 \\ (0.000) \\ 0.499 \\ (0.012) \\ -0.001 \\ (0.000) \\ 0.899 \\ (0.012) \end{pmatrix}$
Matriz de Covariância Σ	$\begin{pmatrix} 1 & & & \\ 0.70 & 1 & & \\ 0.35 & 0.25 & 1 & \\ 0.25 & 0.35 & 0.70 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.933 & & & \\ (0.000) & & & \\ 0.615 & 0.939 & & \\ (0.010) & (0.026) & & \\ 0.305 & 0.217 & 0.933 & \\ (0.008) & (0.010) & (0.000) & \\ 0.213 & 0.301 & 0.599 & 0.919 \\ (0.011) & (0.014) & (0.011) & (0.027) \end{pmatrix}$
Matriz de Correlação	$\begin{pmatrix} 1 & & & \\ 0.70 & 1 & & \\ 0.35 & 0.25 & 1 & \\ 0.25 & 0.35 & 0.70 & 1 \end{pmatrix}$	$\begin{pmatrix} 1.000 & & & \\ (0.000) & & & \\ 0.657 & 1.000 & & \\ (0.005) & (0.000) & & \\ 0.327 & 0.231 & 1.000 & \\ (0.009) & (0.010) & (0.000) & \\ 0.231 & 0.323 & 0.647 & 1.000 \\ (0.011) & (0.011) & (0.005) & (0.000) \end{pmatrix}$

5.3.2.2 Análise utilizando a MEE

As medidas de ajuste do modelo estão listadas na Tabela 5.4.

A Figura 5.6 mostra que a recuperação do parâmetro de discriminação foi inferior à obtida com os métodos da TRI. Já a Figura 5.7 aponta que a recuperação do parâmetro de dificuldade foi tão boa quanto a obtida com os métodos da TRI.

A Tabela 5.5 lista os parâmetros e estimativas da distribuição das variáveis latentes quando estas não são padronizadas, ou seja, quando suas médias e covariâncias não são definidas como 0 e 1,

respectivamente. Os resultados para este caso estão listados na Tabela 5.6. Com a padronização, os valores estimados e simulados são bastantes próximos.

Tabela 5.4 Medidas de ajuste. Modelo multidimensional longitudinal. MEE.

Descrição	Valor
Número de observações	10 000
Número de parâmetros	228
χ^2	6158.557
Graus de Liberdade	7032
valor-p	0.999
CFI	1.000
TLI	1.000
RMSEA	0.000

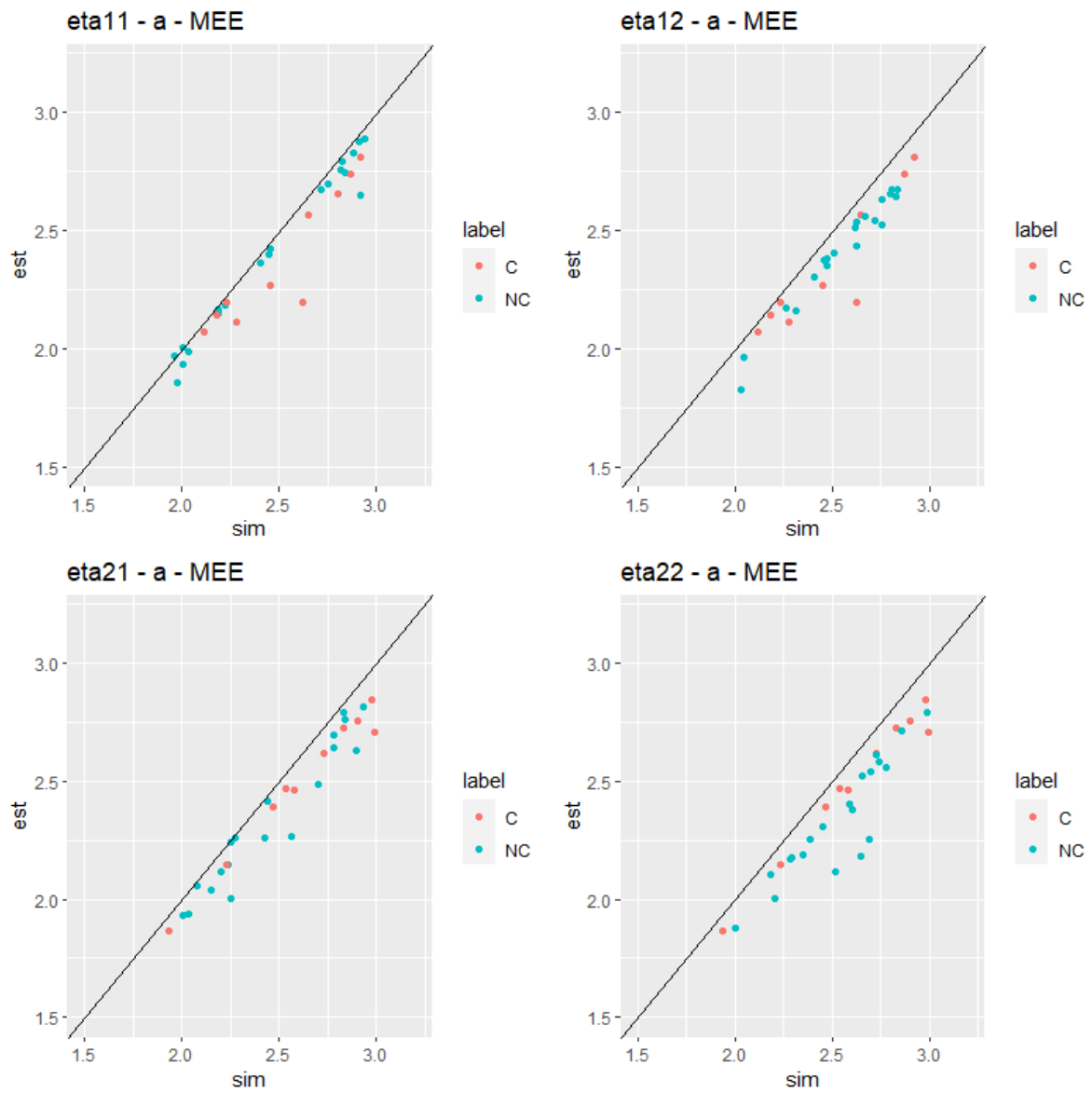


Figura 5.6 Recuperação do parâmetro de discriminação a_{it} utilizando a MEE – Modelo multidimensional longitudinal de dois momentos. C – Item comum. NC – Item não-comum.

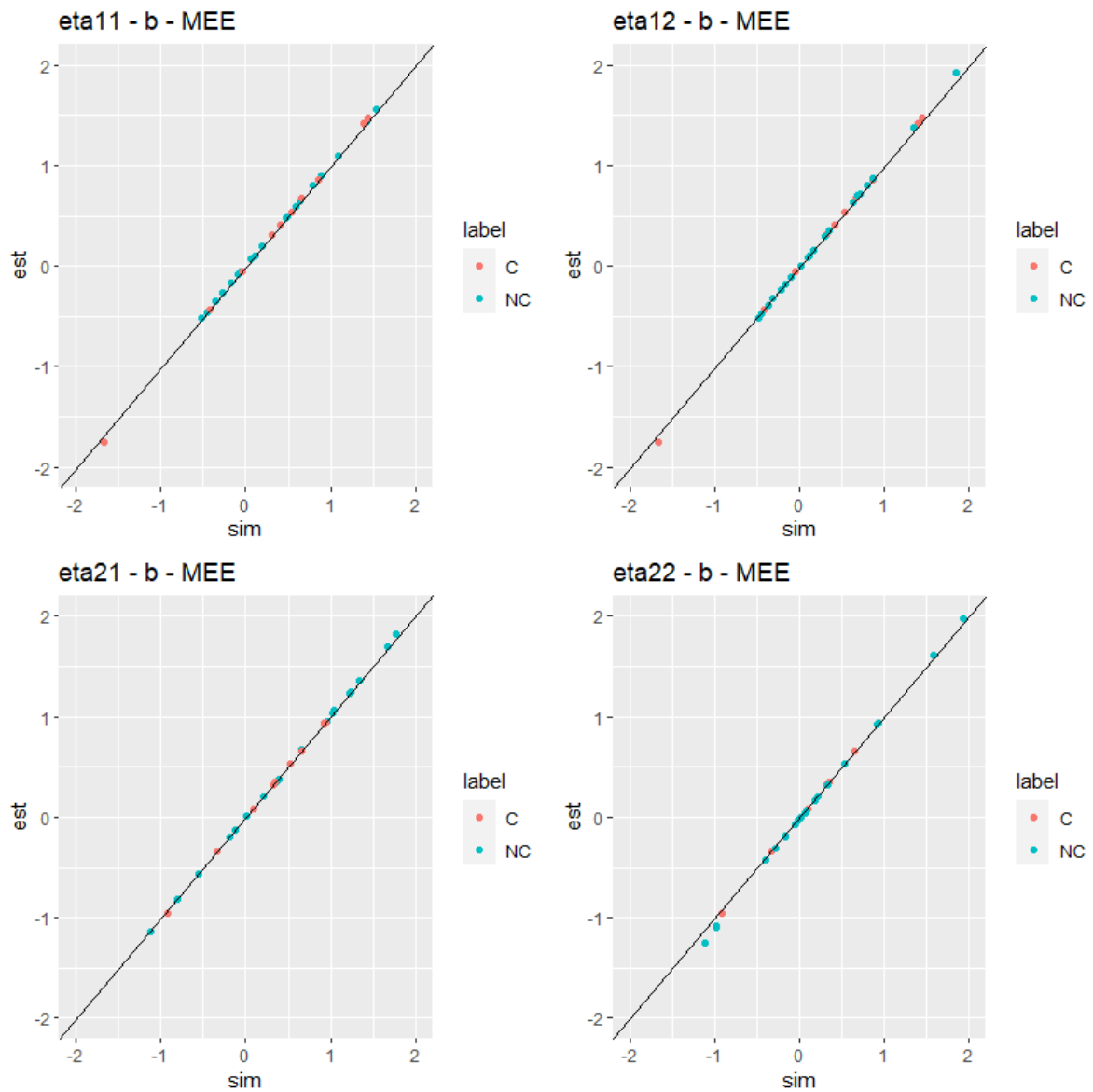


Figura 5.7 Recuperação do parâmetro de dificuldade b_{it} utilizando a MEE – Modelo multidimensional longitudinal de dois momentos. C – Item comum. NC – Item não-comum.

Tabela 5.5 Recuperação dos parâmetros de distribuição das variáveis latentes– Modelo multidimensional longitudinal de dois momentos. Valores brutos.

Parâmetros Populacionais	Valores Simulados	Valores Estimados (Erro padrão)
Média μ	$\begin{pmatrix} 0.00 \\ 0.50 \\ 0.00 \\ 0.90 \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ (0.000) \\ 0.711 \\ (0.016) \\ 0.000 \\ (0.000) \\ 1.281 \\ (0.016) \end{pmatrix}$
Matriz de Covariância	$\begin{pmatrix} 1 & & & \\ 0.70 & 1 & & \\ 0.35 & 0.25 & 1 & \\ 0.25 & 0.35 & 0.70 & 1 \end{pmatrix}$	$\begin{pmatrix} 1.998 & & & \\ (0.356) & & & \\ 1.426 & 2.076 & & \\ (0.027) & (0.043) & & \\ 0.685 & 0.502 & 1.979 & \\ (0.023) & (0.024) & (0.033) & \\ 0.508 & 0.724 & 1.411 & 2.071 \\ (0.027) & (0.035) & (0.029) & (0.048) \end{pmatrix}$
Matriz de Correlação	$\begin{pmatrix} 1 & & & \\ 0.70 & 1 & & \\ 0.35 & 0.25 & 1 & \\ 0.25 & 0.35 & 0.70 & 1 \end{pmatrix}$	$\begin{pmatrix} 1.000 & & & \\ (0.000) & & & \\ 0.700 & 1.000 & & \\ (0.005) & (0.000) & & \\ 0.349 & 0.248 & 1.000 & \\ (0.010) & (0.011) & (0.000) & \\ 0.249 & 0.349 & 0.697 & 1.000 \\ (0.012) & (0.013) & (0.006) & (0.000) \end{pmatrix}$

Tabela 5.6 Recuperação dos parâmetros de distribuição das variáveis latentes– Modelo multidimensional longitudinal de dois momentos. Valores padronizados.

Parâmetros Populacionais	Valores Simulados	Valores estimados padronizados (erro padrão)
Média μ	$\begin{pmatrix} 0.00 \\ 0.50 \\ 0.00 \\ 0.90 \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ (0.000) \\ 0.494 \\ (0.011) \\ 0.000 \\ (0.000) \\ 0.890 \\ (0.013) \end{pmatrix}$
Matriz de Covariância	$\begin{pmatrix} 1 & & & \\ 0.70 & 1 & & \\ 0.35 & 0.25 & 1 & \\ 0.25 & 0.35 & 0.70 & 1 \end{pmatrix}$	$\begin{pmatrix} 1.000 & & & \\ (0.000) & & & \\ 0.700 & 1.000 & & \\ (0.005) & (0.000) & & \\ 0.349 & 0.248 & 1.000 & \\ (0.010) & (0.011) & (0.000) & \\ 0.249 & 0.349 & 0.697 & 1.000 \\ (0.012) & (0.013) & (0.006) & (0.000) \end{pmatrix}$
Matriz de Correlação	$\begin{pmatrix} 1 & & & \\ 0.70 & 1 & & \\ 0.35 & 0.25 & 1 & \\ 0.25 & 0.35 & 0.70 & 1 \end{pmatrix}$	$\begin{pmatrix} 1.000 & & & \\ (0.000) & & & \\ 0.700 & 1.000 & & \\ (0.005) & (0.000) & & \\ 0.349 & 0.248 & 1.000 & \\ (0.010) & (0.011) & (0.000) & \\ 0.249 & 0.349 & 0.697 & 1.000 \\ (0.012) & (0.013) & (0.006) & (0.000) \end{pmatrix}$

Capítulo 6 Aplicação

The time is ripe for our measurement system to shift emphasis from measuring economic production to measuring people's well-being. And measures of well-being should be put in a context of sustainability.

Nic Marks

6.1 Origem dos dados

A metodologia proposta neste trabalho foi aplicada aos dados provenientes da aplicação de provas de Língua Portuguesa e Matemática aos alunos do quinto ano do Ensino Fundamental durante a execução do **Programa Especial de Aumento de Rendimento Escolar – Escola +** da Secretaria Municipal de Educação do município de Cariacica, Espírito Santo, ao longo do ano de 2019. Apêndice E contém o documento autorizando o uso desses dados neste trabalho. O programa tinha o intuito de aumentar a proficiência dos alunos nos descritores pertencentes às Matrizes de Referência de Língua Portuguesa [85] e de Matemática [86] do SAEB do quinto ano do ensino fundamental publicadas em 2001.

Os alunos foram submetidos a um programa de reforço focado nos descritores com menor desempenho detectados na prova diagnóstica inicial. Esse programa se deu na forma de aulas presenciais em contraturno ministradas pelos próprios professores dos alunos. Os professores foram instruídos na interpretação dos relatórios diagnósticos, na construção de itens e nos equívocos mais frequentes nos descritores abordados. O programa teve como foco todos os anos do ciclo final do Ensino Fundamental, mas esta análise focará nos alunos do 5º ano por ser o grupo com maior participação efetiva.

A Tabela 6.1 lista os oito testes aplicados ao longo do programa, mas somente quatro serão utilizados neste estudo longitudinal, os testes S01_LP e S01_MT, aplicados em março de 2019, e os testes S04_LP e S04_MT, aplicados em outubro de 2019, pois estes cobriram todos os descritores das matrizes de avaliação abordadas. Na próxima seção os dados do teste S01_LP são analisados sob um modelo transversal utilizando métodos da TRI e da MEE com o intuito de mostrar a equivalência e as diferenças entre as duas abordagens. Na Seção 6.3, os quatro testes são analisados sob a formulação de modelos longitudinais. Com a TRI, a análise se dará com a combinação de dois modelos unidimensionais longitudinais de dois momentos, um abordando os testes de Língua Portuguesa (S01_LP e S04_LP) e outro, os testes de Matemática (S01_MT e S04_MT). Com a MEE, a análise se dá com um modelo multidimensional longitudinal de dois momentos. Nos dois modelos, a equalização dos dados entre os dois momentos se dá pela presença de itens comuns entre os testes. De todos os alunos que responderam aos testes – ver Tabela 6.1, esta análise se limitará aos 2203 alunos que responderam os quatro testes mencionados. Isso se dá devido uma limitação do pacote `lavaan` que exclui da análise qualquer linha de dados que contenha dados faltantes sistemáticos.

Tabela 6.1 Testes aplicados no Programa Especial de Aumento de Rendimento Escola – Escola +

Teste	Matriz de Competência Abordada	Data de aplicação	Número de alunos	Utilização nesta tese?
S01_LP	SAEB 2001 Língua Portuguesa	Mar/2019	3225	SIM
S01_MT	SAEB 2001 Matemática	Mar/2019	3208	SIM
S02_LP	SAEB 2001 Língua Portuguesa Parcial	Jun/2019	4281*	NÃO
S02_MT	SAEB 2001 Matemática Parcial	Jun/2019	4280*	NÃO
S03_LP	SAEB 2001 Língua Portuguesa Parcial	Ago/2019	5236*	NÃO
S04_MT	SAEB 2001 Matemática Parcial	Ago/2019	5236*	NÃO
S04_LP	SAEB 2001 Língua Portuguesa	Out/2019	3148	SIM
S04_MT	SAEB 2001 Matemática	Out/2019	3164	SIM

* Inclui alunos de todos os anos do ensino fundamental.

6.2 Aplicação em modelos transversais

A Tabela 6.2 contém um breve sumário estatístico do teste S01_LP. Esse teste englobou 30 questões de Língua Portuguesa e teve um escore bruto médio de acertos de 14.046 questões com desvio padrão de 5.428 acertos. A Figura 6.1 mostra o histograma dos escores brutos. O valor de alpha de Cronbach para o teste é de 0.793 indicando que a consistência interna do instrumento é satisfatória, ou seja, que todo item do teste compartilha variância comum com, pelo menos, alguns outros itens do mesmo [87].

6.2.1 Análise utilizando a TRI e a TCT

A Tabela 6.2 contém as estatísticas gerais do teste, a primeira coluna calcula essas estatísticas usando todos os itens do teste, e a segunda utilizando somente itens cuja qualidade foi suficiente para incluí-los no teste final. E a Tabela 6.3 contém estatísticas que auxiliam o pesquisador na escolha de quais itens incluir no teste final. O objetivo é ter no teste somente itens que realmente forneçam informações sobre o traço latente de interesse. A porcentagem de acertos de um item, p , se muito baixa, pode indicar que o item é muito difícil para o público-alvo, especialmente se p ficar abaixo da probabilidade de acerto ao acaso desse item, que, em questões de múltipla escolha com n alternativas, é aproximadamente igual a $\frac{1}{n}$. A correlação ponto-biserial, $pbis$ abaixo de 0,20, indica itens com problemas de elaboração ou itens em que os alunos não estão capacitados a resolver. Esses itens afetam negativamente a precisão e a convergência dos processos de estimação dos parâmetros e, por isso, frequentemente são removidos da modelagem. Outro artifício usado na detecção de itens problemáticos é avaliar o valor de alpha de Cronbach entre os testes que incluem e excluem o item. Se o valor de alpha para o teste que excluir o item for maior, isso pode ser evidência de que o item tem problemas de mensuração. Ver colunas “Alpha se ausente” e “Alpha < 0.793” na Tabela 6.3. Por último, o pesquisador pode tentativamente ajustar o modelo com todos os itens e identificar os itens problemáticos analisando as estimativas dos parâmetros de discriminação (a_{it}) e de dificuldade (b_{it}) dos itens. Nesta aplicação, os dados dicotômicos foram ajustados para o modelo de dois parâmetros da TRI, cuja estimação foi feita utilizando o algoritmo de Esperança-Maximização (EM) utilizando o pacote MIRT desenvolvido na linguagem R por Chalmers [88]. Itens com parâmetros de discriminação próximos de zero ou muito baixos¹³, são itens cuja probabilidade de acerto varia pouco para grandes variações nos traços latentes dos indivíduos e não são desejáveis como parte do teste. Também, são suspeitos de não serem desejáveis, itens com parâmetros de dificuldade cujo valor absoluto excede 4.0 desvios-padrão da escala do traço latente. Isso ocorre pois nesses intervalos da escala, a amostra do teste pode não ser grande o suficiente para garantir uma estimação com erro suficientemente pequeno. Em geral, não existe um critério objetivo para decidir sobre a inclusão ou exclusão de um item ao teste, e, além das estatísticas aqui apresentadas, também se analisa o ajuste das curvas características dos itens aos dados e a validade dos itens quanto à sua capacidade de medir o que ele se propõe a medir. Avaliar a validade de itens está fora do escopo deste trabalho, e o autor sugere consultar o trabalho de Haladyna sobre este vasto campo de pesquisa [89]–[91]. Neste experimento, por uma

¹³ Ver Tabela 3.1 Valores de referência para o parâmetro de discriminação a_i segundo Baker

combinação dos fatores acima listados, sete itens não foram utilizados no instrumento final e, após a exclusão desses itens, o valor de alpha de Cronbach aumentou de 0.793 para 0.818. Esses sete itens estão indicados na coluna “Usar?” na Tabela 6.3. O Item S01_LP_Q24 é um exemplo que ilustra como é subjetiva a decisão da seleção final de itens de um teste, pois pelo critério da *pbis* ele está marginalmente dentro e pelo do alpha de Cronbach, marginalmente fora. Optou-se por incluí-lo, dado que sua porcentagem de acerto é relativamente alta ($p=37.5\%$) e que sua localização na escala não é extrema (1.059).

A Figura 6.1 mostra o histograma dos escores brutos e Figura 6.2 mostra a curva de informação do teste após a seleção dos itens mais apropriados ao instrumento. A curva tem seu máximo próximo à média da escala e é praticamente simétrica em relação a esta. Essa característica é desejável pois indica que o instrumento avalia de maneira similar indivíduos localizados em ambos os lados da escala. Já a Figura 6.3 mostra as curvas características dos itens e nos permite identificar a não existência de itens com discriminação muito baixa ou muito alta. A Tabela 6.4 mostra os valores dos parâmetros de discriminação e dificuldade dos itens selecionados junto com as estatísticas utilizadas em sua seleção.

A Figura 6.4 mostra a comparação de escores latentes obtidos através de três métodos diferentes: máxima verossimilhança (Maximum Likelihood - ML), sob a notação *tri_ml*, pelo máximo da posteriori (maximum a posteriori - MAP), também chamado de Bayes Empírico Modal (Empirical Bayes Modal - EBM), sob a notação *tri_map_ebm*, e pela média da posteriori (Expected a Posteriori - EAP) também chamado de Bayes Empírico (Empirical Bayes – EB), sob a notação *tri_eap_eb* [84, p. 221]. A correlação entre os métodos *tri_map_ebm* e *tri_eap_eb* é linear ao longo de toda a escala, enquanto a relação destes com ML deixa de ser linear nos extremos.

A Figura 6.5 compara a distribuição dos escores latentes preditos nos diferentes métodos. Segundo DeMars [92, p. 65] a variância da população é superestimada pela variância das estimativas de máxima verossimilhança e subestimada pela variância das estimativas bayesianas, e sugere que estimar a variância diretamente é mais preciso do que a estimar indiretamente através do cálculo da variância de escores latentes individuais. Em concordância, na Figura 6.5, a variação dos escores preditos por ML é maior que os calculados pelos métodos Bayesianos.

Tabela 6.2 Estatísticas do teste, S01_LP - TRI.

Estatísticas	Valores	Valores*
Alpha de Cronbach	0.793	0.818
No de Respondentes	3225	3225
No de Itens	30	23
No Mínimo de Acertos	1	0
No Máximo de Acertos	30	23
Média de Acertos	14.046	11.665
Desvio Padrão dos acertos	5.428	4.986
% média de acertos	46.82%	50.72%
* – Valores calculados após a exclusão dos itens com pbis < 0.20.		

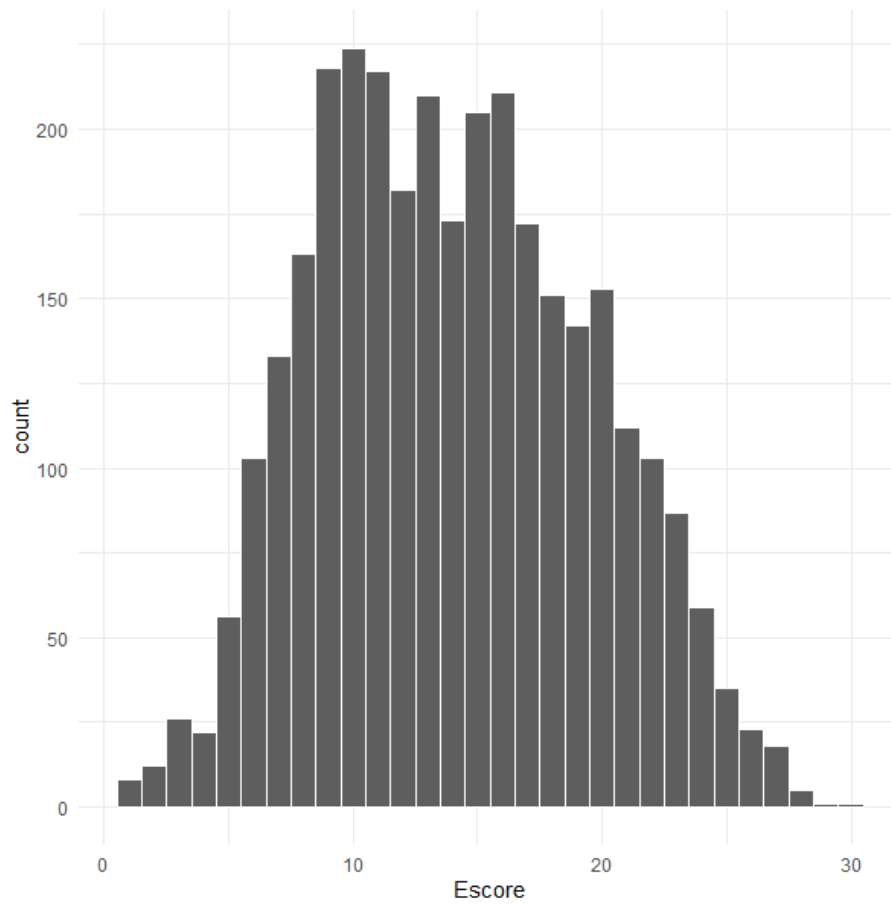


Figura 6.1 Histograma dos escores brutos. S01_LP - TRI

Tabela 6.3 Estatísticas de todos os itens. S01_LP - TRI

Items	p	pbis	a (ep)	b (ep)	Alpha se Ausente	Alpha > 0,793	pbis < 0,2	Usar?
S01_LP_Q1	0.509	0.412	1.166 (0.059)	-0.059 (0.038)	0.783	NÃO	NÃO	SIM
S01_LP_Q2	0.586	0.306	0.795 (0.050)	-0.508 (0.056)	0.788	NÃO	NÃO	SIM
S01_LP_Q3	0.542	0.102	0.221 (0.040)	-0.767 (0.210)	0.797	SIM	SIM	NÃO
S01_LP_Q4	0.729	0.353	1.146 (0.067)	-1.091 (0.059)	0.786	NÃO	NÃO	SIM
CMM_LP_1	0.476	0.277	0.691 (0.046)	0.143 (0.057)	0.789	NÃO	NÃO	SIM
S01_LP_Q6	0.320	0.467	1.540 (0.073)	0.678 (0.039)	0.781	NÃO	NÃO	SIM
S01_LP_Q7	0.223	0.016	0.042 (0.046)	29.446 (32.219)	0.798	SIM	SIM	NÃO
CMM_LP_2	0.720	0.393	1.347 (0.074)	-0.947 (0.048)	0.784	NÃO	NÃO	SIM
S01_LP_Q9	0.461	0.478	1.573 (0.073)	0.116 (0.033)	0.780	NÃO	NÃO	SIM
S01_LP_Q10	0.571	0.451	1.442 (0.070)	-0.299 (0.035)	0.781	NÃO	NÃO	SIM
S01_LP_Q11	0.571	0.431	1.266 (0.064)	-0.318 (0.038)	0.782	NÃO	NÃO	SIM
S01_LP_Q12	0.291	0.309	0.893 (0.053)	1.152 (0.073)	0.787	NÃO	NÃO	SIM
S01_LP_Q13	0.605	0.394	1.125 (0.060)	-0.493 (0.043)	0.784	NÃO	NÃO	SIM
S01_LP_Q14	0.482	0.265	0.634 (0.045)	0.119 (0.061)	0.789	NÃO	NÃO	SIM
S01_LP_Q15	0.694	0.216	0.581 (0.049)	-1.519 (0.131)	0.791	NÃO	NÃO	SIM
CMM_LP_3	0.590	0.506	1.841 (0.088)	-0.343 (0.031)	0.778	NÃO	NÃO	SIM
S01_LP_Q17	0.390	0.126	0.290 (0.041)	1.574 (0.248)	0.795	SIM	SIM	NÃO
S01_LP_Q18	0.529	0.366	0.979 (0.054)	-0.158 (0.043)	0.785	NÃO	NÃO	SIM
S01_LP_Q19	0.413	0.059	0.115 (0.039)	3.059 (1.085)	0.798	SIM	SIM	NÃO
S01_LP_Q20	0.163	0.057	0.181 (0.052)	9.093 (2.587)	0.796	SIM	SIM	NÃO

Items	p	pbis	a (ep)	b (ep)	Alpha se Ausente	Alpha > 0,793	pbis < 0,2	Usar?
S01_LP_Q21	0.377	0.332	0.847 (0.050)	0.674 (0.059)	0.786	NÃO	NÃO	SIM
S01_LP_Q22	0.429	0.348	0.924 (0.052)	0.352 (0.048)	0.786	NÃO	NÃO	SIM
CMM_LP_4	0.392	0.262	0.629 (0.045)	0.755 (0.078)	0.789	NÃO	NÃO	SIM
S01_LP_Q24	0.375	0.212	0.511 (0.044)	1.059 (0.111)	0.792	NÃO	NÃO	SIM
S01_LP_Q25	0.272	0.154	0.392 (0.045)	2.599 (0.297)	0.794	SIM	SIM	NÃO
S01_LP_Q26	0.314	0.275	0.714 (0.048)	1.212 (0.090)	0.789	NÃO	NÃO	SIM
S01_LP_Q27	0.640	0.428	1.296 (0.067)	-0.605 (0.041)	0.782	NÃO	NÃO	SIM
S01_LP_Q28	0.474	0.409	1.132 (0.058)	0.095 (0.040)	0.783	NÃO	NÃO	SIM
S01_LP_Q29	0.378	0.163	0.397 (0.042)	1.299 (0.160)	0.794	SIM	SIM	NÃO
CMM_LP_5	0.532	0.364	0.969 (0.054)	-0.176 (0.044)	0.785	NÃO	NÃO	SIM

p – porcentagem de acerto do item
a – parâmetro de discriminação do item
b – parâmetro de dificuldade do item
p_bis – correlação ponto biserial
ep – erro padrão

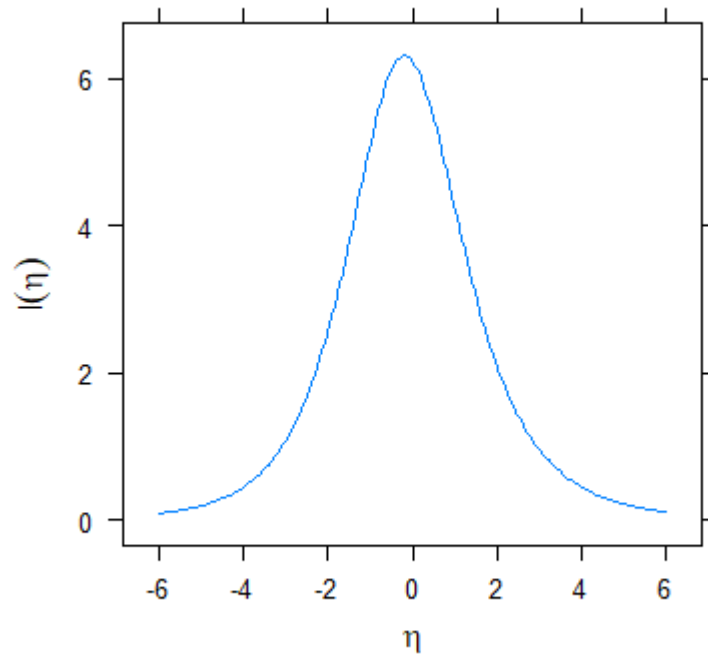


Figura 6.2 Curva de informação do teste. S01_LP - TRI

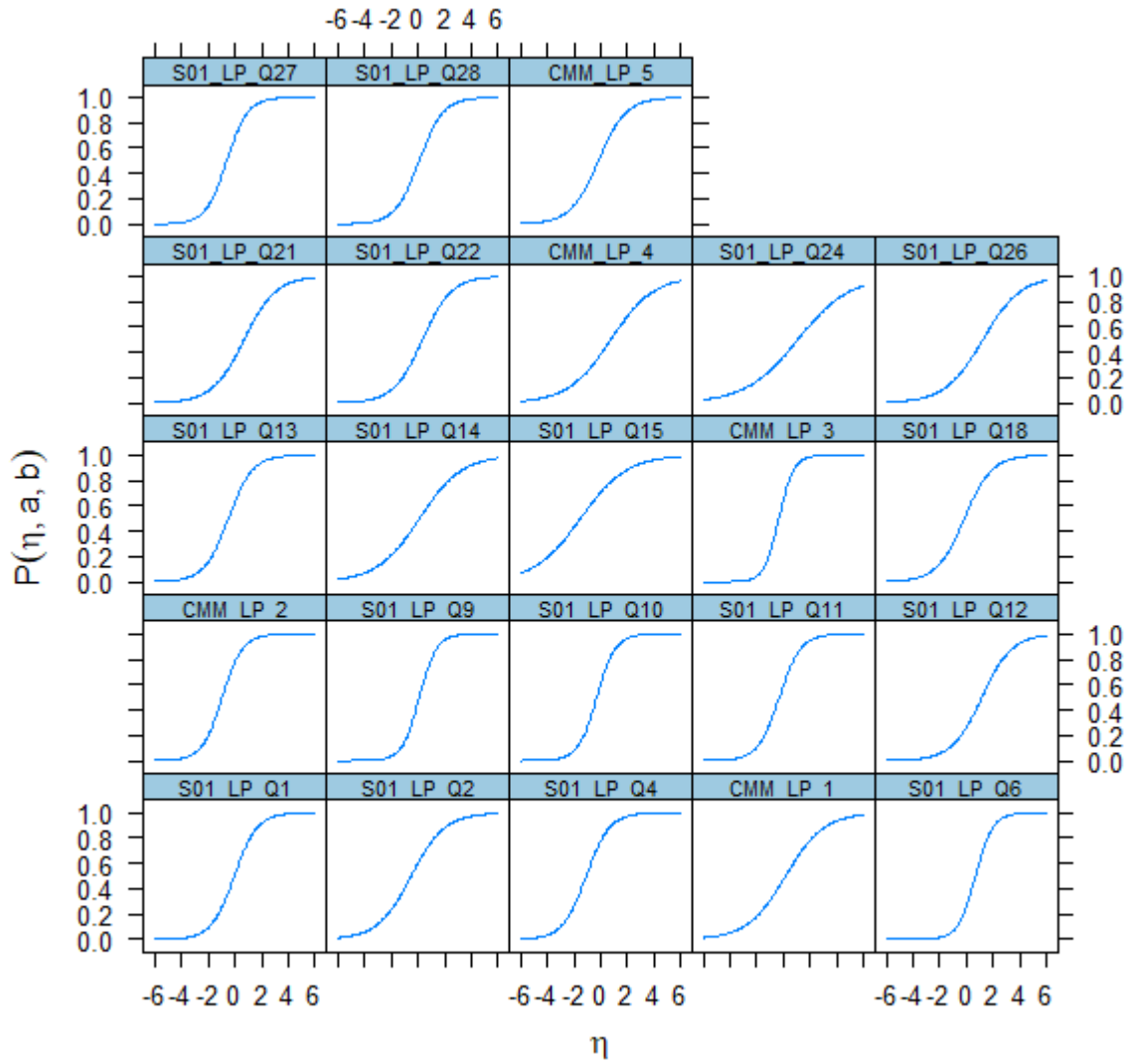


Figura 6.3 Curvas característica dos itens finais, S01_LP - TRI

Tabela 6.4 Estatísticas dos itens finais, S01_LP - TRI

Itens	p	pbis	a (ep)	b (ep)	Alpha se Ausente	Alpha > 0,818	pbis < 0,2	Usar?
S01_LP_Q1	0.509	0.421	1.168 (0.059)	-0.058 (0.038)	0.808	NÃO	NÃO	SIM
S01_LP_Q2	0.586	0.308	0.794 (0.050)	-0.509 (0.056)	0.814	NÃO	NÃO	SIM
S01_LP_Q4	0.729	0.356	1.145 (0.067)	-1.092 (0.059)	0.811	NÃO	NÃO	SIM
CMM_LP_1	0.476	0.283	0.693 (0.046)	0.143 (0.057)	0.815	NÃO	NÃO	SIM
S01_LP_Q6	0.320	0.475	1.546 (0.073)	0.679 (0.039)	0.806	NÃO	NÃO	SIM
CMM_LP_2	0.720	0.396	1.332 (0.073)	-0.954 (0.049)	0.810	NÃO	NÃO	SIM
S01_LP_Q9	0.461	0.492	1.578 (0.073)	0.117 (0.033)	0.805	NÃO	NÃO	SIM
S01_LP_Q10	0.571	0.469	1.450 (0.070)	-0.298 (0.034)	0.806	NÃO	NÃO	SIM
S01_LP_Q11	0.571	0.437	1.263 (0.063)	-0.318 (0.038)	0.808	NÃO	NÃO	SIM
S01_LP_Q12	0.291	0.320	0.896 (0.053)	1.149 (0.072)	0.813	NÃO	NÃO	SIM
S01_LP_Q13	0.605	0.397	1.116 (0.060)	-0.495 (0.043)	0.809	NÃO	NÃO	SIM
S01_LP_Q14	0.482	0.269	0.631 (0.045)	0.120 (0.062)	0.816	NÃO	NÃO	SIM
S01_LP_Q15	0.694	0.219	0.581 (0.049)	-1.519 (0.131)	0.817	NÃO	NÃO	SIM
CMM_LP_3	0.590	0.521	1.831 (0.088)	-0.343 (0.031)	0.804	NÃO	NÃO	SIM
S01_LP_Q18	0.529	0.374	0.979 (0.054)	-0.158 (0.044)	0.811	NÃO	NÃO	SIM
S01_LP_Q21	0.377	0.329	0.840 (0.050)	0.679 (0.059)	0.813	NÃO	NÃO	SIM
S01_LP_Q22	0.429	0.358	0.925 (0.052)	0.352 (0.048)	0.811	NÃO	NÃO	SIM
CMM_LP_4	0.392	0.259	0.625 (0.045)	0.758 (0.079)	0.816	NÃO	NÃO	SIM
S01_LP_Q24	0.375	0.216	0.512 (0.044)	1.058 (0.111)	0.818	SIM	NÃO	SIM
S01_LP_Q26	0.314	0.275	0.709 (0.048)	1.219 (0.091)	0.815	NÃO	NÃO	SIM
S01_LP_Q27	0.640	0.427	1.279 (0.067)	-0.609 (0.041)	0.808	NÃO	NÃO	SIM
S01_LP_Q28	0.474	0.410	1.121 (0.057)	0.097 (0.040)	0.809	NÃO	NÃO	SIM
CMM_LP_5	0.532	0.367	0.961 (0.053)	-0.176 (0.044)	0.811	NÃO	NÃO	SIM

p – porcentagem de acerto do item
 a – parâmetro de discriminação do item
 b – parâmetro de dificuldade do item
 p_bis – correlação ponto biserial
 ep – erro padrão

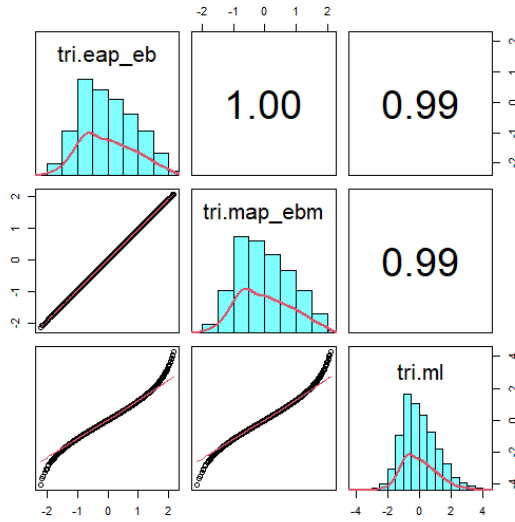


Figura 6.4 Comparação de métodos de estimação dos escores latentes, S01_LP - TRI

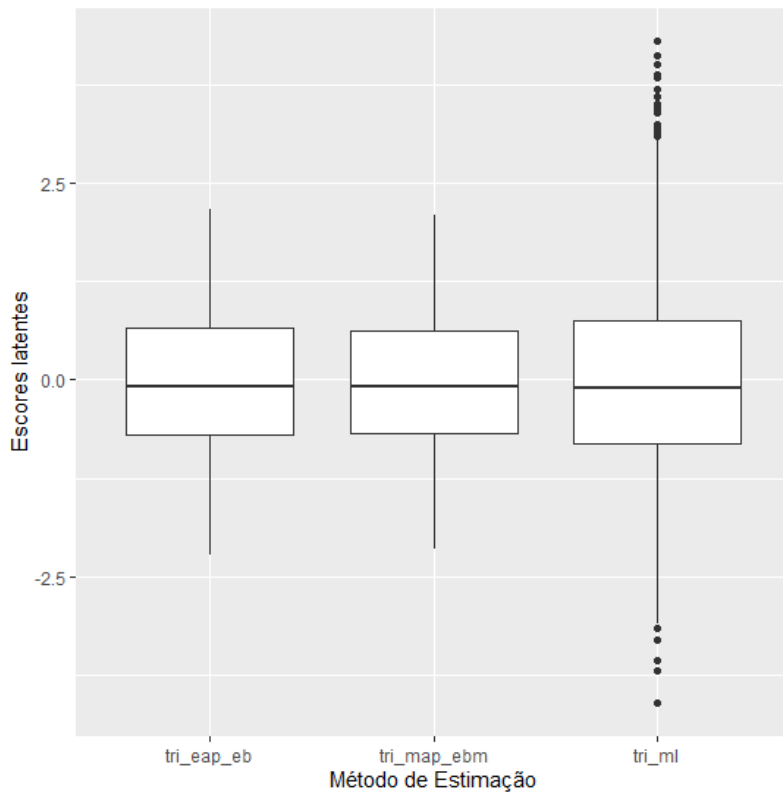


Figura 6.5 Distribuição dos escores latentes nos métodos de estimação. S01_LP - TRI

6.2.2 Análise utilizando a MEE

Nesta seção os mesmos dados da seção anterior são analisados utilizando os métodos da MEE. Os dados são processados utilizando o pacote `lavaan` desenvolvido na linguagem R por Rosseel [93]. Similarmente ao que foi feito na análise da seção anterior, é necessário identificar e eliminar da modelagem itens com reduzida contribuição na mensuração do traço latente.

Conforme pode ser visto na Equação (4.1), a carga fatorial é um o coeficiente do fator comum na equação de regressão que relaciona o fator comum à variável observada e quanto maior a carga fatorial maior a relação entre os valores das variáveis observadas e o fator comum. Desta maneira, itens com baixa carga fatorial devem ser submetidos a uma análise mais criteriosa para se decidir sobre sua inclusão ou não no instrumento final. É importante notar que o valor da carga fatorial depende da parametrização do modelo e, por isso, seu valor deve ser avaliado relativamente à carga fatorial dos outros itens do modelo.

A Tabela 6.5 mostra as cargas amostrais para cada item obtidas sob três diferentes parametrizações: FP_CP_FT, RI_CP_FT e EC_CP_FT, de acordo com a nomenclatura definida na Seção D.6.4- *Eighteen parameterization options* na qual 18 diferentes parametrizações foram apresentadas. Os sete menores valores foram hachurados para facilitar a identificação. Esses são os mesmos sete itens identificados na seção anterior na análise sob a TRI.

O coeficiente de determinação, R^2 , é também é uma estatística bastante útil para discernir que itens têm uma relação mais forte com o fator comum, pois representa a proporção da variância da variável observada que é explicada no modelo de regressão do fator comum - Equação (4.1). Seu valor independe da parametrização utilizada, como pode ser constatado na Tabela 6.6 que mostra o R^2 para cada item. As células hachuradas de cada coluna indicam os sete menores valores de R^2 dentre os itens. Aqui, também esses são os mesmos itens que foram excluídos da análise da seção anterior. Ainda, também o item S01_LP_Q24 é limiar e fica a critério do pesquisador a decisão de incluí-lo ou não. Optou-se por incluí-lo tal qual foi feito na seção anterior.

Tabela 6.5 - Carga fatorial dos itens em três diferentes parametrizações. S01_LP – MEE.

Itens	ec_cp_ft	fp_cp_ft	ri_cp_ft
S01_LP_Q7	0.032	0.016	0.023
S01_LP_Q19	0.143	0.073	0.104
S01_LP_Q20	0.178	0.091	0.129
S01_LP_Q3	0.275	0.141	0.200
S01_LP_Q17	0.352	0.180	0.255
S01_LP_Q25	0.454	0.232	0.329
S01_LP_Q29	0.465	0.238	0.337
S01_LP_Q24	0.625	0.320	0.453
S01_LP_Q15	0.658	0.336	0.477
CMM_LP_4	0.758	0.387	0.549
S01_LP_Q14	0.780	0.399	0.566
CMM_LP_1	0.837	0.428	0.607
S01_LP_Q26	0.846	0.433	0.613
S01_LP_Q2	0.934	0.477	0.677
S01_LP_Q21	1.030	0.527	0.747
S01_LP_Q12	1.057	0.540	0.766
S01_LP_Q22	1.118	0.572	0.811
CMM_LP_5	1.147	0.587	0.832
S01_LP_Q18	1.168	0.597	0.847
S01_LP_Q4	1.217	0.622	0.882
S01_LP_Q13	1.288	0.658	0.934
S01_LP_Q28	1.354	0.692	0.982
S01_LP_Q1	1.379	0.705	1.000
CMM_LP_2	1.413	0.722	1.024
S01_LP_Q27	1.449	0.741	1.050
S01_LP_Q11	1.459	0.746	1.058
S01_LP_Q10	1.650	0.844	1.196
S01_LP_Q9	1.929	0.986	1.398
S01_LP_Q6	1.969	1.007	1.428
CMM_LP_3	2.032	1.039	1.473

Tabela 6.6 – R^2 dos itens em três diferentes parametrizações. S01_LP – MEE.

Itens	ec_cp_ft	fp_cp_ft	ri_cp_ft
S01_LP_Q7	0.000	0.000	0.000
S01_LP_Q19	0.005	0.005	0.005
S01_LP_Q20	0.008	0.008	0.008
S01_LP_Q3	0.019	0.019	0.019
S01_LP_Q17	0.031	0.031	0.031
S01_LP_Q25	0.051	0.051	0.051
S01_LP_Q29	0.054	0.054	0.054
S01_LP_Q24	0.093	0.093	0.093
S01_LP_Q15	0.102	0.102	0.102
S01_LP_Q4	0.130	0.130	0.130
S01_LP_Q14	0.137	0.137	0.137
S01_LP_Q5	0.155	0.155	0.155
S01_LP_Q26	0.158	0.158	0.158
CMM_LP_1	0.186	0.186	0.186
S01_LP_Q21	0.217	0.217	0.217
S01_LP_Q12	0.226	0.226	0.226
S01_LP_Q22	0.246	0.246	0.246
S01_LP_Q5	0.256	0.256	0.256
S01_LP_Q18	0.263	0.263	0.263
S01_LP_Q4	0.279	0.279	0.279
S01_LP_Q13	0.302	0.302	0.302
S01_LP_Q28	0.324	0.324	0.324
S01_LP_Q1	0.332	0.332	0.332
CMM_LP_2	0.343	0.343	0.343
S01_LP_Q27	0.354	0.354	0.354
S01_LP_Q11	0.358	0.358	0.358
S01_LP_Q10	0.416	0.416	0.416
S01_LP_Q9	0.493	0.493	0.493
S01_LP_Q6	0.503	0.503	0.503
S01_LP_Q3	0.519	0.519	0.519

Após a seleção dos itens, 23 itens foram utilizados na mensuração de uma variável latente com a parametrização RI_CP_FT (Reference Indicator, Conditional Parameterization, Free Threshold). As medidas de ajuste do modelo estão listadas na Tabela 6.7. O modelo está falhando no teste qui-quadrado e isso, provavelmente, se dá porque o valor da estatística χ^2 , conforme definição na Equação (4.26), depende do tamanho da amostra e o tamanho da amostra deste experimento é muito grande. Além disso, a análise da matriz de covariância dos resíduos, bem como dos índices de modificação, não indicou discrepâncias que, se corrigidas, poderiam alterar significativamente o valor de χ^2 . Também, foi verificada a sugestão de Kline de que “resíduos de correlação são mais fáceis de se interpretar do que resíduos de covariância, e resíduos de correlação absolutos maiores que 0,1 merecem atenção especial como uma possível evidência de ajuste local fraco” [39, p. 278], mas não foram encontrados resíduos que se encaixassem no critério proposto, o que sugere que o modelo não sofre de falta de ajustes locais.

A Tabela 6.8 mostra os parâmetros da distribuição da variável latente, enquanto a Tabela 6.9 mostra a carga fatorial e o valor limiar dos indicadores nas colunas λ_{it} e τ_{it} , respectivamente. A partir desses valores é possível calcular os parâmetros de discriminação a_{it} e dificuldade b_{it} dos indicadores através da aplicação das Equações (D.14) e (D.15) com $\nu_{it} = 0$ e $\theta_{it} = 1$ dado que a parametrização RI_CP_FT foi escolhida.

A Figura 6.6 mostra a comparação de escores latentes obtidos através de dois métodos diferentes: máxima verossimilhança (Maximum Likelihood - ML) sob a notação *mee_ml*, pelo máximo da posteriori (maximum a posteriori - MAP), também chamado de Bayes Empírico Modal (Empirical Bayes Modal - EBM) sob a notação *mee_map_ebm*. A correlação dos escores obtidos pelos métodos é linear na parte central da escala e gradualmente se torna não linear nos extremos. A Figura 6.7 compara a distribuição dos escores latentes observada nos diferentes métodos. Tal como aconteceu no caso da TRI, a variação dos escores estimados por máxima verossimilhança é maior do que os calculados pelo método bayesiano.

Tabela 6.7 - Medidas de ajuste. S01_LP - MEE.

Descrição	Valor
Número de observações	3225
Número de parâmetros	46
χ^2	426.485
Graus de Liberdade	230
valor-p	6.36E-14
CFI	0.993
TLI	0.992
RMSEA	0.016

Tabela 6.8 Parâmetros da distribuição da variável latente. S01_LP - MEE.

Descrição	Valor (erro padrão)
Média (α)	0 (0.000)
Variância (ψ)	0.503 (0.050)

Tabela 6.9 Parâmetros dos itens. S01_LP - MEE.

Itens	λ_{it}	τ_{it}	a_{it}	b_{it}
S01_LP_Q1	1.000 (0.000)	-0.027 (0.027)	1.206	-0.038
S01_LP_Q2	0.673 (0.050)	-0.241 (0.025)	0.811	-0.505
S01_LP_Q4	0.879 (0.061)	-0.718 (0.029)	1.060	-1.153
CMM_LP_1	0.606 (0.049)	0.065 (0.024)	0.731	0.151
S01_LP_Q6	1.427 (0.103)	0.665 (0.038)	1.721	0.658
CMM_LP_2	1.013 (0.070)	-0.716 (0.030)	1.221	-0.998
S01_LP_Q9	1.401 (0.094)	0.139 (0.032)	1.690	0.140
S01_LP_Q10	1.203 (0.078)	-0.234 (0.029)	1.450	-0.274
S01_LP_Q11	1.054 (0.071)	-0.223 (0.028)	1.272	-0.298
S01_LP_Q12	0.765 (0.061)	0.625 (0.028)	0.922	1.152
S01_LP_Q13	0.925 (0.064)	-0.318 (0.027)	1.115	-0.486
S01_LP_Q14	0.561 (0.048)	0.050 (0.024)	0.676	0.125
S01_LP_Q15	0.476 (0.046)	-0.535 (0.025)	0.574	-1.586
CMM_LP_3	1.468 (0.095)	-0.328 (0.032)	1.770	-0.315
S01_LP_Q18	0.845 (0.061)	-0.085 (0.026)	1.018	-0.142
S01_LP_Q21	0.737 (0.056)	0.353 (0.026)	0.889	0.676
S01_LP_Q22	0.808 (0.058)	0.206 (0.026)	0.974	0.359
CMM_LP_4	0.543 (0.047)	0.294 (0.024)	0.654	0.764
S01_LP_Q24	0.451 (0.043)	0.336 (0.024)	0.544	1.050
S01_LP_Q26	0.606 (0.052)	0.528 (0.026)	0.731	1.229
S01_LP_Q27	1.036 (0.068)	-0.444 (0.028)	1.249	-0.604
S01_LP_Q28	0.969 (0.066)	0.078 (0.027)	1.169	0.113
CMM_LP_5	0.822 (0.058)	-0.094 (0.026)	0.991	-0.162

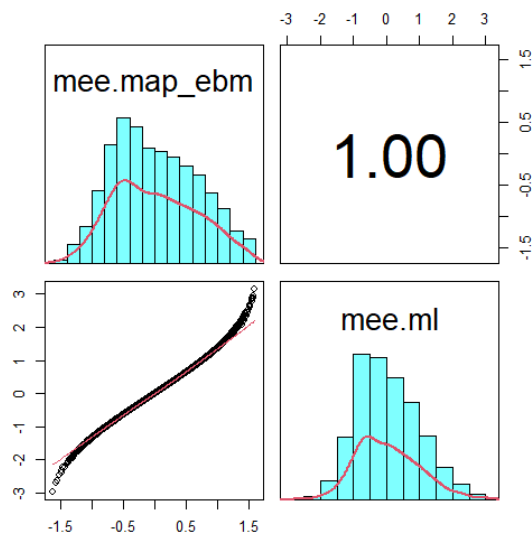


Figura 6.6 Comparação de métodos de estimação dos escores latentes. S01_LP - MEE. N=10 000.

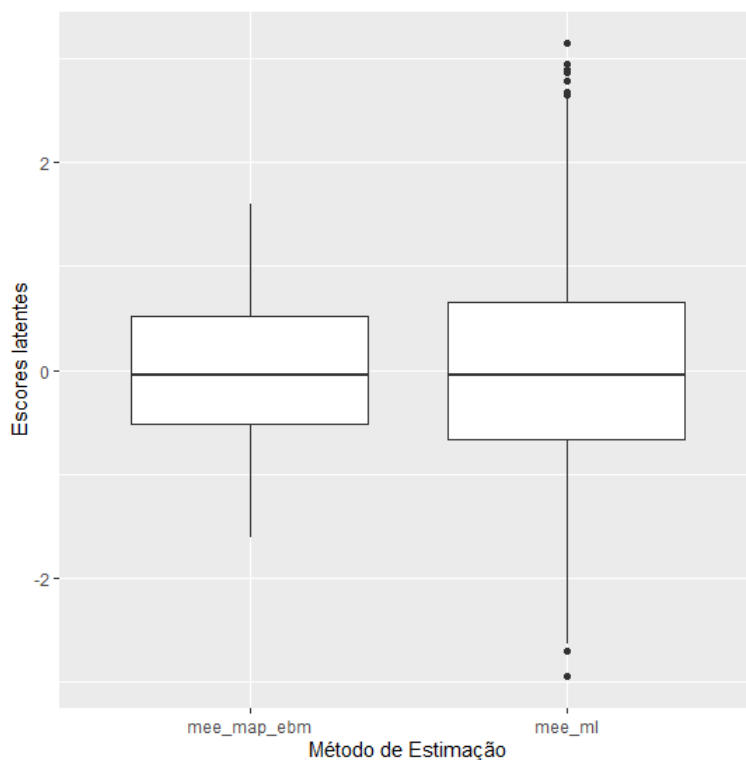


Figura 6.7 Distribuição dos escores latentes nos métodos de estimação. S01_LP - MEE.

6.2.3 Comparação entre os resultados obtidos pela TRI e pela MEE

A Figura 6.8 e Figura 6.9 comparam, respectivamente, os valores obtidos para o parâmetro de discriminação a_{it} e para o parâmetro de dificuldade b_{it} utilizando os métodos da TRI e MEE. Notadamente, em ambas as análises, a consistência das estimativas obtidas para a_{it} é menor do que as obtidas para b_{it} . Os três itens com a maior diferença entre os valores estimados para a_{it} estão identificados nas duas figuras, e essa diferença não parece estar ligada à localização do item na escala ou à quantidade de informação disponível para se utilizar no processo de estimação, pois esses três itens estão localizados na parte central da escala.

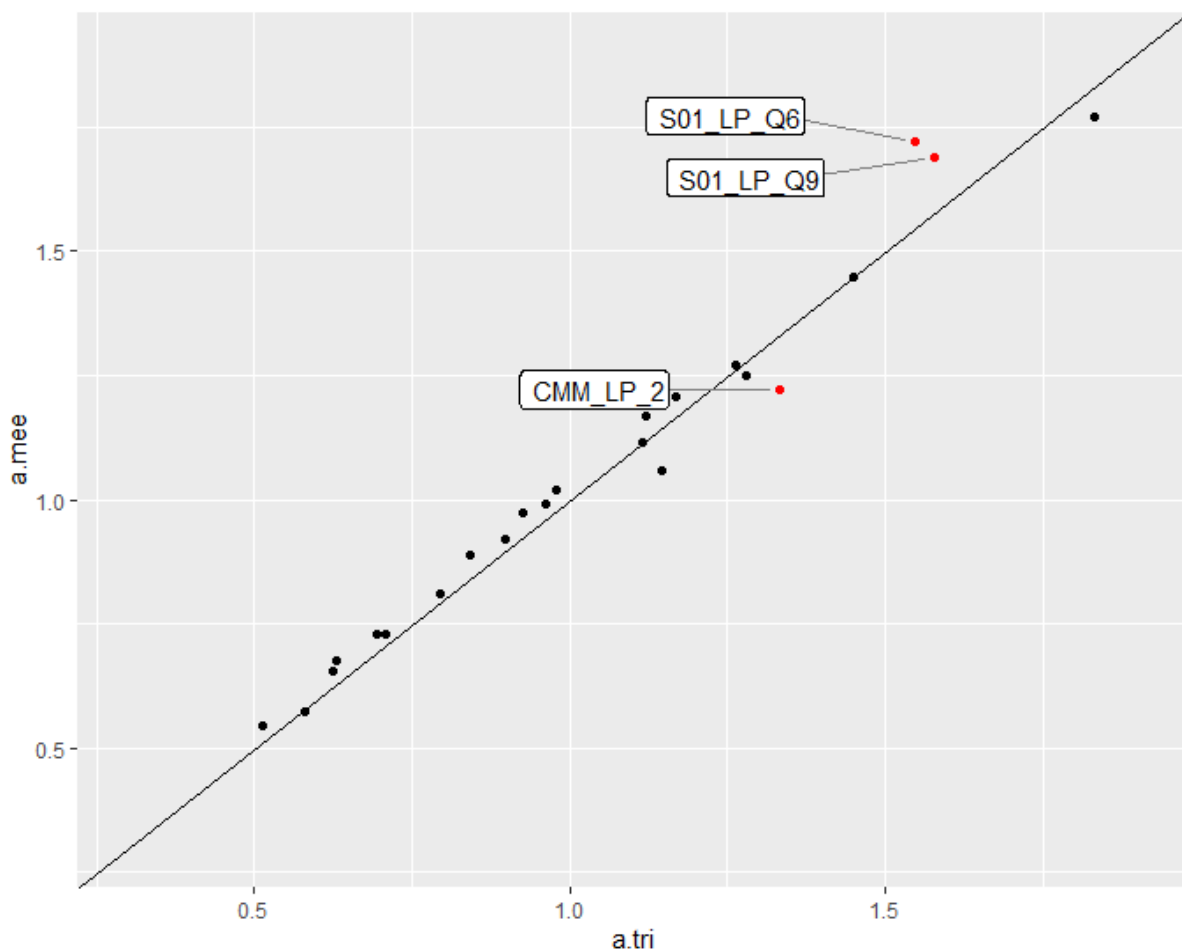


Figura 6.8 - Equivalência entre os valores de a_{it} obtidos com modelos TRI e MEE. S01_LP.

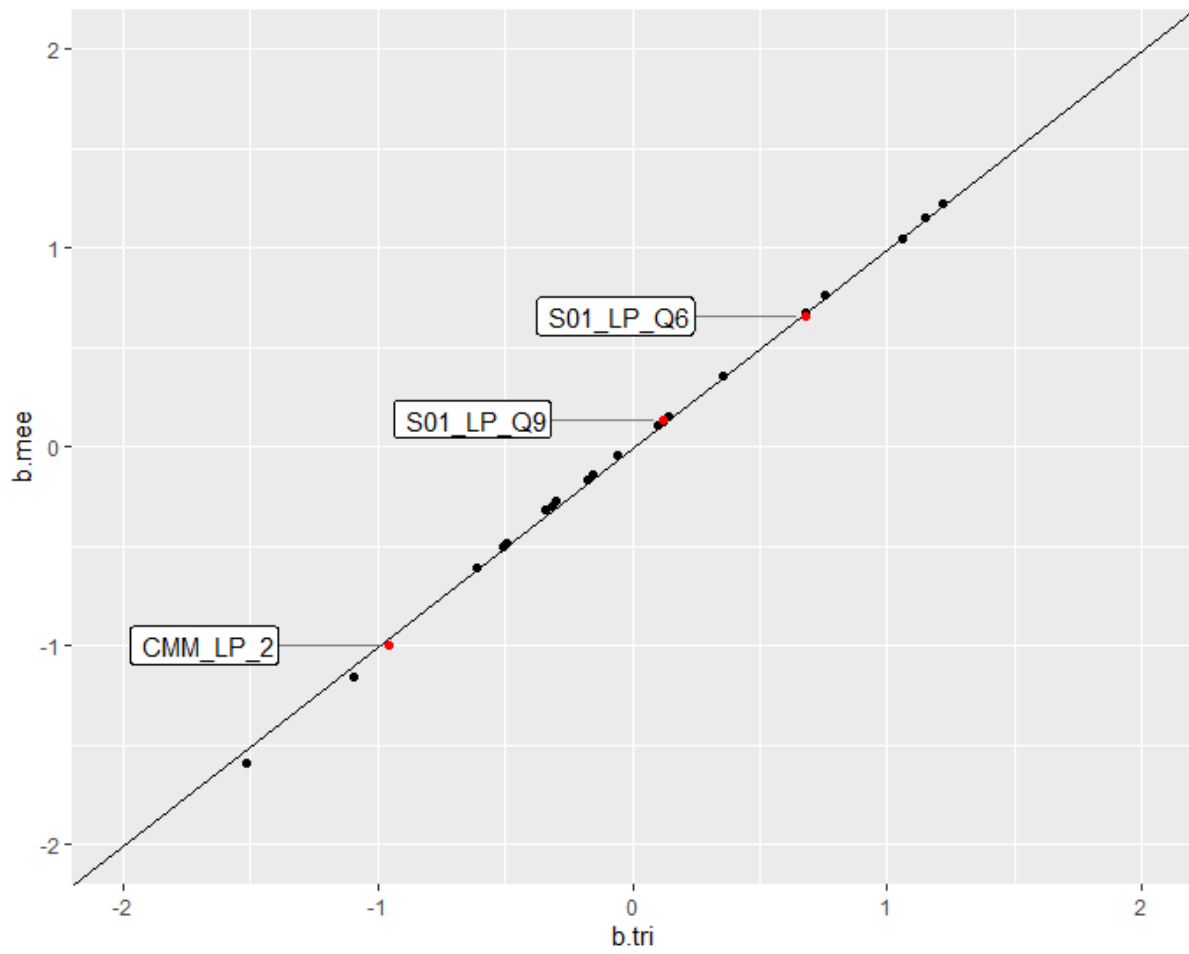


Figura 6.9 - Equivalência entre os valores de b_{it} obtidos com modelos TRI e MEE. S01_LP.

Tabela 6.10 - Valores dos parâmetros dos itens com modelagem TRI e MEE. S01_LP.

Itens	a.tri	a.mee	a.ratio*	b.tri	b.mee	b.delta**
S01_LP_Q1	1.168	1.206	0.969	-0.058	-0.038	-0.020
S01_LP_Q2	0.794	0.811	0.978	-0.509	-0.505	-0.004
S01_LP_Q4	1.145	1.060	1.081	-1.092	-1.153	0.061
CMM_LP_1	0.693	0.731	0.948	0.143	0.151	-0.007
S01_LP_Q6	1.546	1.721	0.898	0.679	0.658	0.021
CMM_LP_2	1.332	1.221	1.090	-0.954	-0.998	0.044
S01_LP_Q9	1.578	1.690	0.934	0.117	0.140	-0.023
S01_LP_Q10	1.450	1.450	1.000	-0.298	-0.274	-0.024
S01_LP_Q11	1.263	1.272	0.993	-0.318	-0.298	-0.019
S01_LP_Q12	0.896	0.922	0.972	1.149	1.152	-0.003
S01_LP_Q13	1.116	1.115	1.001	-0.495	-0.486	-0.010
S01_LP_Q14	0.631	0.676	0.932	0.120	0.125	-0.005
S01_LP_Q15	0.581	0.574	1.012	-1.519	-1.586	0.067
CMM_LP_3	1.831	1.770	1.034	-0.343	-0.315	-0.028
S01_LP_Q18	0.979	1.018	0.961	-0.158	-0.142	-0.016
S01_LP_Q21	0.840	0.889	0.946	0.679	0.676	0.002
S01_LP_Q22	0.925	0.974	0.950	0.352	0.359	-0.008
CMM_LP_4	0.625	0.654	0.956	0.758	0.764	-0.005
S01_LP_Q24	0.512	0.544	0.941	1.058	1.050	0.009
S01_LP_Q26	0.709	0.731	0.970	1.219	1.229	-0.010
S01_LP_Q27	1.279	1.249	1.024	-0.609	-0.604	-0.005
S01_LP_Q28	1.121	1.169	0.959	0.097	0.113	-0.017
CMM_LP_5	0.961	0.991	0.970	-0.176	-0.162	-0.015
<p>* $a.ratio = \frac{a.tri}{a.mee}$</p> <p>** $b.delta = b.tri - b.mee$</p>						

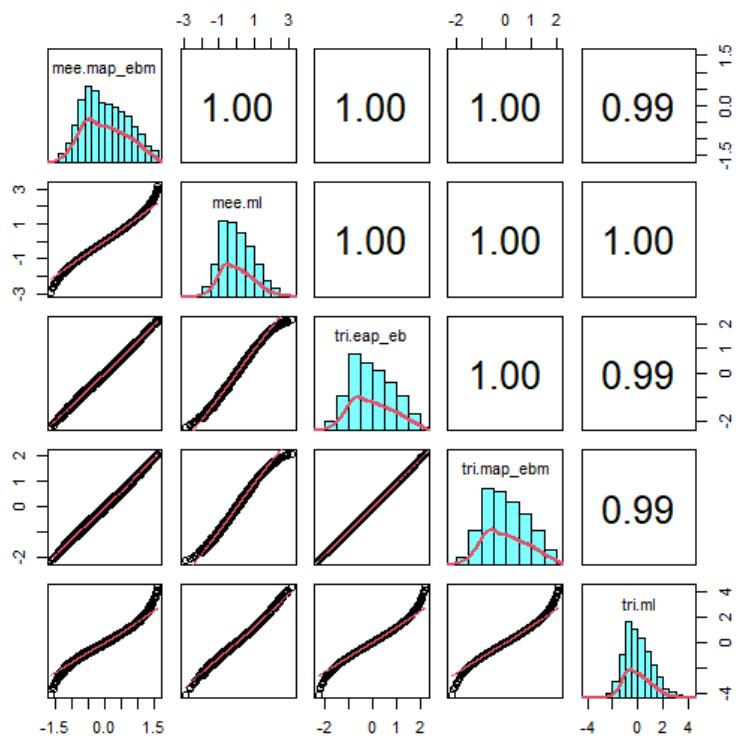


Figura 6.10 Comparação de métodos de estimação dos escores latentes – TRI e MEE. S01_LP.

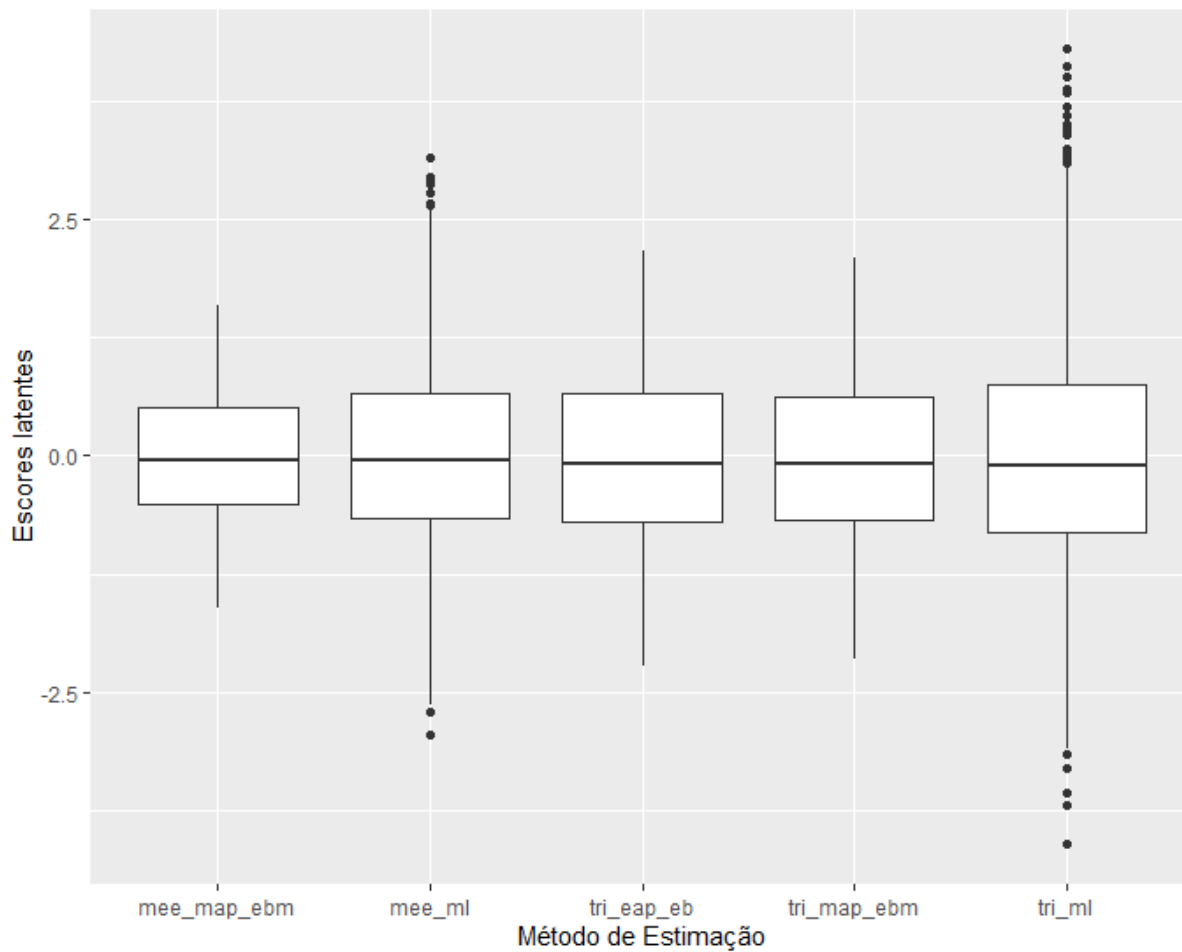


Figura 6.11 Distribuição dos escores latentes nos métodos de estimação – TRI e MEE. S01_LP - MEE.

6.3 Aplicação em modelos longitudinais.

6.3.1 Análise preliminar de seleção de itens

Tal como foi feito para o caso transversal, uma análise sobre que itens devem compor o instrumento final de mensuração deve ser realizada para cada teste. Essa seleção de itens será feita utilizando somente os métodos da TRI já que foi mostrado na seção anterior que se chega a resultados equivalentes quando se usa os métodos da MEE.

6.3.1.1 Teste S01_LP

Esta análise foi feita na seção 6.2.1 e os itens selecionados constam na Tabela 6.4. Uma análise estruturalmente idêntica foi executada para os testes S04_LP, S01_MT e S04_MT e os dados constam nas seguintes três seções.

6.3.1.2 Teste S04_LP

Tabela 6.11 Estatísticas do teste, S04_LP - TRI

Estatísticas	Valores	Valores*
Alpha de Cronbach	0.813	0.834
No de Respondentes	3148	3225
No de Itens	30	25
No Mínimo de Acertos	2	1
No Máximo de Acertos	29	25
Média de Acertos	15.49	13.6
Desvio Padrão dos acertos	5.63	5.35
% média de acertos	51.63%	54.6%
* – Valores calculados após a exclusão dos itens com $p_{bis} < 0.20$.		

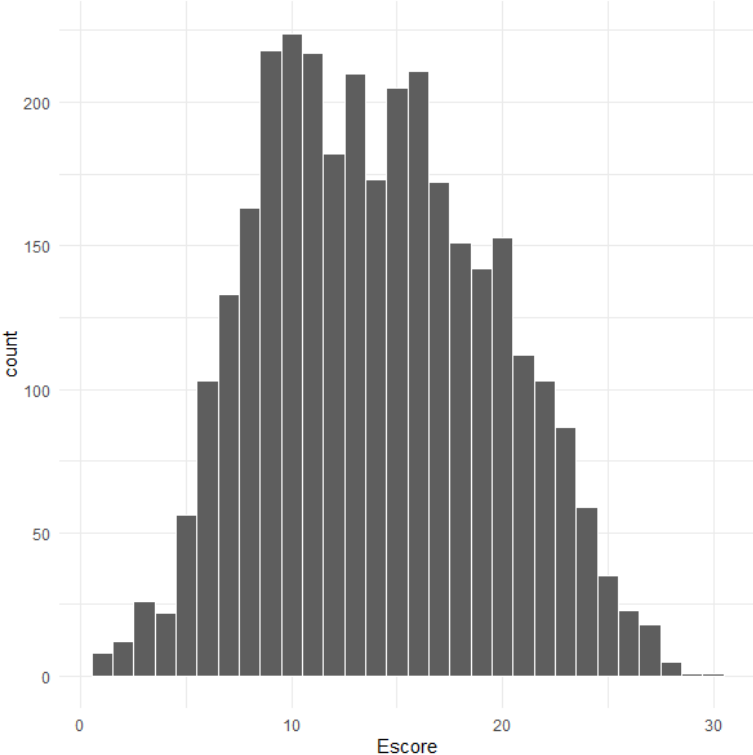


Figura 6.12 Histograma dos escores brutos. S04_LP - TRI

Tabela 6.12 Estatísticas de todos os itens. S04_LP - TRI

Items	p	pbis	a (ep)	b (ep)	Alpha se Ausente	Alpha > 0,813	pbis < 0,2	Usar?
S04_LP_Q1	0.856	0.365	1.749 (0.109)	-1.492 (0.061)	0.807	NÃO	NÃO	SIM
S04_LP_Q2	0.594	0.229	0.565 (0.045)	-0.727 (0.085)	0.811	NÃO	NÃO	SIM
S04_LP_Q3	0.658	0.499	1.823 (0.087)	-0.589 (0.033)	0.800	NÃO	NÃO	SIM
S04_LP_Q4	0.538	0.409	1.140 (0.058)	-0.188 (0.040)	0.804	NÃO	NÃO	SIM
S04_LP_Q5	0.290	0.287	0.795 (0.051)	1.274 (0.086)	0.809	NÃO	NÃO	SIM
S04_LP_Q6	0.654	0.480	1.684 (0.081)	-0.595 (0.035)	0.801	NÃO	NÃO	SIM
S04_LP_Q7	0.536	0.394	1.042 (0.055)	-0.187 (0.042)	0.804	NÃO	NÃO	SIM
S04_LP_Q8	0.474	0.417	1.181 (0.058)	0.094 (0.039)	0.803	NÃO	NÃO	SIM
S04_LP_Q9	0.334	0.222	0.534 (0.045)	1.369 (0.127)	0.811	NÃO	NÃO	SIM
S04_LP_Q10	0.593	0.450	1.372 (0.066)	-0.398 (0.037)	0.802	NÃO	NÃO	SIM
S04_LP_Q11	0.300	0.106	0.263 (0.043)	3.264 (0.536)	0.815	SIM	SIM	NÃO
S04_LP_Q12	0.451	0.274	0.664 (0.046)	0.315 (0.063)	0.809	NÃO	NÃO	SIM
S04_LP_Q13	0.571	0.302	0.759 (0.049)	-0.435 (0.057)	0.808	NÃO	NÃO	SIM
S04_LP_Q14	0.477	-0.054	-0.126 (0.039)	-0.738 (0.363)	0.822	SIM	SIM	NÃO
S04_LP_Q15	0.540	0.158	0.356 (0.041)	-0.471 (0.115)	0.814	SIM	SIM	NÃO
S04_LP_Q16	0.271	0.032	0.081 (0.044)	12.157 (6.508)	0.818	SIM	SIM	NÃO
S04_LP_Q17	0.521	0.401	1.064 (0.055)	-0.117 (0.041)	0.804	NÃO	NÃO	SIM
S04_LP_Q18	0.472	0.411	1.092 (0.056)	0.111 (0.041)	0.804	NÃO	NÃO	SIM
S04_LP_Q19	0.510	0.420	1.151 (0.058)	-0.065 (0.039)	0.803	NÃO	NÃO	SIM
S04_LP_Q20	0.431	0.349	0.914 (0.051)	0.343 (0.049)	0.806	NÃO	NÃO	SIM
S04_LP_Q21	0.255	0.153	0.394 (0.046)	2.814 (0.325)	0.813	SIM	SIM	NÃO
S04_LP_Q22	0.307	0.303	0.793 (0.050)	1.158 (0.081)	0.808	NÃO	NÃO	SIM
CMM_LP_1	0.559	0.380	1.019 (0.055)	-0.302 (0.044)	0.805	NÃO	NÃO	SIM
CMM_LP_2	0.781	0.397	1.491 (0.083)	-1.192 (0.052)	0.805	NÃO	NÃO	SIM

Items	p	pbis	a (ep)	b (ep)	Alpha se Ausente	Alpha > 0,813	pbis < 0,2	Usar?
CMM_LP_3	0.669	0.532	2.168 (0.105)	-0.597 (0.031)	0.799	NÃO	NÃO	SIM
CMM_LP_4	0.503	0.381	0.990 (0.053)	-0.030 (0.043)	0.805	NÃO	NÃO	SIM
S04_LP_Q27	0.730	0.432	1.516 (0.079)	-0.939 (0.044)	0.803	NÃO	NÃO	SIM
S04_LP_Q28	0.725	0.389	1.277 (0.069)	-1.004 (0.051)	0.805	NÃO	NÃO	SIM
S04_LP_Q29	0.392	0.318	0.811 (0.049)	0.609 (0.060)	0.808	NÃO	NÃO	SIM
CMM_LP_5	0.497	0.212	0.486 (0.043)	0.025 (0.078)	0.812	NÃO	NÃO	SIM
p – porcentagem de acerto do item a – parâmetro de discriminação do item b – parâmetro de dificuldade do item p_bis – correlação ponto bisserial ep – erro padrão								

6.3.1.3 Teste S01_MT

Tabela 6.13 Estatísticas do teste, S01_MT

Estatísticas	Valores	Valores*
Alpha de Cronbach	0.734	0.762
No de Respondentes	3208	3208
No de Itens	30	23
No Mínimo de Acertos	1	0
No Máximo de Acertos	28	23
Média de Acertos	12.61	10.68
Desvio Padrão dos acertos	4.52	4.11
% média de acertos	42.05%	46.42%
* – Valores calculados após a exclusão dos itens com pbis < 0.20.		

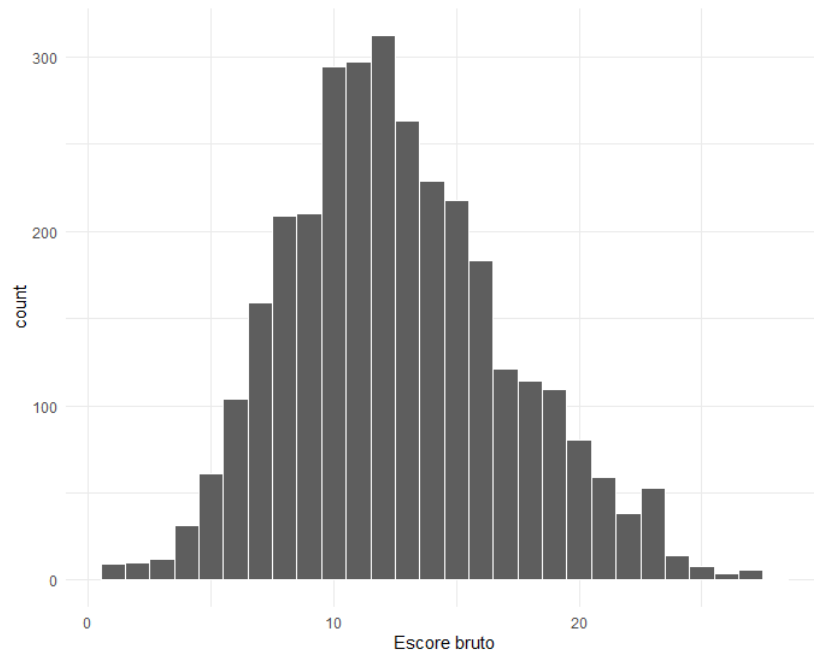


Figura 6.13 Histograma dos escores brutos. S01_MT

Tabela 6.14 Estatísticas de todos os itens. S01_MT

Itens	p	pbis	a (ep)	b (ep)	Alpha se Ausente	Alpha > 0,734	pbis < 0,2	Usar?
S01_MT_Q1	0.756	0.315	1.164 (0.072)	-1.222 (0.066)	0.724	NÃO	NÃO	SIM
S01_MT_Q2	0.448	0.276	0.734 (0.050)	0.307 (0.057)	0.726	NÃO	NÃO	SIM
CMM_MT_1	0.859	0.258	1.272 (0.089)	-1.798 (0.095)	0.728	NÃO	NÃO	SIM
S01_MT_Q4	0.565	0.421	1.485 (0.078)	-0.268 (0.034)	0.716	NÃO	NÃO	SIM
CMM_MT_2	0.322	0.311	0.891 (0.055)	0.964 (0.068)	0.724	NÃO	NÃO	SIM
S01_MT_Q6	0.475	0.235	0.580 (0.046)	0.181 (0.067)	0.729	NÃO	NÃO	SIM
S01_MT_Q7	0.456	0.318	0.852 (0.053)	0.226 (0.050)	0.723	NÃO	NÃO	SIM
S01_MT_Q8	0.153	0.225	0.684 (0.059)	2.727 (0.213)	0.729	NÃO	NÃO	SIM
S01_MT_Q9	0.845	0.316	1.556 (0.100)	-1.507 (0.068)	0.725	NÃO	NÃO	SIM
S01_MT_Q10	0.241	0.158	0.375 (0.048)	3.156 (0.395)	0.733	NÃO	SIM	NÃO
S01_MT_Q11	0.509	0.361	1.029 (0.059)	-0.057 (0.042)	0.720	NÃO	NÃO	SIM
S01_MT_Q12	0.379	0.273	0.705 (0.049)	0.769 (0.073)	0.726	NÃO	NÃO	SIM
S01_MT_Q13	0.673	0.383	1.314 (0.074)	-0.740 (0.044)	0.719	NÃO	NÃO	SIM
S01_MT_Q14	0.337	0.064	0.142 (0.042)	4.793 (1.436)	0.740	SIM	SIM	NÃO
S01_MT_Q15	0.387	0.087	0.209 (0.041)	2.215 (0.464)	0.739	SIM	SIM	NÃO
S01_MT_Q16	0.597	0.378	1.279 (0.070)	-0.422 (0.039)	0.719	NÃO	NÃO	SIM
S01_MT_Q17	0.137	0.094	0.279 (0.057)	6.675 (1.339)	0.736	SIM	SIM	NÃO
S01_MT_Q18	0.301	0.310	0.826 (0.054)	1.154 (0.080)	0.724	NÃO	NÃO	SIM
CMM_MT_3	0.434	0.211	0.524 (0.045)	0.537 (0.084)	0.731	NÃO	NÃO	SIM
S01_MT_Q20	0.278	0.250	0.639 (0.050)	1.621 (0.128)	0.728	NÃO	NÃO	SIM
CMM_MT_4	0.139	0.259	0.845 (0.064)	2.440 (0.160)	0.728	NÃO	NÃO	SIM
S01_MT_Q22	0.325	0.129	0.318 (0.044)	2.353 (0.332)	0.735	SIM	SIM	NÃO
S01_MT_Q23	0.588	0.316	0.921 (0.057)	-0.467 (0.050)	0.723	NÃO	NÃO	SIM
S01_MT_Q24	0.237	0.035	0.104 (0.047)	11.331 (5.090)	0.740	SIM	SIM	NÃO

Itens	p	pbis	a (ep)	b (ep)	Alpha se Ausente	Alpha > 0,734	pbis < 0,2	Usar?
S01_MT_Q25	0.276	0.043	0.112 (0.044)	8.627 (3.415)	0.740	SIM	SIM	NÃO
S01_MT_Q26	0.493	0.365	1.048 (0.059)	0.019 (0.041)	0.720	NÃO	NÃO	SIM
S01_MT_Q27	0.194	0.231	0.642 (0.054)	2.407 (0.190)	0.729	NÃO	NÃO	SIM
S01_MT_Q28	0.089	0.208	0.790 (0.073)	3.268 (0.259)	0.731	NÃO	NÃO	SIM
S01_MT_Q29	0.289	0.320	0.902 (0.056)	1.154 (0.075)	0.723	NÃO	NÃO	SIM
S01_MT_Q30	0.834	0.305	1.442 (0.093)	-1.501 (0.071)	0.725	NÃO	NÃO	SIM
<p>p – porcentagem de acerto do item a – parâmetro de discriminação do item b – parâmetro de dificuldade do item p_bis – correlação ponto bisserial ep – erro padrão</p>								

6.3.1.4 Teste S04_MT

Tabela 6.15 Estatísticas do teste, S04_MT

Estatísticas	Valores	Valores*
Alpha de Cronbach	0.778	0.788
No de Respondentes	3164	3164
No de Itens	30	25
No Mínimo de Acertos	3	2
No Máximo de Acertos	30	25
Média de Acertos	13.87	12.19
Desvio Padrão dos acertos	4.92	4.55
% média de acertos	46.22%	48.78%
* – Valores calculados após a exclusão dos itens com pbis < 0.20.		

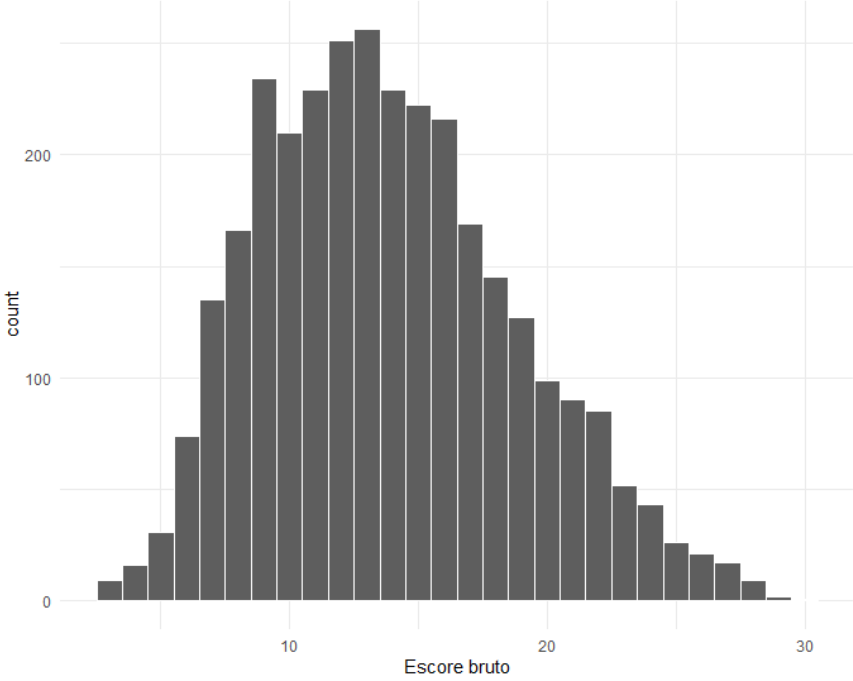


Figura 6.14 Histograma dos escores brutos. S04_MT

Tabela 6.16 Estatísticas de todos os itens. S04_MT

Itens	p	pbis	a (ep)	b (ep)	Alpha se Ausente	Alpha > 0,778	pbis < 0,2	Usar?
S04_MT_Q1	0.239	0.383	1.068 (0.061)	1.312 (0.071)	0.767	NÃO	NÃO	SIM
COMUM_MT_3	0.582	0.327	0.874 (0.054)	-0.456 (0.051)	0.770	NÃO	NÃO	SIM
S04_MT_Q3	0.406	0.472	1.484 (0.073)	0.328 (0.036)	0.762	NÃO	NÃO	SIM
S04_MT_Q4	0.228	0.132	0.343 (0.047)	3.652 (0.495)	0.779	SIM	SIM	NÃO
S04_MT_Q5	0.591	0.160	0.397 (0.044)	-0.969 (0.136)	0.778	SIM	SIM	NÃO
S04_MT_Q6	0.206	0.212	0.522 (0.051)	2.732 (0.254)	0.775	NÃO	NÃO	SIM
S04_MT_Q7	0.228	0.176	0.437 (0.048)	2.905 (0.312)	0.777	NÃO	SIM	NÃO
S04_MT_Q8	0.157	0.261	0.765 (0.059)	2.444 (0.168)	0.773	NÃO	NÃO	SIM
S04_MT_Q9	0.660	0.368	1.222 (0.068)	-0.718 (0.045)	0.768	NÃO	NÃO	SIM
S04_MT_Q10	0.307	0.266	0.650 (0.048)	1.363 (0.108)	0.773	NÃO	NÃO	SIM
S04_MT_Q11	0.213	0.196	0.552 (0.050)	2.516 (0.220)	0.776	NÃO	SIM	SIM
S04_MT_Q12	0.405	0.232	0.572 (0.045)	0.716 (0.085)	0.775	NÃO	NÃO	SIM
S04_MT_Q13	0.266	0.347	0.925 (0.056)	1.275 (0.077)	0.769	NÃO	NÃO	SIM
S04_MT_Q14	0.149	0.269	0.739 (0.059)	2.602 (0.187)	0.773	NÃO	NÃO	SIM
S04_MT_Q15	0.148	0.072	0.194 (0.055)	9.117 (2.549)	0.780	SIM	SIM	NÃO
S04_MT_Q16	0.454	0.210	0.492 (0.044)	0.388 (0.083)	0.776	NÃO	NÃO	SIM
S04_MT_Q17	0.384	0.327	0.828 (0.051)	0.641 (0.060)	0.770	NÃO	NÃO	SIM
S04_MT_Q18	0.479	0.075	0.177 (0.040)	0.473 (0.228)	0.783	SIM	SIM	NÃO
S04_MT_Q19	0.774	0.389	1.733 (0.100)	-1.070 (0.046)	0.767	NÃO	NÃO	SIM
S04_MT_Q20	0.430	0.367	1.004 (0.056)	0.322 (0.046)	0.768	NÃO	NÃO	SIM
S04_MT_Q21	0.810	0.291	1.204 (0.079)	-1.516 (0.079)	0.772	NÃO	NÃO	SIM
S04_MT_Q22	0.816	0.278	1.147 (0.078)	-1.607 (0.087)	0.772	NÃO	NÃO	SIM

Itens	p	pbis	a (ep)	b (ep)	Alpha se Ausente	Alpha > 0,778	pbis < 0,2	Usar?
S04_MT_Q23	0.570	0.264	0.703 (0.049)	-0.456 (0.062)	0.773	NÃO	NÃO	SIM
COMUM_MT_1	0.898	0.266	1.581 (0.118)	-1.861 (0.092)	0.773	NÃO	NÃO	SIM
COMUM_MT_2	0.577	0.361	1.065 (0.059)	-0.381 (0.043)	0.768	NÃO	NÃO	SIM
COMUM_MT_4	0.302	0.373	1.021 (0.058)	0.979 (0.061)	0.767	NÃO	NÃO	SIM
S04_MT_Q27	0.725	0.375	1.436 (0.080)	-0.940 (0.047)	0.767	NÃO	NÃO	SIM
S04_MT_Q28	0.611	0.382	1.252 (0.067)	-0.494 (0.040)	0.767	NÃO	NÃO	SIM
S04_MT_Q29	0.656	0.378	1.258 (0.069)	-0.691 (0.044)	0.767	NÃO	NÃO	SIM
S04_MT_Q30	0.598	0.317	0.898 (0.055)	-0.536 (0.052)	0.770	NÃO	NÃO	SIM
p – porcentagem de acerto do item a – parâmetro de discriminação do item b – parâmetro de dificuldade do item p_bis – correlação ponto bisserial ep – erro padrão								

6.3.2 Análise utilizando a TRI e a TCT

Conforme mencionado anteriormente, a análise com os métodos da TRI se dá com a combinação de dois modelos unidimensionais longitudinais de dois momentos. Um abordando os testes de Língua Portuguesa (S01_LP e S04_LP) e outro, os testes de Matemática (S01_MT e S04_MT).

6.3.2.1 Modelo unidimensional longitudinal aplicado aos testes de Língua Portuguesa

Os itens dos testes de Língua Portuguesa foram equalizados via o processo de estimação simultânea em modelos de grupos múltiplos discutido na Seção 3.6.2. Os testes S01_LP e S04_LP compartilham 5 itens e têm, respectivamente, 23 e 25 itens cada. A estrutura dos dados que é passada para o pacote MIRT é como descrito na Figura 6.15. Apesar de estarmos avaliando o mesmo grupo de indivíduos nos dois momentos, nesse tipo de modelagem os indivíduos são classificados como pertencentes a diferentes grupos para possibilitar a estimação de como os parâmetros populacionais diferem entre eles. Nesse tipo de abordagem, ambos os testes são modelados como

tendo 43 questões. No teste S01_LP, 23 itens são aplicados aos alunos e 20 não, enquanto no teste S04_LP, 25 são aplicados e 18 não. Itens não aplicados diferem de itens respondidos erroneamente em que eles não interferem no escore bruto ou latente do aluno, ou seja, o efeito prático disso é que eles não contribuem com a mensuração do traço latente de interesse.

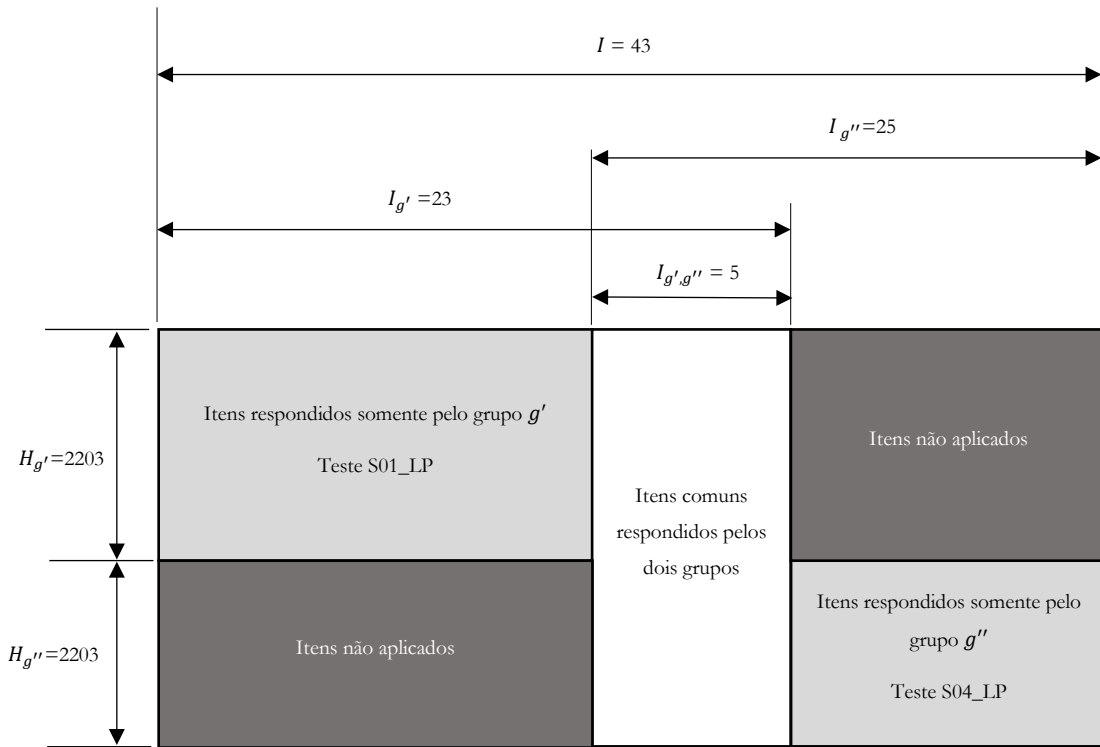


Figura 6.15 Estrutura de grupos múltiplos de dados para equalização simultânea dos testes S01_LP e S04_LP – Modelo unidimensional longitudinal TRI.

Tabela 6.17 Parâmetros dos itens - Modelo unidimensional longitudinal - TRI - S01_LP e S04_LP

Itens	a (ep)	b (ep)	Itens	a (ep)	b (ep)
CMM_LP_01	0.821 (0.041)	-0.054 (0.045)	S01_LP_Q28	1.146 (0.070)	0.007 (0.046)
CMM_LP_02	1.296 (0.063)	-1.094 (0.052)	S04_LP_Q01	1.524 (0.116)	-1.395 (0.087)
CMM_LP_03	1.934 (0.082)	-0.416 (0.031)	S04_LP_Q02	0.459 (0.046)	-0.543 (0.128)
CMM_LP_04	0.786 (0.039)	0.371 (0.047)	S04_LP_Q03	1.550 (0.093)	-0.395 (0.049)
CMM_LP_05	0.576 (0.037)	-0.059 (0.059)	S04_LP_Q04	0.989 (0.062)	0.107 (0.057)
S01_LP_Q01	1.188 (0.072)	-0.149 (0.045)	S04_LP_Q05	0.662 (0.053)	1.840 (0.128)
S01_LP_Q02	0.795 (0.061)	-0.661 (0.073)	S04_LP_Q06	1.370 (0.083)	-0.414 (0.053)
S01_LP_Q04	1.139 (0.080)	-1.160 (0.074)	S04_LP_Q07	0.915 (0.060)	0.079 (0.060)
S01_LP_Q06	1.677 (0.094)	0.551 (0.042)	S04_LP_Q08	0.947 (0.060)	0.437 (0.059)
S01_LP_Q09	1.710 (0.094)	0.007 (0.037)	S04_LP_Q09	0.448 (0.046)	1.929 (0.183)
S01_LP_Q10	1.394 (0.082)	-0.398 (0.043)	S04_LP_Q10	1.149 (0.070)	-0.176 (0.054)
S01_LP_Q11	1.303 (0.078)	-0.394 (0.045)	S04_LP_Q12	0.514 (0.046)	0.681 (0.096)
S01_LP_Q12	0.948 (0.065)	1.048 (0.078)	S04_LP_Q13	0.635 (0.051)	-0.264 (0.086)
S01_LP_Q13	1.155 (0.073)	-0.558 (0.052)	S04_LP_Q17	0.898 (0.059)	0.191 (0.060)
S01_LP_Q14	0.625 (0.054)	0.087 (0.074)	S04_LP_Q18	0.867 (0.057)	0.467 (0.062)
S01_LP_Q15	0.544 (0.058)	-1.638 (0.178)	S04_LP_Q19	0.969 (0.061)	0.269 (0.057)
S01_LP_Q18	0.995 (0.066)	-0.252 (0.052)	S04_LP_Q20	0.710 (0.052)	0.704 (0.075)
S01_LP_Q21	0.879 (0.061)	0.584 (0.066)	S04_LP_Q22	0.657 (0.053)	1.696 (0.120)
S01_LP_Q22	0.987 (0.064)	0.192 (0.053)	S04_LP_Q27	1.338 (0.086)	-0.751 (0.062)
S01_LP_Q24	0.487 (0.052)	0.983 (0.134)	S04_LP_Q28	1.118 (0.075)	-0.879 (0.075)
S01_LP_Q26	0.726 (0.058)	1.107 (0.101)	S04_LP_Q29	0.752 (0.053)	0.931 (0.078)
S01_LP_Q27	1.315 (0.083)	-0.718 (0.051)			

a – parâmetro de discriminação do item
b – parâmetro de dificuldade do item
ep – erro padrão

Tabela 6.18 Parâmetros da variável latente - Modelo unidimensional longitudinal - TRI - S01_LP e S04_LP. Erro padrão entre parêntesis (não calculado para momento de referência).

Momento	Média	Variância
Momento 1	0 (NA)	1 (NA)
Momento 2	0.367 (0.036)	1.375 (0.077)

6.3.2.2 Modelo unidimensional longitudinal aplicado aos testes de Matemática

Os itens dos testes de Matemática foram equalizados via o processo de estimação simultânea em modelos de grupos múltiplos, conforme discutido na Seção 3.6.2. Os testes S01_MT e S04_MT compartilham 4 itens e têm, respectivamente, 23 e 25 itens cada. A estrutura dos dados que é passada para o pacote MIRT é como descrito na Figura 6.16.

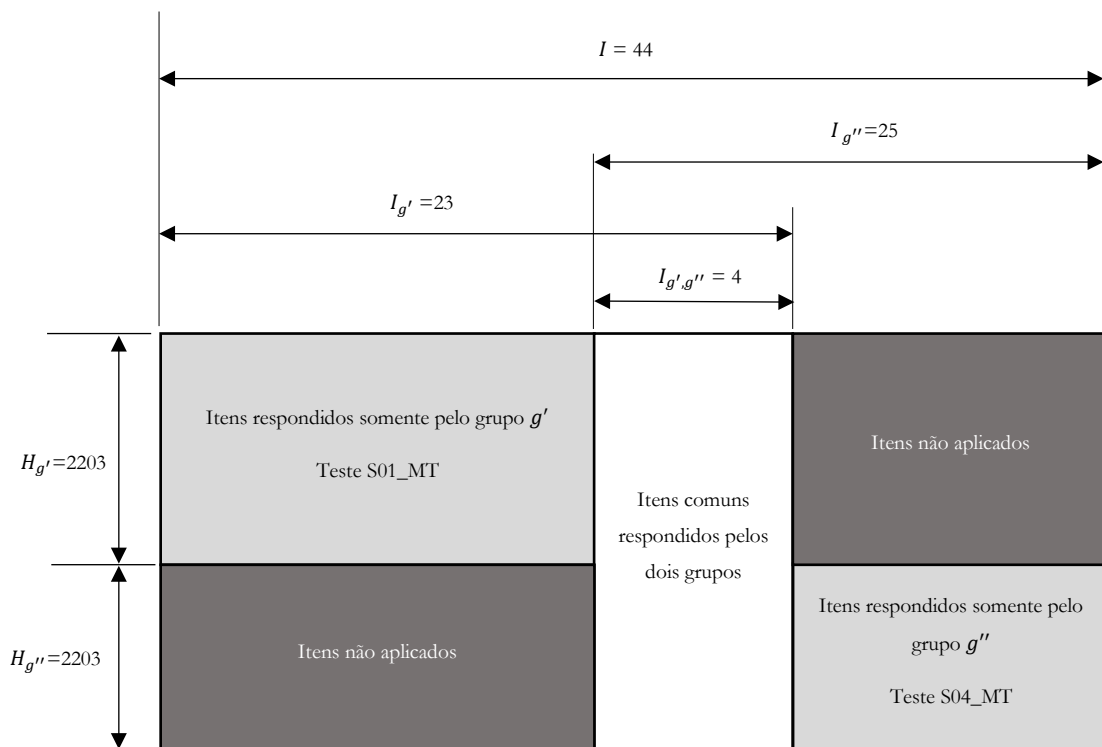


Figura 6.16 Estrutura de grupos múltiplos de dados para equalização simultânea dos testes S01_MT e S04_MT – Modelo unidimensional longitudinal TRI.

Tabela 6.19 Parâmetros dos itens - Modelo unidimensional longitudinal - TRI - S01_MT e S04_MT

Itens	a (ep)	b (ep)	Itens	a (ep)	b (ep)
CMM_MT_01	1.070 (0.071)	-1.961 (0.115)	S01_MT_Q30	1.398 (0.113)	-1.693 (0.100)
CMM_MT_02	0.880 (0.040)	0.703 (0.046)	S04_MT_Q01	0.762 (0.054)	2.990 (0.128)
CMM_MT_03	0.595 (0.033)	0.414 (0.059)	S04_MT_Q03	1.093 (0.067)	1.590 (0.066)
CMM_MT_04	0.818 (0.040)	2.299 (0.085)	S04_MT_Q06	0.355 (0.043)	4.778 (0.422)
S01_MT_Q01	1.138 (0.087)	-1.304 (0.086)	S04_MT_Q08	0.589 (0.053)	4.416 (0.263)
S01_MT_Q02	0.735 (0.060)	0.202 (0.067)	S04_MT_Q09	0.910 (0.062)	0.093 (0.080)
S01_MT_Q04	1.531 (0.098)	-0.338 (0.041)	S04_MT_Q10	0.478 (0.042)	2.819 (0.166)
S01_MT_Q06	0.607 (0.057)	0.124 (0.077)	S04_MT_Q11	0.406 (0.045)	4.659 (0.368)
S01_MT_Q07	0.854 (0.064)	0.164 (0.059)	S04_MT_Q12	0.484 (0.041)	1.935 (0.118)
S01_MT_Q08	0.762 (0.073)	2.394 (0.204)	S04_MT_Q13	0.635 (0.048)	2.892 (0.139)
S01_MT_Q09	1.528 (0.123)	-1.643 (0.092)	S04_MT_Q14	0.499 (0.050)	4.817 (0.338)
S01_MT_Q11	1.027 (0.071)	-0.164 (0.051)	S04_MT_Q16	0.349 (0.038)	1.656 (0.141)
S01_MT_Q12	0.679 (0.059)	0.743 (0.090)	S04_MT_Q17	0.576 (0.044)	1.912 (0.103)
S01_MT_Q13	1.286 (0.088)	-0.781 (0.056)	S04_MT_Q19	1.227 (0.088)	-0.415 (0.085)
S01_MT_Q16	1.266 (0.085)	-0.563 (0.050)	S04_MT_Q20	0.762 (0.051)	1.470 (0.077)
S01_MT_Q18	0.983 (0.071)	0.996 (0.077)	S04_MT_Q21	0.847 (0.069)	-1.033 (0.144)
S01_MT_Q20	0.678 (0.061)	1.461 (0.135)	S04_MT_Q22	0.851 (0.072)	-1.215 (0.158)
S01_MT_Q23	0.949 (0.070)	-0.524 (0.060)	S04_MT_Q23	0.506 (0.043)	0.374 (0.111)
S01_MT_Q26	1.062 (0.072)	-0.094 (0.049)	S04_MT_Q27	1.018 (0.072)	-0.280 (0.088)
S01_MT_Q27	0.736 (0.068)	2.106 (0.180)	S04_MT_Q28	0.892 (0.060)	0.372 (0.074)
S01_MT_Q28	0.757 (0.087)	3.292 (0.330)	S04_MT_Q29	0.882 (0.061)	0.088 (0.082)
S01_MT_Q29	1.001 (0.072)	1.014 (0.077)	S04_MT_Q30	0.648 (0.049)	0.324 (0.093)
a – parâmetro de discriminação do item b – parâmetro de dificuldade do item ep – erro padrão					

Tabela 6.20 Parâmetros da variável latente - Modelo unidimensional longitudinal - TRI - S01_MP e S04_MP. Erro padrão entre parêntesis (não calculado para momento de referência).

Momento	Média (erro padrão)	Variância (erro padrão)
Momento 1	0 (NA)	1 (NA)
Momento 2	1.176 (0.045)	1.961 (0.110)

Tabela 6.21 Parâmetros e correspondentes estimativas da distribuição das variáveis latentes. TRI. Valores estimados. Erro padrão entre parêntesis (não calculado para momento de referência).

Parâmetros populacionais	Notação	Valores Estimados (erro padrão)
Média μ	$\begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{21} \\ \alpha_{22} \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ (NA) \\ 0.367 \\ (0.036) \\ 0.000 \\ (NA) \\ 1.176 \\ (0.045) \end{pmatrix}$
Matriz de Covariância	$\begin{pmatrix} \psi_{11,11} & & & \\ \psi_{21,11} & \psi_{21,21} & & \\ \psi_{12,11} & \psi_{12,21} & \psi_{12,12} & \\ \psi_{22,11} & \psi_{22,21} & \psi_{12,22} & \psi_{22,22} \end{pmatrix}$	$\begin{pmatrix} 1 \\ (NA) & & & \\ NA & 1.375 & & \\ & (0.077) & & \\ NA & NA & 1 & \\ NA & NA & (NA) & 1.961 \\ & & NA & (0.110) \end{pmatrix}$

Tabela 6.22 Parâmetros e correspondentes estimativas da distribuição das variáveis latentes. TRI. Valores calculados a partir dos escores latentes individuais.

Parâmetros populacionais	Notação	Valores estimados
Média μ	$\begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{21} \\ \alpha_{22} \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ 0.367 \\ 0.000 \\ 1.176 \end{pmatrix}$
Matriz de Covariância	$\begin{pmatrix} \psi_{11,11} & & & \\ \psi_{21,11} & \psi_{21,21} & & \\ \psi_{12,11} & \psi_{12,21} & \psi_{12,12} & \\ \psi_{22,11} & \psi_{22,21} & \psi_{12,22} & \psi_{22,22} \end{pmatrix}$	$\begin{pmatrix} 0.828 & & & \\ 0.678 & 1.152 & & \\ 0.500 & 0.531 & 0.776 & \\ 0.623 & 0.826 & 0.693 & 1.152 \end{pmatrix}$
Matriz de Correlação	$\begin{pmatrix} \rho_{11,11} & & & \\ \rho_{21,11} & \rho_{21,21} & & \\ \rho_{12,11} & \rho_{12,21} & \rho_{12,12} & \\ \rho_{22,11} & \rho_{22,21} & \rho_{12,22} & \rho_{22,22} \end{pmatrix}$	$\begin{pmatrix} 1 & & & \\ 0.694 & 1 & & \\ 0.624 & 0.561 & 1 & \\ 0.542 & 0.632 & 0.624 & 1 \end{pmatrix}$

6.3.3 Análise utilizando a MEE

O modelo adotado nesta seção assume um modelo multidimensional longitudinal combinando dois modelos unidimensionais longitudinais ao permitir que as diferentes variáveis latentes interfiram entre si através de uma matriz de covariância como indicado na Figura 5.4. A relação entre as variáveis latentes da figura e os testes do Programa Escola+ é como descrito a seguir:

- η_{11} é o traço latente relacionado às habilidades em Língua Portuguesa medido com o teste S01_LP,
- η_{21} é o traço latente relacionado às habilidades em Matemática medido com o teste S01_MT,
- η_{12} é o traço latente relacionado às habilidades em Língua Portuguesa medido com o teste S04_LP,
- η_{22} é o traço latente relacionado às habilidades em Matemática medido com o teste S04_MT.

A matriz de covariância que relacionada as variáveis latentes do modelo pode ser tão estruturada quanto se queria, mas, neste experimento, foi assumida uma matriz de covariância não estruturada. Foi assumida também que a invariância fatorial das cargas fatoriais e dos valores limiares dos itens comuns é válida. Por último, foi considerada a existência de uma covariância entre os itens comuns correspondentes de diferentes momentos.

Cada variável latente foi medida com os mesmos itens utilizados durante a análise de grupos múltiplos feita na seção anterior utilizando métodos da TRI, mas a estrutura de dados passada ao pacote lavaan é substancialmente diferente da utilizada lá. Por se tratar de uma modelagem longitudinal com um único grupo, todos os indivíduos pertencem a um mesmo grupo e responderam a todos os quatro testes tal como ilustrado na Figura 6.17. A atual versão do pacote lavaan (0.6-14) não consegue processar dados faltantes sistemáticos e, por isso, estudantes que responderam a menos de três provas foram excluídos da análise.

Foram utilizados na mensuração de uma variável latente com a parametrização RI_CP_FT (Reference Indicator, Conditional Parameterization, Free Threshold). As medidas de ajuste do modelo estão listadas na Tabela 6.21. O modelo está falhando no teste qui-quadrado e isso, como ressaltado anteriormente, provavelmente, se dá porque o valor da estatística χ^2 , conforme definição na Equação (4.26), depende do tamanho da amostra e o tamanho da amostra deste experimento é muito grande. Além disso, a análise da matriz de covariância dos resíduos, bem como dos índices de modificação, não indicou discrepâncias que, se corrigidas, poderiam alterar significativamente o valor de χ^2 . Valores dos índices de ajuste CFI, TLI e RMSEA indicam um bom ajuste.

A Tabela 6.23 e a Tabela 6.26 mostram os parâmetros da distribuição das variáveis latentes. Um ponto de grande interesse é o alto valor da correlação cruzada existente entre as variáveis latentes de diferentes disciplinas entre diferentes momentos. Os valores das correlações entre η_{11} e η_{22} , entre η_{21} e η_{12} é superior a 0.68.

A carga fatorial e o valor limiar dos itens associados a cada variável é mostrado nas colunas λ_{it} e τ_{it} do conjunto de tabelas que iniciam em Tabela 6.22 e terminam na Tabela 6.27. A partir desses valores é possível calcular os parâmetros de discriminação a_{it} e dificuldade b_{it} dos indicadores através da aplicação das Equações (E.14) e (E.15) com $v_{it} = 0$ e $\theta_{it} = 1$ dado que a parametrização RI_CP_FT foi escolhida.

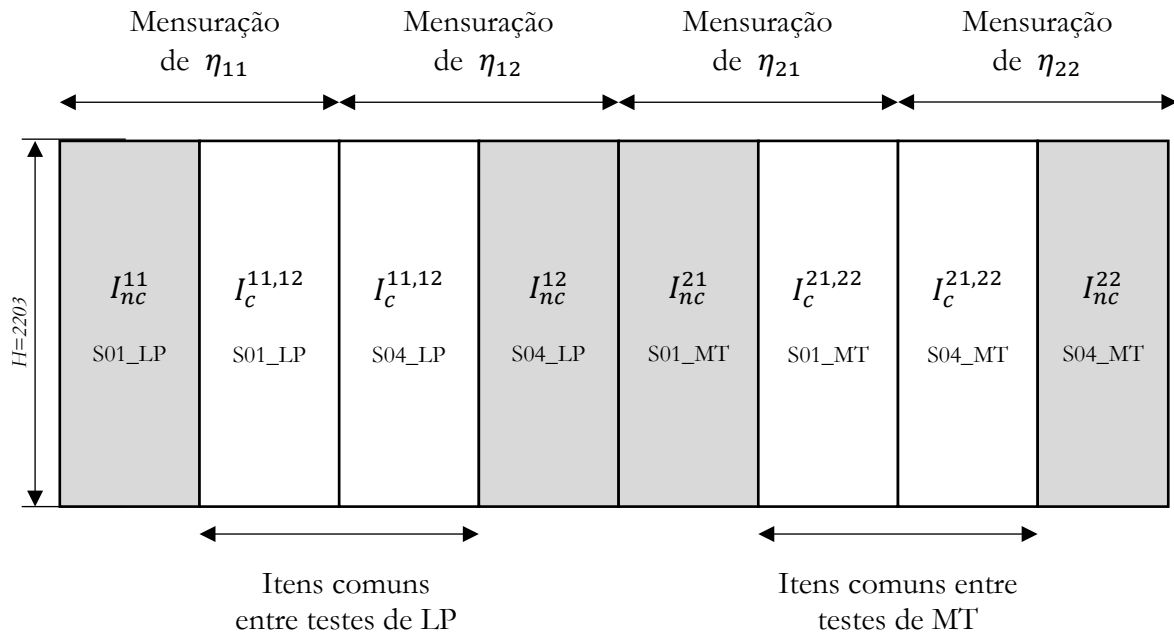


Figura 6.17 Estrutura de dados para equalização simultânea de quatro testes: S01_LP, S04_LP, S01_MT E S04_MT – modelo de um grupo.

Tabela 6.23 Medidas de ajuste. Modelo multidimensional longitudinal. MEE.

Descrição	Valor
Número de observações	2203
Número de parâmetros	191
χ^2	8473.56
Graus de Liberdade	4465
valor-p	0
CFI	0.978
TLI	0.978
RMSEA	0.020

Tabela 6.24 Parâmetros e correspondentes estimativas da distribuição das variáveis latentes. Modelo multidimensional longitudinal. MEE. Valores não padronizados.

Parâmetros populacionais	Notação	Valores Estimados
Média μ	$\begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{21} \\ \alpha_{22} \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ (0.000) \\ 0.187 \\ (0.020) \\ 0.000 \\ (0.000) \\ 0.554 \\ (0.025) \end{pmatrix}$
Matriz de Covariância	$\begin{pmatrix} \psi_{11,11} & & & \\ \psi_{21,11} & \psi_{21,21} & & \\ \psi_{12,11} & \psi_{12,21} & \psi_{12,12} & \\ \psi_{22,11} & \psi_{22,21} & \psi_{12,22} & \psi_{22,22} \end{pmatrix}$	$\begin{pmatrix} 0.444 \\ (0.017) & & & \\ 0.358 & 0.423 & & \\ (0.013) & (0.015) & & \\ 0.286 & 0.249 & 0.302 & \\ (0.011) & (0.010) & (0.013) & \\ 0.256 & 0.284 & 0.248 & 0.326 \\ (0.011) & (0.011) & (0.010) & (0.013) \end{pmatrix}$
Matriz de Correlação	$\begin{pmatrix} \rho_{11,11} & & & \\ \rho_{21,11} & \rho_{21,21} & & \\ \rho_{12,11} & \rho_{12,21} & \rho_{12,12} & \\ \rho_{22,11} & \rho_{22,21} & \rho_{12,22} & \rho_{22,22} \end{pmatrix}$	$\begin{pmatrix} 1 & & & \\ 0.827 & 1 & & \\ 0.781 & 0.696 & 1 & \\ 0.672 & 0.764 & 0.789 & 1 \end{pmatrix}$

Tabela 6.25 Parâmetros e correspondentes estimativas da distribuição das variáveis latentes. Modelo multidimensional longitudinal. MEE. Valores padronizados.

Parâmetros populacionais	Notação	Valores estimados padronizados (erro padrão)
Média μ	$\begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{21} \\ \alpha_{22} \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ (0.000) \\ 0.288 \\ (0.026) \\ 0.000 \\ (0.000) \\ 0.970 \\ (0.047) \end{pmatrix}$
Matriz de Covariância	$\begin{pmatrix} \psi_{11,11} & & & \\ \psi_{21,11} & \psi_{21,21} & & \\ \psi_{12,11} & \psi_{12,21} & \psi_{12,12} & \\ \psi_{22,11} & \psi_{22,21} & \psi_{12,22} & \psi_{22,22} \end{pmatrix}$	$\begin{pmatrix} 1 & & & \\ (0.000) & & & \\ 0.827 & 1 & & \\ (0.015) & (0.000) & & \\ 0.781 & 0.696 & 1 & \\ (0.013) & (0.015) & (0.000) & \\ 0.672 & 0.764 & 0.789 & 1 \\ (0.017) & (0.015) & (0.015) & (0.000) \end{pmatrix}$
Matriz de Correlação	$\begin{pmatrix} \rho_{11,11} & & & \\ \rho_{21,11} & \rho_{21,21} & & \\ \rho_{12,11} & \rho_{12,21} & \rho_{12,12} & \\ \rho_{22,11} & \rho_{22,21} & \rho_{12,22} & \rho_{22,22} \end{pmatrix}$	$\begin{pmatrix} 1 & & & \\ (0.000) & & & \\ 0.827 & 1 & & \\ (0.015) & (0.000) & & \\ 0.781 & 0.696 & 1 & \\ (0.013) & (0.015) & (0.000) & \\ 0.672 & 0.764 & 0.789 & 1 \\ (0.017) & (0.015) & (0.015) & (0.000) \end{pmatrix}$

Tabela 6.26 Parâmetros dos itens do teste S01_LP. Modelo multidimensional longitudinal MEE.

Itens	λ_{it} (ep)	τ_{it} (ep)	a_{it}	b_{it}
S01_LP_Q01	1.093 (0.062)	-0.094 (0.033)	1.239	-0.129
S01_LP_Q02	0.729 (0.052)	-0.316 (0.030)	0.827	-0.650
S01_LP_Q04	0.864 (0.056)	-0.745 (0.034)	0.980	-1.294
S01_LP_Q06	1.499 (0.091)	0.564 (0.043)	1.699	0.564
S01_LP_Q09	1.464 (0.081)	0.036 (0.037)	1.660	0.037
S01_LP_Q10	1.177 (0.065)	-0.304 (0.034)	1.334	-0.388
S01_LP_Q11	1.101 (0.061)	-0.284 (0.034)	1.248	-0.387
S01_LP_Q12	0.914 (0.067)	0.612 (0.035)	1.036	1.005
S01_LP_Q13	1.013 (0.058)	-0.372 (0.033)	1.148	-0.551
S01_LP_Q14	0.580 (0.050)	0.036 (0.029)	0.658	0.093
S01_LP_Q15	0.437 (0.049)	-0.538 (0.029)	0.495	-1.849
S01_LP_Q18	0.882 (0.055)	-0.142 (0.031)	1.000	-0.242
S01_LP_Q21	0.801 (0.055)	0.317 (0.031)	0.908	0.593
S01_LP_Q22	0.974 (0.059)	0.126 (0.032)	1.104	0.194
S01_LP_Q24	0.484 (0.049)	0.299 (0.029)	0.548	0.927
S01_LP_Q26	0.630 (0.055)	0.488 (0.031)	0.714	1.163
S01_LP_Q27	1.073 (0.061)	-0.530 (0.035)	1.217	-0.741
S01_LP_Q28	1.104 (0.064)	0.018 (0.033)	1.251	0.024
CMM_LP_01	0.785 (0.040)	-0.035 (0.025)	0.890	-0.067
CMM_LP_02	1.141 (0.048)	-0.818 (0.031)	1.293	-1.076
CMM_LP_03	1.800 (0.068)	-0.441 (0.039)	2.040	-0.368
CMM_LP_04	0.771 (0.039)	0.169 (0.024)	0.873	0.329
CMM_LP_05	0.684 (0.036)	-0.020 (0.023)	0.775	-0.043
λ – carga fatorial τ – valor limiar de corte (threshold) a – parâmetro de discriminação do item b – parâmetro de dificuldade do item ep – erro padrão				

Tabela 6.27 Parâmetros dos itens do teste S04_LP. Modelo multidimensional longitudinal MEE.

Itens	λ_{it}	τ_{it}	a_{it}	b_{it}
S04_LP_Q01	1.223 (0.068)	-1.147 (0.047)	1.386	-1.408
S04_LP_Q02	0.490 (0.049)	-0.163 (0.031)	0.556	-0.499
S04_LP_Q03	1.731 (0.090)	-0.355 (0.054)	1.962	-0.308
S04_LP_Q04	1.122 (0.066)	0.065 (0.041)	1.271	0.088
S04_LP_Q05	0.790 (0.065)	0.751 (0.042)	0.896	1.426
S04_LP_Q06	1.483 (0.078)	-0.333 (0.048)	1.681	-0.337
S04_LP_Q07	0.995 (0.060)	0.038 (0.039)	1.128	0.057
S04_LP_Q08	1.134 (0.067)	0.264 (0.043)	1.286	0.349
S04_LP_Q09	0.483 (0.050)	0.520 (0.033)	0.547	1.617
S04_LP_Q10	1.344 (0.073)	-0.115 (0.045)	1.523	-0.128
S04_LP_Q12	0.536 (0.049)	0.201 (0.032)	0.607	0.564
S04_LP_Q13	0.741 (0.053)	-0.105 (0.033)	0.840	-0.213
S04_LP_Q17	1.068 (0.064)	0.112 (0.040)	1.211	0.157
S04_LP_Q18	0.989 (0.060)	0.247 (0.039)	1.122	0.375
S04_LP_Q19	1.107 (0.064)	0.160 (0.041)	1.255	0.217
S04_LP_Q20	0.766 (0.055)	0.293 (0.035)	0.868	0.575
S04_LP_Q22	0.686 (0.056)	0.656 (0.037)	0.777	1.435
S04_LP_Q27	1.275 (0.069)	-0.582 (0.042)	1.446	-0.685
S04_LP_Q28	1.014 (0.061)	-0.579 (0.038)	1.150	-0.856
S04_LP_Q29	0.841 (0.058)	0.421 (0.038)	0.954	0.752
CMM_LP_01	0.785 (0.040)	-0.035 (0.025)	0.890	-0.067
CMM_LP_02	1.141 (0.048)	-0.818 (0.031)	1.293	-1.076
CMM_LP_03	1.800 (0.068)	-0.441 (0.039)	2.040	-0.368
CMM_LP_04	0.771 (0.039)	0.169 (0.024)	0.873	0.329
CMM_LP_05	0.684 (0.036)	-0.020 (0.023)	0.775	-0.043
λ – carga fatorial τ – valor limiar de corte (threshold) a – parâmetro de discriminação do item b – parâmetro de dificuldade do item ep – erro padrão				

Tabela 6.28 Parâmetros dos itens do teste S01_MT. Modelo multidimensional longitudinal MEE.

Itens	λ_{it}	τ_{it}	a_{it}	b_{it}
S01_MT_Q01	1.281 (0.082)	-0.892 (0.039)	1.197	-1.268
S01_MT_Q02	0.755 (0.061)	0.094 (0.029)	0.705	0.227
S01_MT_Q04	1.508 (0.085)	-0.278 (0.035)	1.409	-0.335
S01_MT_Q06	0.632 (0.059)	0.049 (0.028)	0.591	0.141
S01_MT_Q07	0.864 (0.063)	0.090 (0.030)	0.808	0.190
S01_MT_Q08	0.674 (0.085)	1.051 (0.037)	0.629	2.842
S01_MT_Q09	1.329 (0.095)	-1.359 (0.051)	1.242	-1.862
S01_MT_Q11	1.123 (0.072)	-0.093 (0.031)	1.049	-0.151
S01_MT_Q12	0.832 (0.066)	0.318 (0.030)	0.777	0.695
S01_MT_Q13	1.277 (0.080)	-0.573 (0.035)	1.193	-0.817
S01_MT_Q16	1.155 (0.071)	-0.393 (0.033)	1.080	-0.620
S01_MT_Q18	1.035 (0.079)	0.591 (0.034)	0.967	1.039
S01_MT_Q20	0.781 (0.073)	0.609 (0.032)	0.729	1.421
S01_MT_Q23	0.923 (0.065)	-0.287 (0.030)	0.862	-0.566
S01_MT_Q26	1.347 (0.081)	-0.053 (0.033)	1.259	-0.072
S01_MT_Q27	0.721 (0.082)	0.915 (0.036)	0.674	2.311
S01_MT_Q28	0.663 (0.103)	1.401 (0.046)	0.620	3.846
S01_MT_Q29	0.985 (0.075)	0.602 (0.033)	0.921	1.112
S01_MT_Q30	1.134 (0.085)	-1.267 (0.045)	1.060	-2.033
CMM_MT_01	1.044 (0.066)	-1.160 (0.034)	0.975	-2.024
CMM_MT_02	1.040 (0.052)	0.374 (0.030)	0.972	0.654
CMM_MT_03	0.779 (0.044)	0.163 (0.026)	0.728	0.382
CMM_MT_04	1.117 (0.066)	1.171 (0.042)	1.044	1.908
λ – carga fatorial τ – valor limiar de corte (threshold) a – parâmetro de discriminação do item b – parâmetro de dificuldade do item ep – erro padrão				

Tabela 6.29 Parâmetros dos itens do teste S04_MT. Modelo multidimensional longitudinal MEE.

Itens	λ_{it}	τ_{it}	a_{it}	b_{it}
S04_MT_Q01	0.925 (0.075)	1.301 (0.068)	0.864	2.561
S04_MT_Q03	1.720 (0.104)	1.256 (0.089)	1.607	1.330
S04_MT_Q06	0.430 (0.066)	1.003 (0.054)	0.402	4.249
S04_MT_Q08	0.793 (0.087)	1.549 (0.081)	0.741	3.555
S04_MT_Q09	1.315 (0.081)	0.158 (0.058)	1.229	0.219
S04_MT_Q10	0.627 (0.062)	0.821 (0.052)	0.586	2.386
S04_MT_Q11	0.532 (0.068)	1.137 (0.058)	0.497	3.890
S04_MT_Q12	0.670 (0.060)	0.598 (0.049)	0.626	1.624
S04_MT_Q13	0.885 (0.072)	1.142 (0.064)	0.827	2.348
S04_MT_Q14	0.473 (0.071)	1.291 (0.061)	0.442	4.972
S04_MT_Q16	0.442 (0.056)	0.349 (0.044)	0.413	1.437
S04_MT_Q17	0.814 (0.065)	0.712 (0.054)	0.760	1.594
S04_MT_Q19	1.709 (0.104)	-0.166 (0.065)	1.597	-0.176
S04_MT_Q20	1.213 (0.078)	0.823 (0.065)	1.133	1.235
S04_MT_Q21	1.189 (0.079)	-0.432 (0.054)	1.111	-0.661
S04_MT_Q22	0.996 (0.073)	-0.574 (0.050)	0.931	-1.049
S04_MT_Q23	0.793 (0.061)	0.194 (0.047)	0.741	0.444
S04_MT_Q27	1.664 (0.103)	0.012 (0.067)	1.555	0.013
S04_MT_Q28	1.460 (0.090)	0.379 (0.067)	1.365	0.472
S04_MT_Q29	1.370 (0.082)	0.188 (0.058)	1.280	0.250
S04_MT_Q30	0.998 (0.067)	0.223 (0.052)	0.932	0.407
CMM_MT_01	1.044 (0.066)	-1.160 (0.034)	0.975	-2.024
CMM_MT_02	1.040 (0.052)	0.374 (0.030)	0.972	0.654
CMM_MT_03	0.779 (0.044)	0.163 (0.026)	0.728	0.382
CMM_MT_04	1.117 (0.066)	1.171 (0.042)	1.044	1.908
λ – carga fatorial τ – valor limiar de corte (threshold) a – parâmetro de discriminação do item b – parâmetro de dificuldade do item ep – erro padrão				

6.3.4 Comparação entre os resultados obtidos pela TRI e pela MEE

A Figura 6.18 compara os valores obtidos para o parâmetro de discriminação a_{it} obtidos utilizando os métodos da TRI e MEE. No momento 1, apesar de não haver uma correspondência exata entre os valores obtidos nos dois métodos, os valores dos diferentes itens parecem se distribuir ao redor da reta identidade, o que já não acontece no momento 2 em que se observa uma clara tendência dos métodos MEE de sobre-estimar os valores de a_{it} . A Figura 6.19 compara os valores obtidos para o parâmetro de dificuldade b_{it} utilizando os métodos da TRI e MEE. Aqui a estimação se comportou de maneira diferentes dependendo da disciplina avaliada. Observa-se uma correspondência quase exata entre os valores estimados para os itens de Língua Portuguesa e para os itens de Matemática localizados no intervalo $[-2.5; +2.5]$. Os valores do parâmetro b_{it} dos itens S04_MT_Q8, S04_MT_Q11, S04_MT_Q14 são maiores que 4.5 (Tabela 6.33) e isso parece comprometer a linearidade da conversão. Interessantemente, esses itens haviam passado o critério de seleção de itens que elimina itens com valores de b_{it} maiores que ± 3 (Tabela 6.16) e só passaram a ter esses valores exacerbados após a equalização. Aqui talvez fosse desejável a aplicação de uma segunda filtragem de itens para eliminar essas distorções. Outro ponto que pode explicar essa melhor correspondência entre os itens de Língua Portuguesa do que entre os de Matemática, é o fato de que usou-se cinco itens comuns na provas de Língua Portuguesa e quatro nas provas de Matemática, e essa pequena diferença pode ter prejudicado a qualidade da equalização.

A Figura 6.20 mostra o resultados que avaliam a sensibilidade do processo de estimação dos escores latentes utilizando parâmetros de itens estimados pela TRI e pela MEE. A correlação entre eles é altíssima como era de se esperar, e é interessante notar que a melhora na média das variáveis associadas à Matemática (η_{21} e η_{22}) é próximo do triplo da melhora observadas nas variáveis associadas à Língua Portuguesa (η_{11} e η_{12}) - $\frac{0.970}{0.288} = 3.368$ na Tabela 6.25. Essa melhora relativa é próxima da observada na análise feito com os métodos da TRI - $\frac{1.176}{0.367} = 3.204$ na Tabela 6.18 e Tabela 6.20.

A carga fatorial e o valor limiar dos itens associados a cada variável é mostrado nas colunas λ_{it} e τ_{it} do conjunto de tabelas que iniciam em Tabela 6.30 e terminam na Tabela 6.33. A partir desses valores é possível calcular os parâmetros de discriminação a_{it} e dificuldade b_{it} dos indicadores através da aplicação das Equações (E.14) e (E.15) com $v_{it} = 0$ e $\theta_{it} = 1$ dado que a parametrização RI_CP_FT foi escolhida.

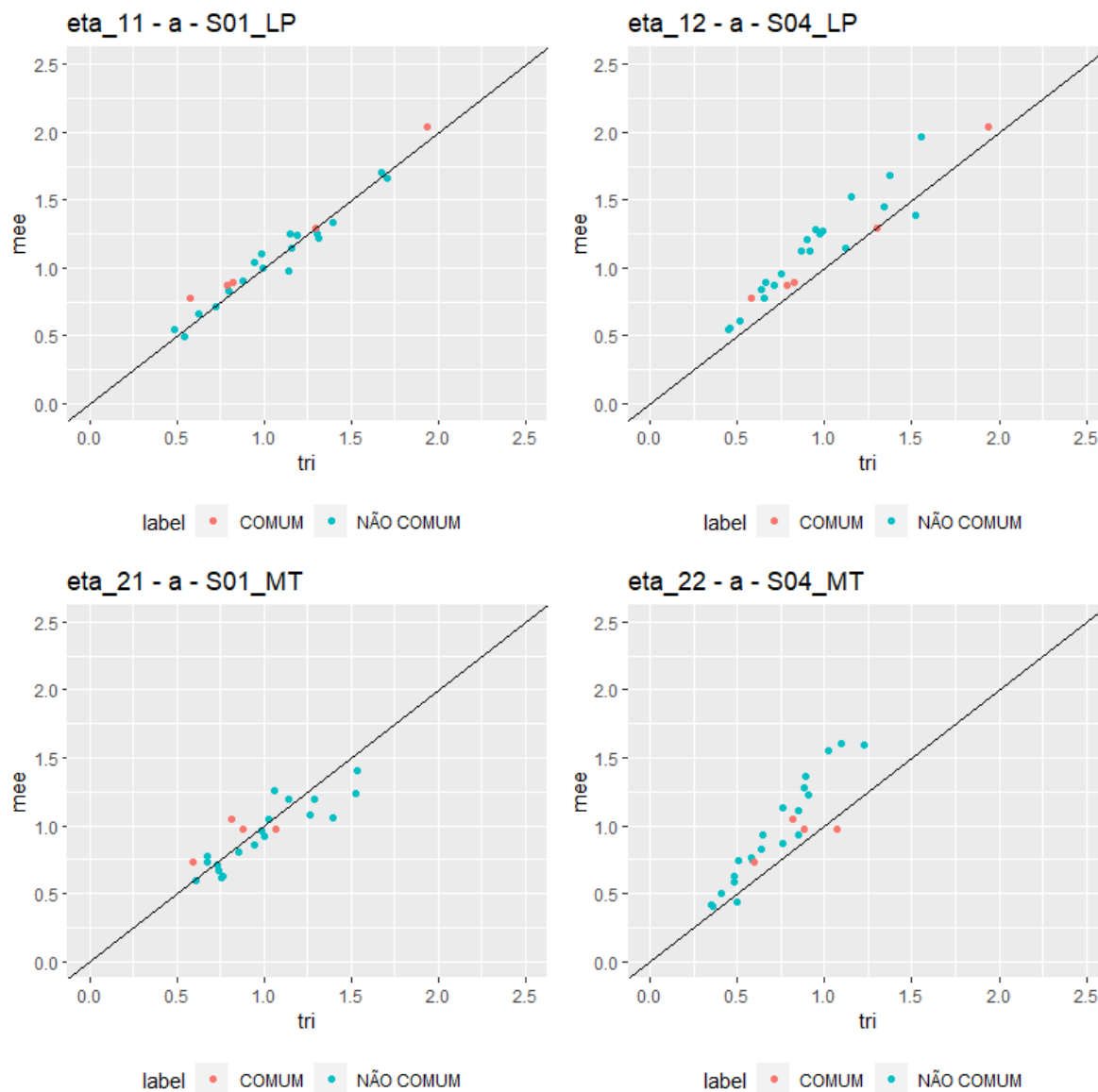


Figura 6.18 Comparação das estimativas do parâmetro de discriminação a_{it} . Modelos longitudinais. TRI vs. MEE

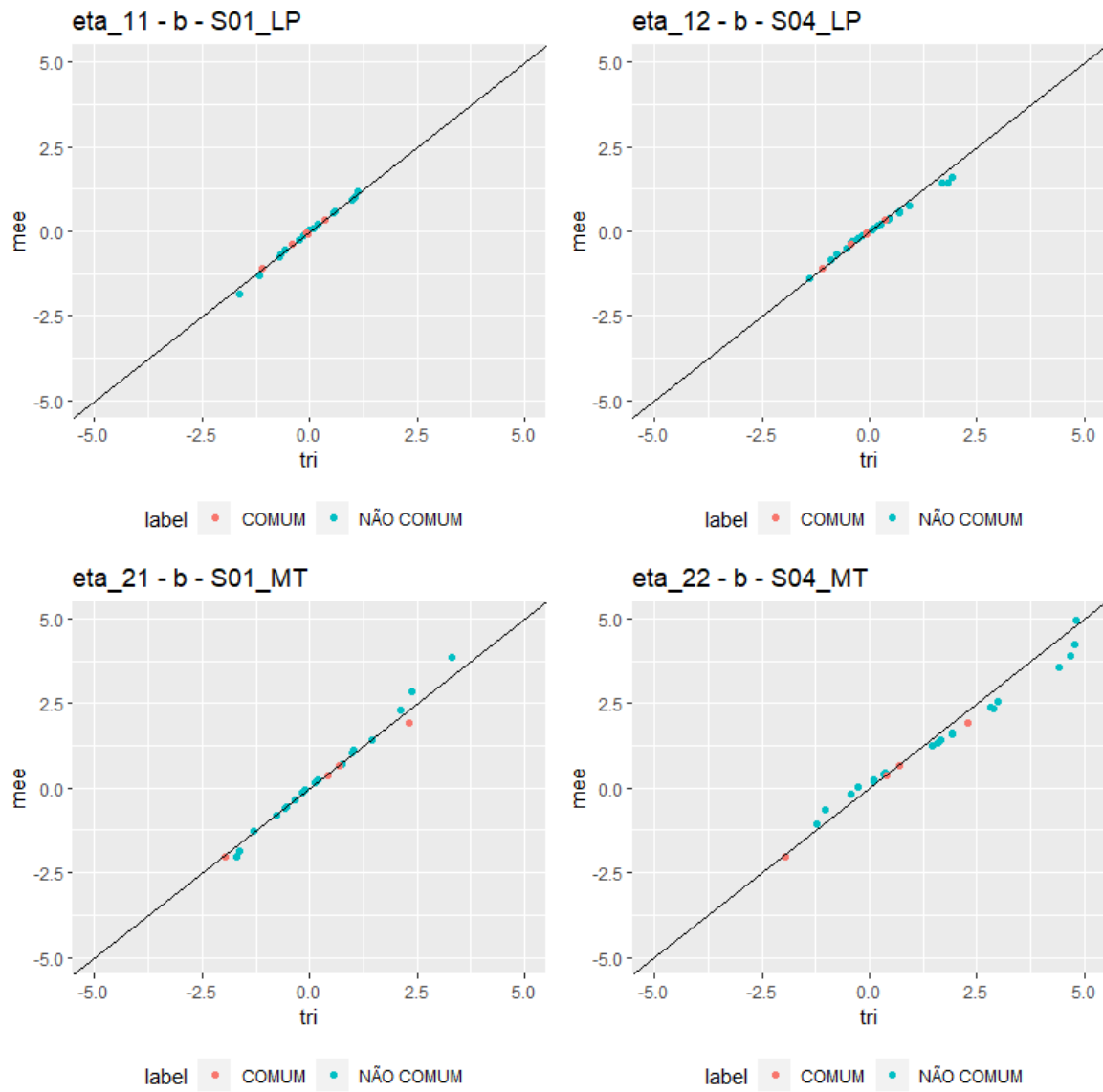


Figura 6.19 Comparação das estimativas do parâmetro de dificuldade b_{it} . Modelos longitudinais. TRI vs. MEE

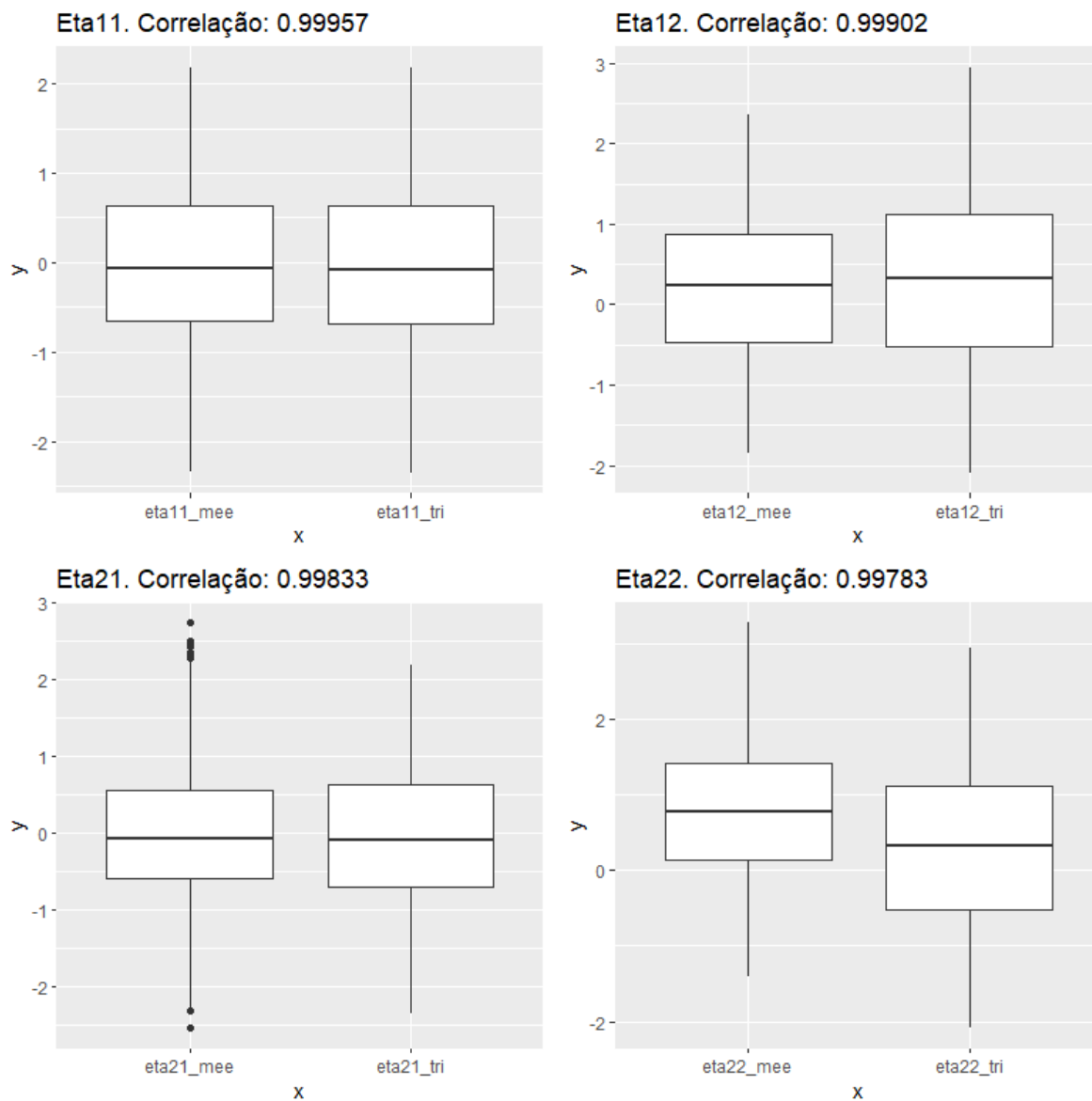


Figura 6.20 Estudo de sensibilidade do processo de estimação dos escores latentes utilizando parâmetros de itens estimados pela TRI e pela MEE. Método de estimação: MAP/EBM.

Tabela 6.30 Comparação das estimativas do parâmetro dos itens do teste S01_LP. Modelos longitudinais. TRI vs. MEE

Itens	a_{it} TRI	a_{it} MEE	$\frac{a_{it}TRI}{a_{it}MEE}$	b_{it} TRI	b_{it} MEE	$b_{it} TRI - b_{it} MEE$
S01_LP_Q01	1.188	1.239	0.958	-0.149	-0.129	-0.052
S01_LP_Q02	0.795	0.827	0.962	-0.661	-0.650	-0.031
S01_LP_Q04	1.139	0.980	1.163	-1.160	-1.294	0.160
S01_LP_Q06	1.677	1.699	0.987	0.551	0.564	-0.022
S01_LP_Q09	1.710	1.660	1.030	0.007	0.037	0.050
S01_LP_Q10	1.394	1.334	1.045	-0.398	-0.388	0.060
S01_LP_Q11	1.303	1.248	1.044	-0.394	-0.387	0.055
S01_LP_Q12	0.948	1.036	0.915	1.048	1.005	-0.088
S01_LP_Q13	1.155	1.148	1.006	-0.558	-0.551	0.007
S01_LP_Q14	0.625	0.658	0.950	0.087	0.093	-0.033
S01_LP_Q15	0.544	0.495	1.100	-1.638	-1.849	0.049
S01_LP_Q18	0.995	1.000	0.995	-0.252	-0.242	-0.005
S01_LP_Q21	0.879	0.908	0.968	0.584	0.593	-0.029
S01_LP_Q22	0.987	1.104	0.894	0.192	0.194	-0.117
S01_LP_Q24	0.487	0.548	0.889	0.983	0.927	-0.061
S01_LP_Q26	0.726	0.714	1.017	1.107	1.163	0.012
S01_LP_Q27	1.315	1.217	1.081	-0.718	-0.741	0.098
S01_LP_Q28	1.146	1.251	0.916	0.007	0.024	-0.105
CMM_LP_01	0.821	0.890	0.922	-0.054	-0.067	-0.069
CMM_LP_02	1.296	1.293	1.002	-1.094	-1.076	0.003
CMM_LP_03	1.934	2.040	0.948	-0.416	-0.368	-0.106
CMM_LP_04	0.786	0.873	0.900	0.371	0.329	-0.087
CMM_LP_05	0.576	0.775	0.743	-0.059	-0.043	-0.199

Tabela 6.31 Comparação das estimativas do parâmetro dos itens do teste S04_LP. Modelos longitudinais. TRI vs. MEE

Itens	a_{it} TRI	a_{it} MEE	$\frac{a_{it}TRI}{a_{it}MEE}$	b_{it} TRI	b_{it} MEE	$b_{it} TRI - b_{it} MEE$
S04_LP_Q01	1.524	1.386	1.099	-1.395	-1.408	0.137
S04_LP_Q02	0.459	0.556	0.825	-0.543	-0.499	-0.097
S04_LP_Q03	1.550	1.962	0.790	-0.395	-0.308	-0.413
S04_LP_Q04	0.989	1.271	0.778	0.107	0.088	-0.282
S04_LP_Q05	0.662	0.896	0.739	1.840	1.426	-0.234
S04_LP_Q06	1.370	1.681	0.815	-0.414	-0.337	-0.312
S04_LP_Q07	0.915	1.128	0.811	0.079	0.057	-0.213
S04_LP_Q08	0.947	1.286	0.737	0.437	0.349	-0.338
S04_LP_Q09	0.448	0.547	0.818	1.929	1.617	-0.100
S04_LP_Q10	1.149	1.523	0.754	-0.176	-0.128	-0.374
S04_LP_Q12	0.514	0.607	0.846	0.681	0.564	-0.093
S04_LP_Q13	0.635	0.840	0.757	-0.264	-0.213	-0.205
S04_LP_Q17	0.898	1.211	0.741	0.191	0.157	-0.313
S04_LP_Q18	0.867	1.122	0.773	0.467	0.375	-0.255
S04_LP_Q19	0.969	1.255	0.772	0.269	0.217	-0.286
S04_LP_Q20	0.710	0.868	0.818	0.704	0.575	-0.158
S04_LP_Q22	0.657	0.777	0.845	1.696	1.435	-0.121
S04_LP_Q27	1.338	1.446	0.925	-0.751	-0.685	-0.108
S04_LP_Q28	1.118	1.150	0.972	-0.879	-0.856	-0.032
S04_LP_Q29	0.752	0.954	0.788	0.931	0.752	-0.202
CMM_LP_01	0.821	0.890	0.922	-0.054	-0.067	-0.069
CMM_LP_02	1.296	1.293	1.002	-1.094	-1.076	0.003
CMM_LP_03	1.934	2.040	0.948	-0.416	-0.368	-0.106
CMM_LP_04	0.786	0.873	0.900	0.371	0.329	-0.087
CMM_LP_05	0.576	0.775	0.743	-0.059	-0.043	-0.199

Tabela 6.32 Comparação das estimativas do parâmetro dos itens do teste S01_MT. Modelos longitudinais. TRI vs. MEE

Itens	a_{it} TRI	a_{it} MEE	$\frac{a_{it}TRI}{a_{it}MEE}$	b_{it} TRI	b_{it} MEE	$b_{it}TRI - b_{it}MEE$
S01_MT_Q01	1.138	1.197	0.951	-1.304	-1.268	-0.059
S01_MT_Q02	0.735	0.705	1.041	0.202	0.227	0.029
S01_MT_Q04	1.531	1.409	1.086	-0.338	-0.335	0.122
S01_MT_Q06	0.607	0.591	1.027	0.124	0.141	0.016
S01_MT_Q07	0.854	0.808	1.057	0.164	0.190	0.046
S01_MT_Q08	0.762	0.629	1.211	2.394	2.842	0.133
S01_MT_Q09	1.528	1.242	1.231	-1.643	-1.862	0.286
S01_MT_Q11	1.027	1.049	0.979	-0.164	-0.151	-0.022
S01_MT_Q12	0.679	0.777	0.873	0.743	0.695	-0.099
S01_MT_Q13	1.286	1.193	1.078	-0.781	-0.817	0.093
S01_MT_Q16	1.266	1.080	1.173	-0.563	-0.620	0.187
S01_MT_Q18	0.983	0.967	1.017	0.996	1.039	0.016
S01_MT_Q20	0.678	0.729	0.930	1.461	1.421	-0.051
S01_MT_Q23	0.949	0.862	1.100	-0.524	-0.566	0.087
S01_MT_Q26	1.062	1.259	0.844	-0.094	-0.072	-0.196
S01_MT_Q27	0.736	0.674	1.093	2.106	2.311	0.063
S01_MT_Q28	0.757	0.620	1.221	3.292	3.846	0.137
S01_MT_Q29	1.001	0.921	1.087	1.014	1.112	0.080
S01_MT_Q30	1.398	1.060	1.320	-1.693	-2.033	0.339
CMM_MT_01	1.070	0.975	1.097	-1.961	-2.024	0.095
CMM_MT_02	0.880	0.972	0.905	0.703	0.654	-0.092
CMM_MT_03	0.595	0.728	0.818	0.414	0.382	-0.133
CMM_MT_04	0.818	1.044	0.783	2.299	1.908	-0.226

Tabela 6.33 Comparação das estimativas do parâmetro dos itens do teste S04_MT. Modelos longitudinais. TRI vs. MEE

Itens	a_{it} TRI	a_{it} MEE	$\frac{a_{it} TRI}{a_{it} MEE}$	b_{it} TRI	b_{it} MEE	$b_{it} TRI - b_{it} MEE$
S04_MT_Q01	0.762	0.864	0.881	2.990	2.561	-0.103
S04_MT_Q03	1.093	1.607	0.680	1.590	1.330	-0.514
S04_MT_Q06	0.355	0.402	0.884	4.778	4.249	-0.046
S04_MT_Q08	0.589	0.741	0.794	4.416	3.555	-0.152
S04_MT_Q09	0.910	1.229	0.741	0.093	0.219	-0.319
S04_MT_Q10	0.478	0.586	0.816	2.819	2.386	-0.108
S04_MT_Q11	0.406	0.497	0.816	4.659	3.890	-0.092
S04_MT_Q12	0.484	0.626	0.773	1.935	1.624	-0.142
S04_MT_Q13	0.635	0.827	0.767	2.892	2.348	-0.193
S04_MT_Q14	0.499	0.442	1.129	4.817	4.972	0.057
S04_MT_Q16	0.349	0.413	0.845	1.656	1.437	-0.064
S04_MT_Q17	0.576	0.760	0.758	1.912	1.594	-0.184
S04_MT_Q19	1.227	1.597	0.768	-0.415	-0.176	-0.370
S04_MT_Q20	0.762	1.133	0.673	1.470	1.235	-0.371
S04_MT_Q21	0.847	1.111	0.763	-1.033	-0.661	-0.264
S04_MT_Q22	0.851	0.931	0.914	-1.215	-1.049	-0.080
S04_MT_Q23	0.506	0.741	0.683	0.374	0.444	-0.235
S04_MT_Q27	1.018	1.555	0.655	-0.280	0.013	-0.537
S04_MT_Q28	0.892	1.365	0.654	0.372	0.472	-0.473
S04_MT_Q29	0.882	1.280	0.689	0.088	0.250	-0.398
S04_MT_Q30	0.648	0.932	0.695	0.324	0.407	-0.285
CMM_MT_01	1.070	0.975	1.097	-1.961	-2.024	0.095
CMM_MT_02	0.880	0.972	0.905	0.703	0.654	-0.092
CMM_MT_03	0.595	0.728	0.818	0.414	0.382	-0.133
CMM_MT_04	0.818	1.044	0.783	2.299	1.908	-0.226

Capítulo 7 Comentários e pesquisas futuras

Measuring progress is often like watching grass grow.
While it's difficult to detect movement on a daily
basis, it's simple to see growth over time.

Frank Sonnenberg

Neste capítulo são feitas algumas considerações finais e sugestões de pesquisas futuras relacionadas ao desenvolvimento do presente trabalho.

7.1 Considerações finais

A literatura em equivalência da análise longitudinal de dados multidimensionais através da TRI e da MEE, não deixa claro que essas duas abordagens podem ser formuladas como alternativas de resolver o mesmo problema com métodos diferentes e complementares. Conforme mencionado neste trabalho, a MEE é tradicionalmente utilizada na construção simultânea de múltiplas variáveis latentes e no estudo da estrutura latente que melhor explica como elas interagem. Neste contexto, a MEE tem uma literatura mais vasta sobre testes de ajuste locais e globais, condições de identificabilidade e invariância fatorial que a da TRI. Já a TRI explora em mais detalhe como a interação entre as respostas aos itens e o traço latente associado pode resultar em um processo de mensuração mais preciso. Isso, pelo menos intuitivamente, justifica os diferentes modelos de curvas sigmóides propostos que possibilitam uma conexão flexível e mais precisa entre as respostas aos itens e o traço latente, com uma melhor exploração das informações contidas nos dados.

As simulações realizadas no Capítulo 5 abordaram modelos transversais e longitudinais. Ao tentar confirmar resultados anteriores em modelos transversais sobre a equivalência entre o modelo do fator comum e modelos lineares generalizados com funções de ligação probit e logit [31][33], este trabalho contribuiu na expansão da literatura corrente em duas frentes. Primeiro, na elucidação do jargão utilizado na MEE em que frequentemente diferentes termos se referem ao mesmo conceito. Isso pode gerar confusão ou ambiguidade na comunicação entre pesquisadores. Por exemplo, uma das alternativas para se fixar a escala do fator comum é fixar sua variância na unidade, sendo essa parametrização chamada de *Unit Variance Identification* (UVI) por Kline [39], de *Standardized Factor* por Kamata e Bauer [36] e de *fixed factor method* por Little [29]. Um outro exemplo seria como os diferentes autores se referem à invariância fatorial relacionada à carga dos fatores em diferentes grupos: *Metric Invariance* por Steenkamp e Baumgartner [83], e *Loading Invariance* por Little [29] e *Weak Invariance* por Kline [39]. Um iniciante no ramo certamente se beneficiará destes esclarecimentos. Segundo, na generalização da fórmula proposta Kamata e Bauer em 2008 [36] para aumentar o número de parametrizações suportadas por ela de 4 para 18. Na MEE, a escolha da parametrização é bastante flexível, leva a diferentes estimativas de parâmetros e afeta a interpretabilidade dos resultados. Este trabalho identificou e sistematizou 18 diferentes maneiras de se parametrizar um modelo unidimensional e como a fórmula em seu formato original só suporta quatro, sendo que, um usuário não cuidadoso pode facilmente aplicar a fórmula num

contexto não suportado. A generalização da fórmula para suportar as diferentes parametrizações identificadas minimiza a probabilidade desse cenário acontecer. Além disso, ilustramos o uso e implementação computacional destas soluções sob diferentes cenários de simulação de dados.

Os estudos de simulação considerados neste trabalho mostraram que a estimação dos parâmetros de dificuldade b_{it} dos itens foi bastante precisa na parte central da escala em que os valores simulados e estimados são praticamente idênticos. Nos extremos da escala, a discrepância aumentava devido à menor quantidade de dados disponíveis para viabilizar uma estimação mais precisa - TABLE D.7 do Apêndice D. Quanto à estimação dos parâmetros de discriminação a_{it} , os desvios ficaram na segunda casa decimal – Ver TABLE D.6 do Apêndice D. Essas estimativas foram obtidas a partir de modelos que utilizaram as 18 parametrizações mencionadas no parágrafo anterior e as mesmas tendências foram observadas, independentemente da parametrização utilizada, o que demonstra a solidez da generalização da fórmula.

Nas simulações abordando modelos longitudinais, a estrutura latente proposta foi modelada na TRI como a combinação de dois modelos unidimensionais longitudinais, e na MEE com um único modelo multidimensional longitudinal. Com a TRI, os parâmetros dos itens foram recuperados com precisão, mas a estrutura latente não. Com a MEE, a estimação dos parâmetros de dificuldade (b_{it}) dos itens foi satisfatória, com desvios maiores observados na estimação dos parâmetros de discriminação (a_{it}). Apesar de DeMars [92, p. 34] mencionar que a estimação de b_{it} quando comparada ao da de a_{it} converge mais rapidamente e requer tamanhos de amostras menores, isso não justifica as discrepâncias observadas aqui dado que as simulações assumiram amostras com 10 000 indivíduos. Existem algumas hipóteses que podem explicar essas diferenças. Primeiro, a fórmula utilizada na obtenção de a_{it} foi derivada num contexto unidimensional transversal e aplicada num contexto multidimensional. Em um contexto multidimensional, não é incoerente assumir que a resposta a um item além de ser influenciada pela variável latente diretamente associada, é também suscetível aos efeitos indiretos de outras variáveis latentes que existam no modelo. Assim, a fórmula utilizada pode estar falhando na consideração desses efeitos indiretos. Uma segunda possibilidade pode estar no método de estimação utilizado, enquanto na TRI procura-se os valores de a_{it} que maximizam a verossimilhança de um padrão de respostas acontecer (método EM), na MEE, procura-se por parâmetros que minimizem a diferença entre Σ e \mathbf{S} . Ou seja, a estimação dos valores dos parâmetros tem por objetivo a reconstrução da matriz de covariância observada e não uma melhor representação da interação resposta/traço latente, apesar de haver uma sobreposição considerável dos efeitos desses diferentes tipos de modelagem. No que se refere à recuperação da estrutura latente, esta foi recuperada com bastante precisão na

modelagem MEE com estimação direta dos parâmetros populacionais procurados. O mesmo não aconteceu na modelagem TRI, já que a covariância entre as variáveis latentes não pode ser estimada diretamente, e seu cálculo foi feito a partir dos escores latentes individuais.

Uma contribuição deste trabalho foi mostrar que na MEE é possível efetuar a calibração simultânea de itens do momento 2, na escala de itens do momento 1 utilizando itens comuns entre esses momentos em um cenário com um grupo, permitindo que as covariâncias intra- e entre-momentos existentes entre os traços latentes influenciem o processo de estimação dos parâmetros. Assim, fica estabelecido que a equalização de itens também é possível utilizando métodos da MEE, o que permite estudos longitudinais que utilizem diferentes, mas comparáveis, instrumentos de mensuração nos diferentes momentos, como é requerido em qualquer avaliação longitudinal educacional de larga escala. É importante ressaltar que na TRI, a calibração simultânea é comumente feita em cenários de grupos múltiplos nos quais se assume que cada grupo é originário de populações independentes.

Outra contribuição foi mostrar a utilidade da equação das médias em permitir a observação do comportamento das médias latentes através dos diferentes momentos em modelos MEE. Combinada com a equação da covariância, a equação das médias pode ser usada para modelar e testar hipóteses sobre as relações entre as variáveis observáveis e latentes em estudos longitudinais, como foi feito nesta tese.

Vale aqui ressaltar como o modelo MEE proposto difere de outra classe de modelos MEE conhecidos como Modelos de Curva de Crescimento (MCC). Os MCC são frequentemente usados no estudo do comportamento de variáveis latentes em estudos longitudinais, mas como as variáveis latentes modeladas são os coeficientes das curvas de tendência, por exemplo, coeficientes de uma equação linear ou quadrática, perde-se a capacidade de estimação dos escores latentes individuais, quase sempre desejáveis em aplicações educacionais.

Outro ponto relevante que merece menção é que, apesar deste trabalho ter focado na análise longitudinal de itens dicotômicos, a análise de longitudinal de itens politômicos é também possível com os mesmos métodos desenvolvidos para os modelos aqui analisados em linha com as análises feitas por Wirth e Edwards [37] para o caso unidimensional transversal.

No Capítulo 6 apresentamos uma aplicação dos métodos pesquisados nesta tese com dados reais coletados no município de Cariacica em 2019. A prova inicial de Português foi avaliada utilizando técnicas da TRI e da MEE para identificar quais itens devem compor o instrumento final, e, independentemente da teoria aplicada, os itens identificados como não satisfatórios foram os mesmos. O mesmo procedimento utilizado no processamento de modelos longitudinais no

capítulo de simulações foi aplicado aqui. Tal como foi observado nas simulações, as estimativas para o parâmetro de dificuldade são bastante próximas (Figura 6.19), enquanto as estimativas para o parâmetro de discriminação divergem levemente (Figura 6.18). Essa observação pode suscitar o questionamento sobre se a escala obtida com os modelos TRI é a mesma obtida com os modelos MEE e que influência isso pode ter na análise dos resultados. Por isso, foi feito um estudo de sensibilidade do processo de estimação dos escores latentes utilizando parâmetros de itens estimados pela TRI e pela MEE e notou-se que a correlação entre os escores individuais é próxima de 1. Em ambos os métodos se observou que a melhora na média das variáveis latentes associadas à Matemática (η_{21} e η_{22}) é próximo do triplo da melhora observada nas variáveis associadas à Língua Portuguesa (η_{11} e η_{12}). Ou seja, em termos gerais, as conclusões obtidas à partir dos dados estudados foram as mesmas independentemente da escolha do método de análise escolhido: TRI ou MEE. Tanto em Língua Portuguesa quanto em Matemática se observou uma significativa melhora na performance dos alunos após seis meses de treinamento direcionado pelas deficiências identificadas na primeira aplicação. Apesar de uma tendência de melhora ser, teoricamente, esperada, a escala dessa melhora pode estar ligeiramente sobrestimada por dois motivos: (a) a amostra de dados contém somente alunos que participaram das quatro aplicações, ou seja, é uma amostra que contém somente os alunos mais assíduos e/ou disciplinados e (b) a existência de itens comuns (cinco em Língua Portuguesa e quatro em Matemática) entre as duas aplicações pode ter colaborado para um aumento na porcentagem de acertos nas provas caso os alunos tenham sido treinados nesses itens.

Assim, mesmo a um custo computacional significativamente maior, a capacidade da MEE de incorporar estruturas latentes complexas na modelagem longitudinal aliado à capacidade de utilizar instrumentos de mensuração distintos e comparáveis em diferentes momentos, como foi proposto nesta tese, faz da MEE uma alternativa interessante para a análise de dados longitudinais em que diferentes momentos sejam avaliados com diferentes instrumentos de mensuração.

Por fim, com o intuito de promover a reprodução dos resultados e aumentar a alcance de audiência, todas as simulações foram feitas utilizando os pacotes `MIRT` e `lavaan` da linguagem R, ambos disponíveis sem custo e no formato *open source*. Scripts de simulação e resultados brutos estão disponíveis sob demanda.

7.2 Sugestões para pesquisas futuras

Alguns aspectos não foram abordados neste presente trabalho e entram no escopo de pesquisas futuras.

7.2.1 Aplicabilidade dos modelos propostos a microdados multidimensionais

As aplicações do ENEM entre os anos 1998 e 2008 contêm itens conhecidamente multidimensionais. Esses exames se estruturavam em 5 competências e 21 habilidades, as quais eram avaliadas com um conjunto de 63 questões. Essas habilidades tinham um caráter abrangente e, segundo o INEP, “algumas das habilidades se apresentam claramente relacionadas a certas competências, outras habilidades podem ser mais próximas de uma ou de outra competência” [94, p. 12]. Deste modo, o estudo da aplicabilidade dos modelos propostos nesta tese para esse tipo de dados é de grande valia.

7.2.2 Estudo de sensibilidade dos resultados apresentados aos parâmetros das simulações

Resultados apresentados nesta tese se limitaram a dois momentos, com cada momento sendo avaliado com 30 itens, com 10 itens comuns entre os momentos e com uma estrutura única de médias e covariâncias. Verificar se os resultados se mantêm sob outras condições de simulação seria de enorme utilidade.

7.2.3 Suporte para dados faltantes sistemáticos em modelos MEE

As avaliações educacionais de larga escala no Brasil frequentemente utilizam o sistema BIB (Blocos Incompletos Balanceados) na construção dos cadernos de respostas aplicados aos alunos [95]. Nessas condições, o sistema BIB permite que diferentes conjuntos de itens sejam distribuídos entre os cadernos, ao mesmo tempo que impõe uma quantidade mínima de itens comuns entre estes para possibilitar suas equalizações. A construção desses cadernos exige cuidado pois é necessário manter a equivalência destes em termos de dificuldade e distribuição temática. Uma consequência da utilização do sistema BIB é que cada respondente é apresentado somente a um subconjunto de itens extraído do conjunto total de itens utilizado na avaliação de todos os respondentes. Apesar da equalização dos itens dos diferentes cadernos puder ser feita de par em par utilizando métodos de calibração de itens *a posteriori* descritos na Seção 3.6.3, se o número de cadernos for grande, a

estimação e equalização simultânea desses cadernos é preferível para minimizar a propagação de erros de equalização. Na TRI, isso é corriqueiramente realizado utilizando a abordagem detalhada na Seção 3.6.2 - Equalização via estimação simultânea, em que os dados referentes aos itens não apresentados (dados faltantes ou *missing data*) são convenientemente ignorados pelo algoritmo de estimação na composição da equação de verossimilhança. Já na MEE, os algoritmos de estimação dos parâmetros não conseguem diretamente absorver dados faltantes sistemáticos que passam a compor sua entrada de dados devido à adoção do sistema BIB. Tem-se aqui um ótimo tema para pesquisa futura, possivelmente, envolvendo efeitos aleatórios específicos na formulação dos MEE que incorporem dados faltantes.

7.2.4 Otimização de algoritmos visando melhor uso de recursos computacionais.

Os processos de estimação da TRI são ordens de magnitude mais rápidas do que os da MEE. O desenvolvimento de software, algoritmos e uso de aceleração de processamento através do uso de GPUs (Graphic Processing Units) é um possível tópico de estudo. Não se pode interromper uma determinada área de pesquisa atual devido à sua natureza computacionalmente intensiva. Embora os computadores atuais possam ser considerados poderosos, ainda há limitações na capacidade de processamento e armazenamento de dados que podem impedir o progresso e aplicação da MEE em avaliações educacionais cotidianas. A TRI mesmo é um dos exemplos de áreas que antes eram computacionalmente inviáveis, mas que entraram para cotidiano de educadores com o desenvolvimento de tecnologias computacionais.

7.2.5 Equivalência entre o modelo de fator comum e o modelo de três parâmetros da TRI

Atualmente, somente está estabelecida a equivalência entre o modelo do fator comum e modelos lineares generalizados com link probit e logit. Dentre todos os modelos propostos na TRI, o modelo de 3 parâmetros, que inclui o parâmetro de acerto casual, é o mais utilizado nas avaliações de larga escala no Brasil. A definição teórica sobre como esse modelo se relaciona com o modelo do fator comum é de grande interesse.

7.2.6 Desenvolvimento de critérios de seleção de itens comuns em cenários multidimensionais.

Um dos pontos mais críticos de qualquer análise longitudinal de dados é garantir que escalas obtidas por meio de diferentes instrumentos de mensuração em diferentes momentos sejam comparáveis. A grande premissa em qualquer método de equalização de escalas é que os itens comuns, entre outros pontos, tenham invariância fatorial, estejam uniformemente distribuídos na escala e tenham parâmetros adequadamente estimados e com valores não extremos. Zhang and Kolen [96] pesquisaram o efeito do número de itens comuns na precisão da equalização e um método para determinar o número mínimo de itens comuns necessários dada uma requerida precisão de equalização para métodos de equalização lineares e encadeados. Michaelides [97] estudou esses efeitos de itens comuns defeituosos no processo de equalização. Estudos similares num contexto multidimensional utilizando métodos de calibração simultânea são de grande interesse prático e teórico.

7.2.7 Estudo sobre *Intensive Longitudinal Data Models*

O estudo de *Intensive Longitudinal Data Models* [98], uma categoria nova de modelos existente entre modelos de séries temporais (uma amostra, muitos momentos) e modelos longitudinais (muitas amostras, poucos momentos) tornou-se possível com o desenvolvimento da tecnologia que permite a coleta automática de dados sobre indivíduos ou grupos em alta frequência por longos períodos (muitas amostras, muitos momentos). A possibilidade de aplicação desses modelos em contextos educacionais parece ter grande potencial prático.

7.2.8 Distribuições assimétricas de traços latentes

Este trabalho assumiu a distribuição normal para os traços latentes, mas essa é uma premissa que nem sempre é válida. Em um estudo longitudinal, a distribuição pode adquirir caudas assimétricas com a inserção e eliminação de indivíduos ao longo do tempo. Ley [99] reporta três diferentes famílias de distribuições assimétricas, dentre as quais a distribuição *skew-normal* foi aplicada por Azevedo, Bolfarine e Andrade [100] e Bazán, Branco, Bolfarine [101] na criação de modelos TRI unidimensionais transversais. Uma revisão do modelo do fator comum que levasse em consideração distribuições assimétricas e seu relacionamento com modelos TRI que abordam essa assimetria seria de grande valia.

Referências Bibliográficas

- [1] F. de J. Moreira Junior, “Aplicações da Teoria da Resposta ao Item (TRI) no Brasil,” *Revista Brasileira de Biometria*, vol. 28, no. 4, 2010.
- [2] C. Spearman, “Correlations of sums and differences,” *British Journal of Psychology*, vol. 5, pp. 417–426, 1913.
- [3] C. Spearman, “Demonstration of formulae for true measurement of correlation,” *American Journal of Psychology*, vol. 18, pp. 161–169, 1907.
- [4] C. Spearman, “General Intelligence objectively determined and measured.,” *American Journal of Psychology*, vol. 15, pp. 201–293, 1904.
- [5] A. Binet and T. Simon, “Le developpement de l’intelligence chez les enfants,” *Annee psychologique*, vol. 14, pp. 1–94, 1908.
- [6] A. Binet and T. Simon, “The development of intelligence in young children.” 1916.
- [7] L. M. Terman, *The measurement of intelligence*. Boston: Houghton Mifflin, 1916.
- [8] D. N. Lawley, “The Application of the Maximum Likelihood Method to Factor Analysis,” *British Journal of Psychology. General Section*, vol. 33, no. 3, 1943, doi: 10.1111/j.2044-8295.1943.tb01052.x.
- [9] F. Lord, “A theory of test scores.,” *Psychometric monographs*, 1952.
- [10] R. K. Hambleton and R. W. Jones, “Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development,” Philadelphia, PA, United States of America, 1993.
- [11] E. Andersen and M. Madsen, “Estimating the parameters of the latent population distribution,” *Psychometrika*, vol. 42, no. 3, 1977, doi: 10.1007/BF02293656.
- [12] L. Sanathanan and S. Blumenthal, “The logistic model and estimation of latent structure,” *J Am Stat Assoc*, vol. 73, no. 364, 1978, doi: 10.1080/01621459.1978.10480101.
- [13] J. de Leeuw and N. Verhelst, “Maximum likelihood estimation in generalized Rasch models,” *Journal of Educational and Behavioral Statistics*, vol. 11, no. 3, pp. 183–196, 1986.
- [14] B. Lindsay, C. C. Clogg, and J. Grego, “Semiparametric estimation in the rasch model and related exponential response models, including a simple latent class model for item analysis,” *J Am Stat Assoc*, vol. 86, no. 413, 1991, doi: 10.1080/01621459.1991.10475008.

- [15] R. J. Mislevy, “Estimating latent distributions,” *Psychometrika*, vol. 49, no. 3, 1984, doi: 10.1007/BF02306026.
- [16] R. D. Bock and M. F. Zimowski, “Multiple group IRT,” in *Handbook of Modern Item response Theory*, W. J. der Linden and R. K. Hambleton, Eds., 1 st. New York: Springer-Verlag, 1997.
- [17] M. J. Kolen and R. L. Brennan, *Test Equating, Scaling, and Linking: Methods and Practices*. in Statistics for Social and Behavioral Sciences. Springer New York, 2014. [Online]. Available: <https://books.google.com.br/books?id=AzO5BAAAQBAJ>
- [18] D. F. de Andrade, H. R. Tavares, and R. da C. Valle, “A Teoria da Resposta ao Item: Conceitos e Aplicações,” *4º Sinape*, no. September, 2000.
- [19] J. J. McArdle, K. J. Grimm, F. Hamagami, R. P. Bowles, and W. Meredith, “Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement,” *Psychol Methods*, vol. 14, pp. 126–149, Sep. 2009.
- [20] A. D. F. Tavares H. R., “Item Response Theory for Longitudinal Data: Item Parameter Estimation,” University of São Paulo, São Paulo, Brazil, 2001.
- [21] C. L. N. Azevedo, J. P. Fox, and D. F. Andrade, “Bayesian longitudinal item response modeling with restricted covariance pattern structures,” *Stat Comput*, vol. 26, no. 1–2, pp. 443–460, Jan. 2016, doi: 10.1007/s11222-014-9518-5.
- [22] M. D. Reckase, *Multidimensional Item Response Theory*, 1st ed. Springer Publishing Company, Incorporated, 2009.
- [23] J. B. Simpson, “A model for testing with multidimensional items,” in *Proceedings of the 1977 Computerized Adaptive Testing Conference*, Minneapolis, 1978.
- [24] D. M. Bolt and V. F. Lall, “Estimation of Compensatory and Noncompensatory Multidimensional Item Response Models Using Markov Chain Monte Carlo,” *Appl Psychol Meas*, vol. 27, no. 6, 2003, doi: 10.1177/0146621603258350.
- [25] L. Pasquali, *Psicometria: Teoria dos testes na Psicologia e na Educação*, 5th ed. Editora Vozes, 2003.
- [26] R. J. de Ayala, *The Theory and Practice of Item Response Theory*. in Methodology in the Social Sciences. Guilford Publications, 2013. [Online]. Available: <https://books.google.com.br/books?id=NchXAQAAQBAJ>
- [27] N. N. Vieira, “As Provas das Quatro Áreas do Enem Vistas como Prova Única na Ótica de Modelos da Teoria da Resposta ao Item Uni e Multidimensional,” Universidade Federal de Santa Catarina, 2016.

- [28] C. L. N. Azevedo, “Modelos longitudinais de grupos múltiplos multiníveis na teoria da resposta ao item: métodos de estimação e seleção estrutural sob uma perspectiva bayesiana,” Universidade de São Paulo, São Paulo, 2008.
- [29] Todd D. Little, *Longitudinal Structural Equation Modeling*. Guilford Press, 2013.
- [30] J. T. Newsom, *Longitudinal Structural Equation Modeling*. 2015. doi: 10.4324/9781315871318.
- [31] Y. Takane and J. de Leeuw, “On the relationship between item response theory and factor analysis of discretized variables,” *Psychometrika*, vol. 52, no. 3, pp. 393–408, 1987.
- [32] B. Muthén, “A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators,” *Psychometrika*, vol. 49, pp. 115–132, 1984.
- [33] B. Muthén, “Latent variable structural equation modeling with categorical data,” *J Econom*, vol. 22, 1983.
- [34] B. David John, “Latent variable models for ordered categorical data,” *J Econom*, vol. 22, pp. 229–243, 1983.
- [35] A. Glockner-Rist and H. Hoijtink, “The best of both worlds: Factor analysis of dichotomous data using item response theory and structural equation modeling,” *Structural Equation Modeling*, vol. 10, no. 4, pp. 544–565, 2003.
- [36] A. Kamata and D. J. Bauer, “A Note on the Relation Between Factor Analytic and Item Response Theory Models,” *Structural Equation Modeling*, vol. 15, pp. 136–153, 2008.
- [37] R. J. Wirth and M. C. Edwards, “Item factor analysis: current approaches and future directions,” *Psychol Methods*, vol. 12, no. 1, pp. 58–79, Sep. 2007.
- [38] K. A. Bollen, “Latent variables in psychology and the social sciences,” *Annu Rev Psychol*, vol. 53, no. 1, pp. 605–634, 2002.
- [39] R. B. Kline, “Principles and Practice of Structural Equation Modeling, Fourth Edition,” *The Guilford Press*, vol. 8, no. 5, 2016.
- [40] J. Loehinger, “Objective Tests as Instruments of Psychological Theory,” *Psychol Rep*, vol. 3, no. 3, 1957, doi: 10.2466/pr0.1957.3.3.635.
- [41] D. Bartholomew, M. Knott, and I. Moustaki, *Latent Variable Models and Factor Analysis: A Unified Approach: 3rd Edition*. 2011. doi: 10.1002/9781119970583.
- [42] A. Skrondal and S. Rabe-Hesketh, *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. 2004. doi: 10.1198/tech.2005.s263.

- [43] F. M. Lord and M. R. Novick, *Statistical theories of mental test scores*. in Addison-Wesley series in behavioral science. Addison-Wesley Pub. Co., 1968. [Online]. Available: <https://books.google.com.br/books?id=FW5EAAAIAAJ>
- [44] P. M. Bentler and D. G. Weeks, “Linear structural equations with latent variables,” *Psychometrika*, vol. 45, no. 3, pp. 289–308, 1980, doi: 10.1007/BF02293905.
- [45] K. A. Bollen, *Structural Equations with Latent Variables*. Wiley, 1998.
- [46] R. P. McDonald, *Test theory: A unified treatment*. 2013. doi: 10.4324/9781410601087.
- [47] M. R. Sanches, “Indicadores formativos em modelos de equações estruturais,” Master Thesis, Universidade de São Paulo, São Paulo, 2005.
- [48] F. B. Baker and S.-H. Kim, *Item Response Theory: Parameter Estimation Techniques, Second Edition*. in Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 2004. [Online]. Available: https://books.google.com.br/books?id=y-Q_Q7pasJ0C
- [49] F. B. Baker, *The Basics of Item Response Theory*, Second Edition. ERIC Clearinghouse on Assessment and Evaluation, 2001.
- [50] A. Birnbaum, “Some Latent Trait Models and Their Use in Inferring an Examinee’s Ability,” in *Statistical Theories of Mental Test Scores*, Reading: Addison-Wesley Pub. Co., 1968, pp. 397–479.
- [51] G. L. Marco, “Item Characteristic Curve Solutions to Three Intractable Testing Problems,” *J Educ Meas*, vol. 14, no. 2, 1977, doi: 10.1111/j.1745-3984.1977.tb00033.x.
- [52] M. A. Barton and F. M. Lord, “An Upper Asymptote for the Three-parameter Logistic Item-response Model,” *ETS Research Report Series*, vol. 1981, no. 1, 1981, doi: 10.1002/j.2333-8504.1981.tb01255.x.
- [53] E. Loken and K. L. Rulison, “Estimation of a 4-parameter Item Response Theory model,” *Br J Math Stat Psychol*, vol. 63, no. 3, pp. 509–525, 2010.
- [54] H. R. Tavares, D. F. de Andrade, and C. A. de Bragança Pereira, “Detection of determinant genes and diagnostic via item response theory,” *Genet Mol Biol*, vol. 27, pp. 679–685, 2004.
- [55] T. C. Guimarães, A. N. Moraes, and H. R. Tavares, “Estimação no Modelo Logístico de 5 Parâmetros,” in *Anais do VI CONBRATRI: Métodos para detecção de fraudes em testes*, Juiz de Fora, MG: ABAVE, Dec. 2018.
- [56] D. C. Haley, “Estimation of the Dosage Mortality Relationship When the Dose is Subject to Error,” Stanford, CA, 1952.

- [57] J. A. Spray, T. C. Davey, M. D. Reckase, T. A. Ackerman, and J. E. Carlson, “Comparison of Two Logistic Multidimensional Item Response Theory Models,” Iowa City, IA, 1990.
- [58] W.-C. Wang and P.-H. Chen, “Implementation and Measurement Efficiency of Multidimensional Computerized Adaptive Testing,” *Appl Psychol Meas*, vol. 28, no. 5, pp. 295–316, 2004, doi: 10.1177/0146621604265938.
- [59] B. G. E. Babcock, “Estimating A Noncompensatory IRT Model Using A Modified Metropolis Algorithm,” UNIVERSITY OF MINNESOTA, 2009.
- [60] D. F. Andrade, H. R. Tavares, and R. C. Valle, “Introdução a Teoria da resposta ao Item: Conceitos e Aplicações. 14o SINAPE: Caxambu.” MG, 2000.
- [61] D. F. de Andrade, H. R. Tavares, and R. da Cunha Valle, “Teoria da Resposta ao Item: conceitos e aplicações,” *ABE, Sao Paulo*, 2000.
- [62] R. J. Mislevy and R. D. Bock, “BILOG 3: Item Analysis and Test Scoring with binary logistic models,” Mooresville, IN, 1989.
- [63] E. B. Andersen, “Asymptotic Properties of Conditional Maximum-Likelihood Estimators,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 33, no. 1, 1971, doi: 10.1111/j.2517-6161.1971.tb00869.x.
- [64] J. Neyman and E. L. Scott, “Consistent Estimates Based on Partially Consistent Observations,” *Econometrica*, vol. 16, no. 1, 1948, doi: 10.2307/1914288.
- [65] D. N. M. de Gruijter, “A comment on ‘some standard errors in item response theory,’” *Psychometrika*, vol. 49, no. 2. 1984. doi: 10.1007/BF02294177.
- [66] R. D. Bock and M. Aitkin, “Marginal maximum likelihood estimation of item parameters: An application of a EM algorithm,” *Psychometrika*, vol. 46, pp. 433–459, 1981.
- [67] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*, 4th ed. Pearson, 2012.
- [68] J. M. de Andrade, J. A. Laros, and V. V. Gouveia, “O uso da teoria de Resposta ao Item em avaliações educacionais: diretrizes para pesquisadores,” *Avaliação Psicológica*, vol. 9, no. 3, 2010.
- [69] B. H. Loyd and H. D. Hoover, “Vertical Equating Using the Rasch Model,” *J Educ Meas*, vol. 17, no. 3, 1980, doi: 10.1111/j.1745-3984.1980.tb00825.x.
- [70] S. L. Hershberger, “The Growth of Structural Equation Modeling: 1994-2001,” *Structural Equation Modeling*, vol. 10, no. 1, pp. 35–46, 2003.

- [71] R. Pilati and J. A. Laros, “Modelos de equações estruturais em psicologia: conceitos e aplicações,” *Psicologia: teoria e pesquisa*, vol. 23, no. 2, pp. 205–216, 2007.
- [72] M. W. Browne, “Asymptotically distribution-free methods for the analysis of covariance structures,” *British Journal of Mathematical and Statistical Psychology*, vol. 37, no. 1, pp. 62–83, May 1984, doi: 10.1111/j.2044-8317.1984.tb00789.x.
- [73] K. G. Jöreskog and I. Moustaki, “Factor Analysis of Ordinal Variables: A Comparison of Three Approaches,” *Journal Multivariate Behavioral Research*, vol. 36, no. 3, pp. 347–387, Sep. 2001.
- [74] C. Varin, “On composite marginal likelihoods,” *Advances in Statistical Analysis*, vol. 98, no. 1, pp. 1–28, 2008.
- [75] R. E. Schumacker and R. G. Lomax, *A Beginner’s Guide to Structural Equation Modeling*. 2004. doi: 10.4324/9781410610904.
- [76] J. H. Steiger and J. C. Lind, “Statistically based tests for the number of common factors.,” in *The Annual Meeting of the Psychometric Society*, 1980.
- [77] L. R. Tucker and C. Lewis, “A reliability coefficient for maximum likelihood factor analysis,” *Psychometrika*, vol. 38, no. 1, pp. 1–10, Mar. 1973, doi: 10.1007/BF02291170.
- [78] P. M. Bentler and D. G. Bonett, “Significance tests and goodness of fit in the analysis of covariance structures.,” *Psychol Bull*, vol. 88, no. 3, pp. 588–606, Nov. 1980, doi: 10.1037/0033-2909.88.3.588.
- [79] B. G. Whitaker and J. L. McKinney, “Assessing the measurement invariance of latent job satisfaction ratings across survey administration modes for respondent subgroups: A MIMIC modeling approach,” *Behav Res Methods*, vol. 39, no. 3, pp. 502–509, Aug. 2007, doi: 10.3758/BF03193019.
- [80] H. Goldstein, “Measuring Changes in Educational Attainment over Time: Problems And Possibilities,” *J Educ Meas*, vol. 20, no. 4, pp. 369–377, Dec. 1983, doi: 10.1111/j.1745-3984.1983.tb00214.x.
- [81] C. S. Wells, M. J. Subkoviak, and R. C. Serlin, “The Effect of Item Parameter Drift on Examinee Ability Estimates,” *Appl Psychol Meas*, vol. 26, no. 1, pp. 77–87, Mar. 2002, doi: 10.1177/0146621602261005.
- [82] W. Meredith, “Measurement invariance, factor analysis and factorial invariance,” *Psychometrika*, vol. 58, no. 4, pp. 525–543, Dec. 1993, doi: 10.1007/BF02294825.

- [83] J. E. M. Steenkamp and H. Baumgartner, “Assessing Measurement Invariance in Cross-National Consumer Research,” *Journal of Consumer Research*, vol. 25, no. 1, pp. 78–107, Jun. 1998, doi: 10.1086/209528.
- [84] C. D. Desjardins and O. Bulut, *Handbook of Educational Measurement and Psychometrics Using R*, 1st Edition. CRC Press, 2018.
- [85] BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), “Matriz de Referência de Língua Portuguesa: SAEB – 2001,” Brasília, Apr. 2022. Accessed: Nov. 28, 2022. [Online]. Available: https://download.inep.gov.br/educacao_basica/saeb/matriz-de-referencia-de-lingua-portuguesa_2001.pdf
- [86] BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), “Matriz de Referência de Matemática: SAEB - 2001,” Brasília, Apr. 2022. Accessed: Nov. 28, 2022. [Online]. Available: https://download.inep.gov.br/educacao_basica/saeb/matriz-de-referencia-de-matematica_2001.pdf
- [87] K. S. Taber, “The Use of Cronbach’s Alpha When Developing and Reporting Research Instruments in Science Education,” *Res Sci Educ*, vol. 48, no. 6, pp. 1273–1296, Dec. 2018, doi: 10.1007/s11165-016-9602-2.
- [88] R. P. Chalmers, “mirt: A Multidimensional Item Response Theory Package for the R Environment,” *J Stat Softw*, vol. 48, no. 6, pp. 1–29, 2012, doi: 10.18637/jss.v048.i06.
- [89] T. M. Haladyna and M. C. Rodriguez, *Developing and validating test items*. New York, NY: Routledge, 2013.
- [90] T. M. Haladyna, *Developing and Validating Multiple-choice Test Items*. Taylor & Francis, 2012. [Online]. Available: <https://books.google.com.br/books?id=twJsqoAFsNcC>
- [91] T. M. Haladyna, *Writing Test Items to Evaluate Higher Order Thinking*. Allyn and Bacon, 1997. [Online]. Available: <https://books.google.com.br/books?id=C0sOAQAAMAAJ>
- [92] C. DeMars, *Item Response Theory*. in Understanding Statistics. Oxford University Press, 2010. [Online]. Available: <https://books.google.com.br/books?id=KOADeYBt7sIC>
- [93] Y. Rosseel, “lavaan: An R Package for Structural Equation Modeling,” *J Stat Softw*, vol. 48, no. 2, pp. 1–36, 2012, [Online]. Available: <http://www.jstatsoft.org/v48/i02/>

- [94] “EXAME NACIONAL DO ENSINO MÉDIO - ENEM, ESCALAS DE PROFICIÊNCIA 1998/2008,” Nov. 2018.
- [95] R. M. Bekman, “Aplicação dos blocos incompletos balanceados na Teoria de Resposta ao Item,” *Estudos em Avaliação Educacional*, no. 24, p. 119, Dec. 2001, doi: 10.18222/ae02420012202.
- [96] M. Zhang and M. J. Kolen, “Effects of the Number of Common Items on Equating Precision and Estimates of the Lower Bound to the Number of Common Items Needed,” Aug. 2013.
- [97] M. Michaelides, “A review of the effects on IRT item parameter estimates with a focus on misbehaving common items in test equating,” *Front Psychol*, vol. 1, 2010, doi: 10.3389/fpsyg.2010.00167.
- [98] T. A. Walls and J. L. Schafer, Eds., *Models for Intensive Longitudinal Data*, 1st ed. 2006.
- [99] C. Ley, “Flexible modelling in statistics: past, present and future,” *Journal de la société française de statistique*, vol. 156, pp. 76–96, 2015.
- [100] C. L. N. Azevedo, H. Bolfarine, and D. F. Andrade, “Bayesian inference for a skew-normal IRT model under the centred parameterization,” *Computational Statistics Data Analysis*, vol. 55, no. 1, pp. 353–365, Jan. 2011.
- [101] J. L. Bazán, M. D. Branco, and H. Bolfarine, “Extensions of the skew-normal ogive item response model,” *Braz J Probab Stat*, vol. 28, pp. 1–23, 2014.

Apêndice A Simulações

A.1 Simulações Unidimensionais Transversais

A.1.1 O Modelo

Em simulações unidimensionais transversais ($K = 1, T = 1$), utilizando o modelo de 2 parâmetros e assumindo que o traço latente dos H indivíduos segue uma distribuição normal $\eta_{11} \sim (\mu_{11}, \sigma_{11}^2)$, o conjunto de equações que definem o modelo é o seguinte.

$$Y_{i1} \mid (\eta_{11}, \zeta_i) \sim \text{Bernoulli}(P_{ih11}) \quad (\text{A.1})$$

$$P_{ih11} = P(Y_{i1} = 1 \mid \eta_{h11}, \zeta_i) = \frac{1}{1 + e^{a_i D(\eta_{h11} - b_i)}} \quad (\text{A.2})$$

$$\zeta_i = (a_i, b_i) \quad (\text{A.3})$$

$$\eta_{11} \mid \boldsymbol{\pi}_{11} \sim N(\mu_{11}, \sigma_{11}^2) \quad (\text{A.4})$$

A.1.2 Geração de dados simulados

Os passos necessários para se gerar as respostas a cada item por cada indivíduo são os seguintes.

1. Gere H amostras η_{h11} de $N(\mu_{11}, \sigma_{11}^2)$.
2. Gere I pares de parâmetros ζ_i . Os valores de a_i são amostras de uma variável uniforme $U[0.6; 1.6]$. Os valores de b_i variaram de -3 a $+3$ em intervalos de 0.5 .
3. Calcule P_{ih11} para cada item e indivíduo conforme a Equação (A.2).
4. Utilizando o método da transformada inversa gera-se os valores y_{ih1} para todos os I itens e H indivíduos segundo abaixo.
 - a. Gera-se $U \sim U(0,1)$, então
 - b. Se definirmos $Y_{i1} = 1_{U \leq P_{ih11}}$, então
 - i. $P(Y_{i1} = 1) = P(U \leq P_{ih11}) = P_{ih11}$
 - ii. $P(Y_{i1} = 0) = P(U > P_{ih11}) = 1 - P_{ih11}$

O código R utilizado na geração destes dados é o seguinte:

```
I = 30
N = 2000
a = runif(I,0.6, 1.65)      # U(0.6 , 1.6)
b = runif(I,-3.0, 3.0)
c = rep(0,I)
d = -a*b # MIRT trabalha com o intercepto (d=-ab)
eta = MASS::mvrnorm(n=N, mu=0, Sigma=1 )
D = 1
eta = eta[,1]%*% t(D*a[1:30]) + matrix(D*d[1:30],N,30,byrow=TRUE);
eta =t(eta) # N x I (a'theta+d)
P = c[1:30] + (1-c[1:30])/(1+exp(-eta)); P=t(P) # n x I
X = runif(N*30); dim(X)=c(N,30) # matriz n x I de U(0,1)
U1 = 1*(X<P) ; U1=as.data.frame(U1) # AQUI TEMOS OS DADOS DICOTÔMICOS
colnames(U1)=paste0("Item.",1:30,".t1")
```

ou, alternativamente, a função `simdata` do pacote `MIRT` também pode ser utilizada.

```
I=30
N = 2000
a = runif(I,0.6, 1.65)      # U(0.6 , 1.6)
b = runif(I,-3.0, 3.0)
c = rep(0,I)
d =-a*b # MIRT trabalha com o intercepto (d=-ab)
eta = MASS::mvrnorm(n=N, mu=0, Sigma=1 )
U1 = mirt::simdata(a=a,d=d,N=N,itemtype = '2PL', Theta = eta )
```

A.1.3 Utilizando a TRI

Os dados dicotômicos gerados na seção anterior são ajustados em um modelo unidimensional de dois parâmetros e os parâmetros dos itens são estimados.

```
mod <-mirt(data=U1,itemtype = "2PL",model=1)
par<-coef(mod,simplify=TRUE,IRTpars = TRUE)

ggplot2::ggplot(data = data.frame(SIM = a[1:30], EST = par$items[, "a"]),
  aes(x = SIM, y = EST)) +
  geom_point() +
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Simulações do parâmetro a")

ggplot2::ggplot(data = data.frame(SIM = b[1:30], EST = par$items[, "b"]),
  aes(x = SIM, y = EST)) +
  geom_point() +
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Simulações do parâmetro b")
```

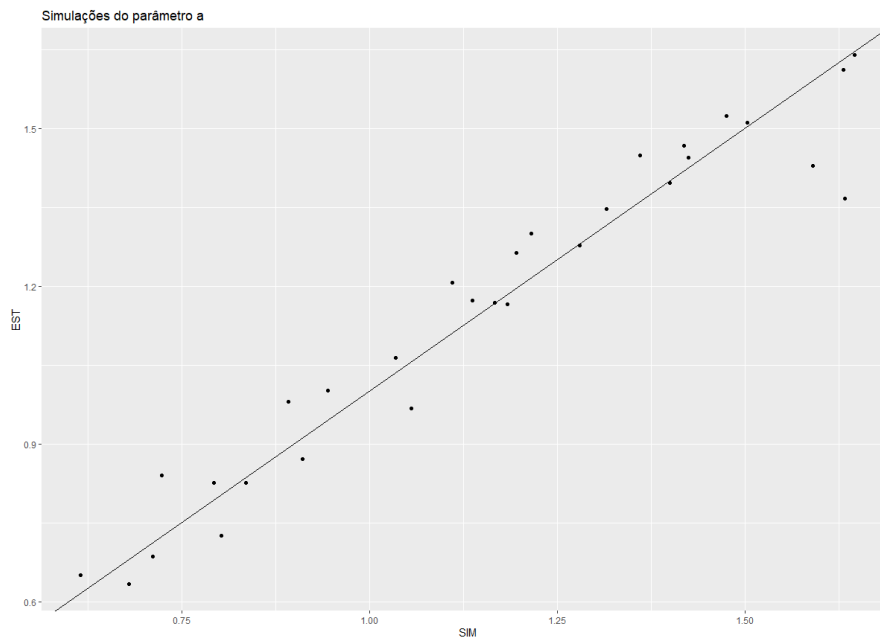



Figura A.1 Recuperação dos valores do parâmetro a utilizando a TRI. ML2P. 30 itens.

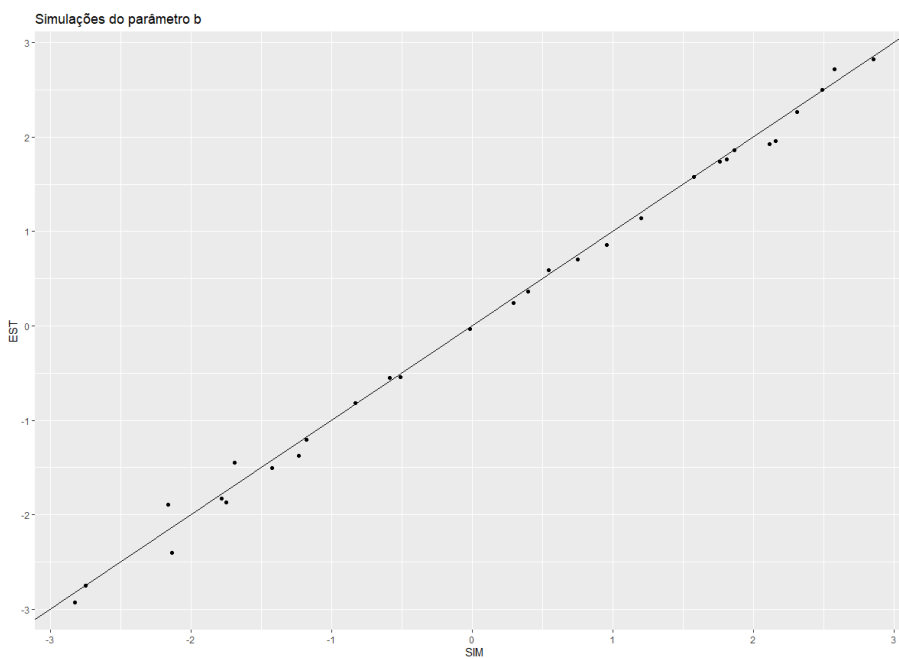


Figura A.2 Recuperação dos valores do parâmetro b utilizando a TRI. ML2P. 30 itens.

A.1.4 Utilizando a MEE

```
get_lmb_tau_alpha_psi <- function( fitted.model){  
  
  lambda <- lavInspect(fitted.model,what = 'est')$lambda  
  colnames(lambda) <- paste0("lambda", 1:ncol(lambda))  
  
  #getting tau values  
  tau <- lavInspect(fitted.model,what = 'est')$tau  
  colnames(tau) <- c("tau")  
  
  alpha <- lavInspect(fitted.model,what = 'mean.lv')  
  psi <- lavInspect(fitted.model,what = 'cov.lv')  
  
  return(list(lambda=lambda,tau=tau,alpha=alpha,psi=psi))  
}  
  
get_a_d_b <- function( tn.sem.param,D=1.701){  
  
  n <- length(tn.sem.param$tau)  
  a <- rep(NA,n)  
  d <- rep(NA,n)  
  
  for(i in seq(1,n,1)){  
    a[i] <- tn.sem.param$lambda[i,1]*sqrt(tn.sem.param$psi[1,1])*D  
    d[i] <- (-tn.sem.param$tau[i] +tn.sem.param$lambda[i,1]*tn.sem.param$alpha[1]) *D  
  }  
  b = -d/a  
  
  return(list(a=a,d=d,b=b))  
}
```

```

lavaan.model.t1.free.tag <- function(){}
lavaan.model.t1.free <-'

d1      =~      lambda1.t1*Item.1.t1+      lambda2.t1*Item.2.t1+      lambda3.t1*Item.3.t1+
lambda4.t1*Item.4.t1+ lambda5.t1*Item.5.t1+ lambda6.t1*Item.6.t1+ lambda7.t1*Item.7.t1+
lambda8.t1*Item.8.t1+ lambda9.t1*Item.9.t1+ lambda10.t1*Item.10.t1+ lambda11.t1*Item.11.t1+
lambda12.t1*Item.12.t1+      lambda13.t1*Item.13.t1+      lambda14.t1*Item.14.t1+
lambda15.t1*Item.15.t1+      lambda16.t1*Item.16.t1+      lambda17.t1*Item.17.t1+
lambda18.t1*Item.18.t1+      lambda19.t1*Item.19.t1+      lambda20.t1*Item.20.t1+
lambda21.t1*Item.21.t1+      lambda22.t1*Item.22.t1+      lambda23.t1*Item.23.t1+
lambda24.t1*Item.24.t1+      lambda25.t1*Item.25.t1+      lambda26.t1*Item.26.t1+
lambda27.t1*Item.27.t1+      lambda28.t1*Item.28.t1+      lambda29.t1*Item.29.t1+
lambda30.t1*Item.30.t1
lambda1.t1+lambda2.t1+lambda3.t1+lambda4.t1+lambda5.t1+lambda6.t1+lambda7.t1+lambda8.t1+la
mbda9.t1+lambda10.t1+lambda11.t1+lambda12.t1+lambda13.t1+lambda14.t1+lambda15.t1+lambda16.
t1+lambda17.t1+lambda18.t1+lambda19.t1+lambda20.t1+lambda21.t1+lambda22.t1+lambda23.t1+lamb
bda24.t1+lambda25.t1+lambda26.t1+lambda27.t1+lambda28.t1+lambda29.t1+lambda30.t1==30

Item.1.t1 | tlambda1.t1*t1
Item.2.t1 | tlambda2.t1*t1
Item.3.t1 | tlambda3.t1*t1
Item.4.t1 | tlambda4.t1*t1
Item.5.t1 | tlambda5.t1*t1
Item.6.t1 | tlambda6.t1*t1
Item.7.t1 | tlambda7.t1*t1
Item.8.t1 | tlambda8.t1*t1
Item.9.t1 | tlambda9.t1*t1
Item.10.t1 | tlambda10.t1*t1
Item.11.t1 | tlambda11.t1*t1
Item.12.t1 | tlambda12.t1*t1
Item.13.t1 | tlambda13.t1*t1
Item.14.t1 | tlambda14.t1*t1
Item.15.t1 | tlambda15.t1*t1
Item.16.t1 | tlambda16.t1*t1
Item.17.t1 | tlambda17.t1*t1
Item.18.t1 | tlambda18.t1*t1
Item.19.t1 | tlambda19.t1*t1
Item.20.t1 | tlambda20.t1*t1
Item.21.t1 | tlambda21.t1*t1
Item.22.t1 | tlambda22.t1*t1
Item.23.t1 | tlambda23.t1*t1
Item.24.t1 | tlambda24.t1*t1
Item.25.t1 | tlambda25.t1*t1
Item.26.t1 | tlambda26.t1*t1
Item.27.t1 | tlambda27.t1*t1
Item.28.t1 | tlambda28.t1*t1
Item.29.t1 | tlambda29.t1*t1
Item.30.t1 | tlambda30.t1*t1
'

lavaan.model.fit <- lavaan(lavaan.model.t1.free,
                           data = U1,
                           int.ov.free = TRUE,
                           int.lv.free = FALSE,
                           meanstructure = TRUE,
                           std.lv =FALSE,
                           auto.fix.first = FALSE,
                           auto.var = TRUE,
                           auto.th = TRUE,
                           auto.delta = TRUE,
                           auto.cov.y = TRUE,
                           auto.var = TRUE,
                           ordered = colnames(U1),
                           parameterization = "theta")

summary ( lavaan.model.fit , standardized = TRUE )
fitMeasures(lavaan.model.fit)[c("cfi","tli","rmsea")]

```

```
t1.sem.param<-get_lmb_tau_alpha_psi(lavaan.model.fit)
t1.tri.param<-get_a_d_b(t1.sem.param,D=1.701)

ggplot2::ggplot(data = data.frame(SIM = a[1:30], EST = t1.tri.param$a[1:30]),
  aes(x = SIM, y = EST)) +
  geom_point() +
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Simulações do parâmetro a")

ggplot2::ggplot(data = data.frame(SIM = b[1:30], EST = t1.tri.param$b[1:30]),
  aes(x = SIM, y = EST)) +
  geom_point() +
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Simulações do parâmetro b")
```

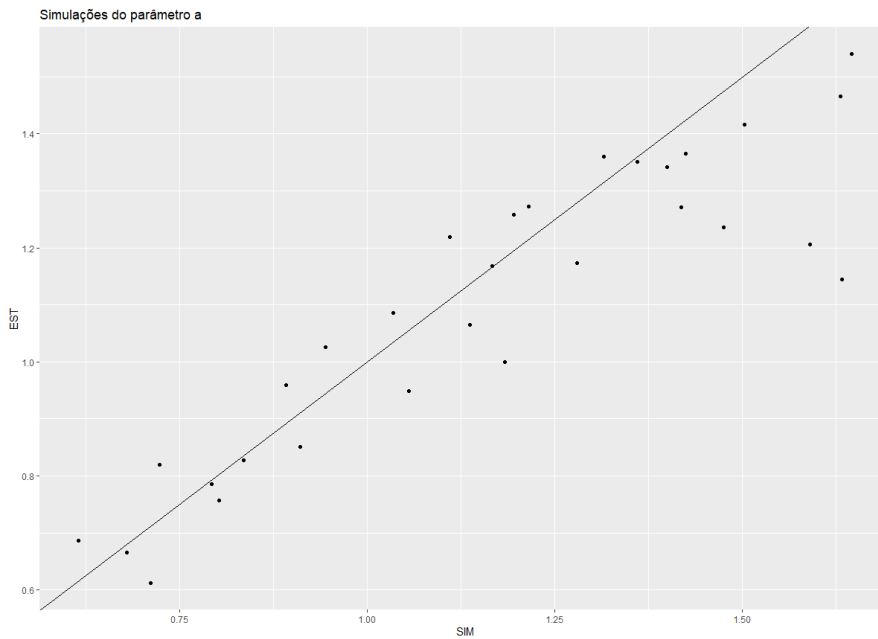


Figura A.3 Recuperação dos valores do parâmetro a utilizando a MEE. ML2P. 30 itens.

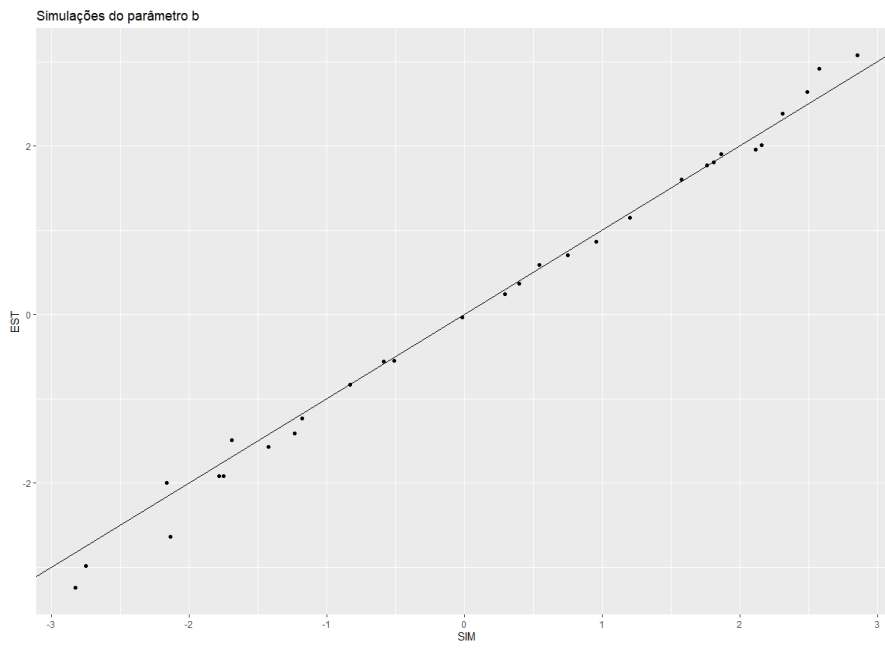


Figura A.4 Recuperação dos valores do parâmetro b utilizando a MEE. ML2P. 30 itens.

A.2 Simulações Unidimensionais Longitudinais com dois momentos

A.2.1 O modelo

O conjunto de equações que definem o modelo utilizado em simulações unidimensionais longitudinais ($K = 1, T = 2$), utilizando o modelo de 2 parâmetros e assumindo que a distribuição dos traços latentes dos H indivíduos nos dois momentos segue uma distribuição normal bidimensional multivariada $\boldsymbol{\eta}$ é definido pelas equações abaixo listadas

$$Y_{iht} \mid (\eta_{1t}, \boldsymbol{\zeta}_i) \sim \text{Bernoulli}(P_{ihkt}), t = \{1, 2\} \quad (\text{A.5})$$

$$P_{ihkt} = P(y_{iht} = 1 \mid \eta_{h1t}, \boldsymbol{\zeta}_i) = \frac{1}{1 + e^{a_i D(\eta_{h1t} - b_i)}}, t = \{1, 2\} \quad (\text{A.6})$$

$$\boldsymbol{\zeta}_i = (a_i, b_i), i = \{1, \dots, I\} \quad (\text{A.7})$$

$$\eta_{11} \mid \boldsymbol{\pi}_{11} \sim N(\mu_{11}, \sigma_{11}^2)$$

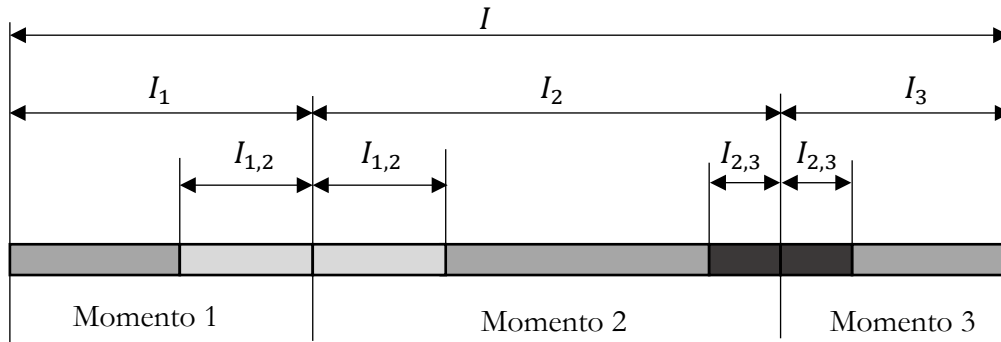
$$\eta_{12} \mid \boldsymbol{\pi}_{12} \sim N(\mu_{12}, \sigma_{12}^2)$$

$$\boldsymbol{\eta} \sim N(\boldsymbol{\mu}, \boldsymbol{\Psi}) \quad (\text{A.8})$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_{11} \\ \mu_{12} \end{pmatrix} \text{ e } \boldsymbol{\Psi} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{11,12}^2 \\ \sigma_{11,12}^2 & \sigma_{12}^2 \end{pmatrix}$$

Neste modelo assume-se que os parâmetros do conjunto de itens utilizados nos diferentes momentos estão calibrados na mesma escala do traço latente, ou, alternativamente, que existem itens comuns entre os dois momentos que permitam colocar os dois conjuntos em uma mesma escala através de métodos de equalização. Seja I o número total de itens aplicados nos três momentos, I_t o número de itens no momento t e I_{t_1, t_2} o número de itens comuns entre os momentos t_1 e t_2 , o número de itens distintos I_D é calculado pela seguinte equação. Figura A.5 mostra exemplo de como essas definições são utilizadas em um desenho longitudinal de três momentos.

$$I_D = I - \sum_{k=1}^{T-1} I_{t_k, t_{k+1}} \quad (\text{A.9})$$



$$I_D = I - I_{1,2} - I_{2,3}$$

Figura A.5 Exemplo de organização de itens em desenho longitudinal de 3 momentos.

A.2.2 Geração de dados simulados

Os passos necessários para se gerar as respostas a cada item por cada indivíduo são os seguintes.

1. Gere H amostras η .
2. Gere I_D pares de parâmetros ζ_i . Os valores de a_i são amostras de uma variável uniforme $U[0,6; 1,6]$. Os valores de b_i variaram de -3 a $+3$ em intervalos de 0.5 . Entre momentos, assume-se que os últimos I_{t_1, t_2} itens do momento t_1 são iguais aos primeiros I_{t_1, t_2} itens do momento seguinte.
3. Calcule P_{ihkt} para cada item e indivíduo conforme a Equação (A.6).
4. Utilizando o método da transformada inversa gera-se os valores Y_{it} para todos os I itens e H indivíduos segundo abaixo. O resultado é salvo em uma matriz $H \times I$.
 - a. Gera-se $U \sim U(0,1)$, então
 - b. Se definirmos $Y_{i1} = \mathbf{1}_{U \leq P_{iht}}$, então
 - i. $P(Y_{it} = 1) = P(U \leq P_{ihkt}) = P_{ihkt}, t = 1,2$
 - ii. $P(Y_{it} = 0) = P(U > P_{ihkt}) = 1 - P_{iht}$

O código utilizado na geração de dados para um desenho longitudinal de dois momentos ($T = 2$) com $I = 60, I_{1,2} = 10, I_D = 50$ é exemplificado a seguir. Após o carregamento das pacotes utilizados, os parâmetros dos itens e amostras da distribuição normal bidimensional multivariada $\tilde{\eta}_2$ são gerados

```
rm(list=ls())
if(!require(mirt)) install.packages("mirt"); library(mirt)
if(!require(mirtCAT)) install.packages("mirtCAT"); library(mirtCAT)
if(!require(MASS)) install.packages("MASS"); library(MASS)
if(!require(rockchalk)) install.packages("rockchalk"); library(rockchalk)
if(!require(lavaan)) install.packages("lavaan"); library(lavaan)
if(!require(ggplot2)) install.packages("ggplot2"); library(ggplot2)

N <- 2000 ## subjects
I= 50 # Number of Items
Ig = 30 #Items per group
Ic = 10 # Common Items
D= 1.0

PL=2 # Logistic Model (1,2,3 parameters)
rho<-0.7

coefs <- matrix(ncol=6,nrow=I)
colnames(coefs)=c("a1","b1","c1","a2","b2","c2")

if (PL==1) {a = rep(1,I)} else {a = runif(I,0.7, 1.6)} # U(0.5 , 2.5)
if (PL<=2) {c = rep(0,I)} else {c = runif(I,0.0, 0.3)} # U(0 , 0.3)
b = c(runif(20,-2, 2.0),runif(10,-0.5, 0.5),
      runif(20,-2, 2.0) )
d=-a*b # MIRT trabalha com o intercepto (d=-ab) e n?o com a dificuldade (b)

Mu <- c(0,0.5)
Sigma <- matrix( c(1,0.7,0.7,1),ncol=2)
Theta <- MASS::mvrnorm(n=N, mu=Mu, Sigma =Sigma)

colMeans(Theta)
cov(Theta)
```

Para cada momento t , os dados dicotômicos são gerados da mesma maneira executado no exemplo transversal. Uma matriz $N \times I_1$ é gerada para o momento 1 e uma matriz $N \times I_2$ é gerada para o momento 2, com $I_1 = I_2 = 30$ e $I_g = 10$ correspondentes aos itens 21 a 30. Itens 1 a 30 são utilizados no momento 1 enquanto que os itens 21 a 50 o são no momento 2.

```
#####
# momento 1
#####
eta = Theta[,1]*%*% t(D*a[1:30]) + matrix(D*d[1:30],N,30,byrow=TRUE);
eta=t(eta) # N x I (a'theta+d)
P = c[1:30] + (1-c[1:30])/(1+exp(-eta)); P=t(P) # n x I
X = runif(N*30); dim(X)=c(N,30) # matriz n x I de U(0,1)
U1 = 1*(X<P) ; U1=as.data.frame(U1) # AQUI TEMOS OS DADOS DICOT?MICOS
colnames(U1)=paste0("Item.",1:30,".t1")

#####
# momento 2
#####
eta = Theta[,2]*%*% t(a[21:50]) + matrix(d[21:50],N,30,byrow=TRUE);
eta=t(eta) # N x I (a'theta+d)
P = c[21:50] + (1-c[21:50])/(1+exp(-eta)); P=t(P) # n x I
X = runif(N*30); dim(X)=c(N,30) # matriz n x I de U(0,1)
U2 = 1*(X<P) ; U2=as.data.frame(U2) # AQUI TEMOS OS DADOS DICOT?MICOS
colnames(U2)=paste0("Item.",21:50,".t2")
```


A.2.3 Utilizando a TRI

Os dados dicotômicos gerados na seção A.2.2 são ajustados em um modelo de grupos múltiplos com ML2P itens. É importante notar que a estrutura dos dados utilizada nesse ajuste é como na Figura A.7. A organização é feita como se 50 itens fossem apresentados a todos os respondentes nos dois momentos, mas destes só 30 pudessem ser respondidos a cada momento.

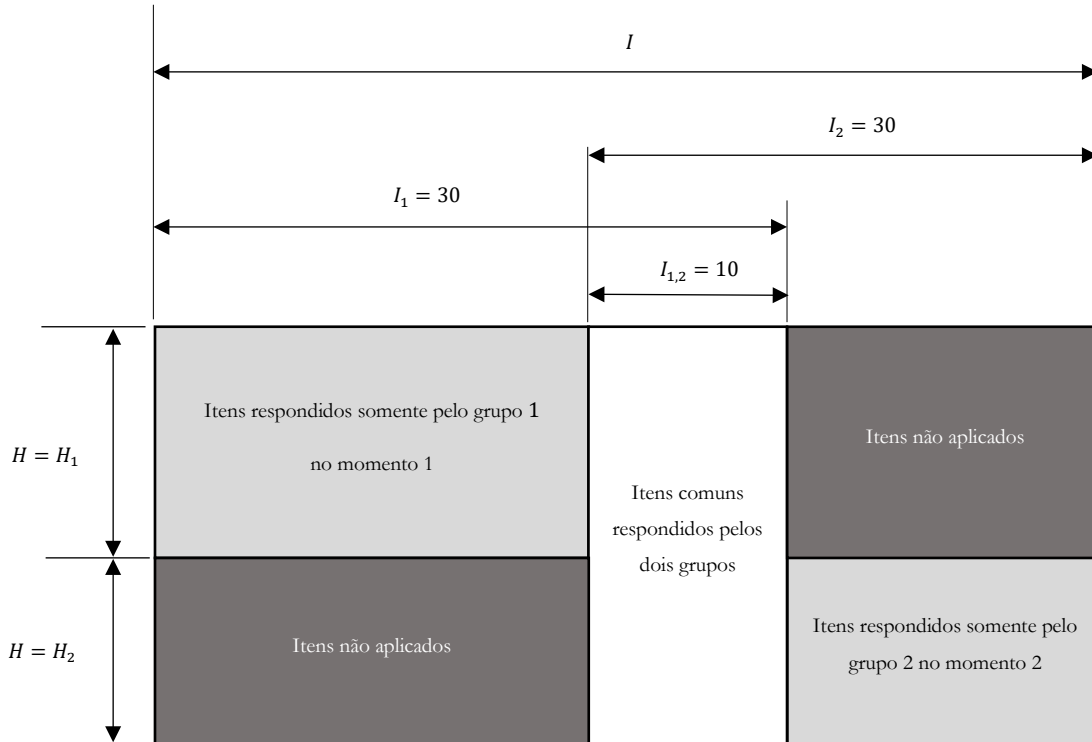


Figura A.6 Estrutura de entrada de dados para modelos de grupos múltiplos

O código utilizado na reestruturação dos dados gerados na seção A.2 é mostrado a seguir.

```
#Estrutura de dados para processamento de grupos múltiplos
mNA = rep(NA, (N*20));#filler
dim(mNA)=c(N,20) # matriz n x I de U(0,1)
dim(mNA)

itemnames <- c(1:50)
U1t1 <- cbind(U1,mNA)
names(U1t1)<-itemnames
U2t2 <- cbind(mNA,U2)
names(U2t2)<-itemnames
Ut <-rbind(U1t1,U2t2)
Ut[(N-10):(N+10),15:35]
dim(Ut)

#group names
group_names= paste0("G",rep(1:2, each=N)) #G1...G1,G2...G2, G3..G3
group_names[0:10]
group_names[(N-10):(N+10)]
models <- 'F1 = 1-50'

mod <- mirt::multipleGroup(
  model=models,
  data = Ut,
  itemtype='2PL',
  group = group_names,
  invariance = c(itemnames,'free_means','free_var','slopes','intercepts'))

par<-coef(mod,simplify=TRUE,IRTpars = TRUE)
par
```

A recuperação dos parâmetros de discriminação é confirmada pelo seguinte código e gráfico.

```
plot(a[1:50],par$G1[1][['items']][,1],
  main="Recuperação dos parâmetros de discriminação (a)",
  xlab="Valores verdadeiros",
  ylab="Estimativas",
  asp = 1);
lines(c(-4,4),c(-4,4), col = "blue",)
```

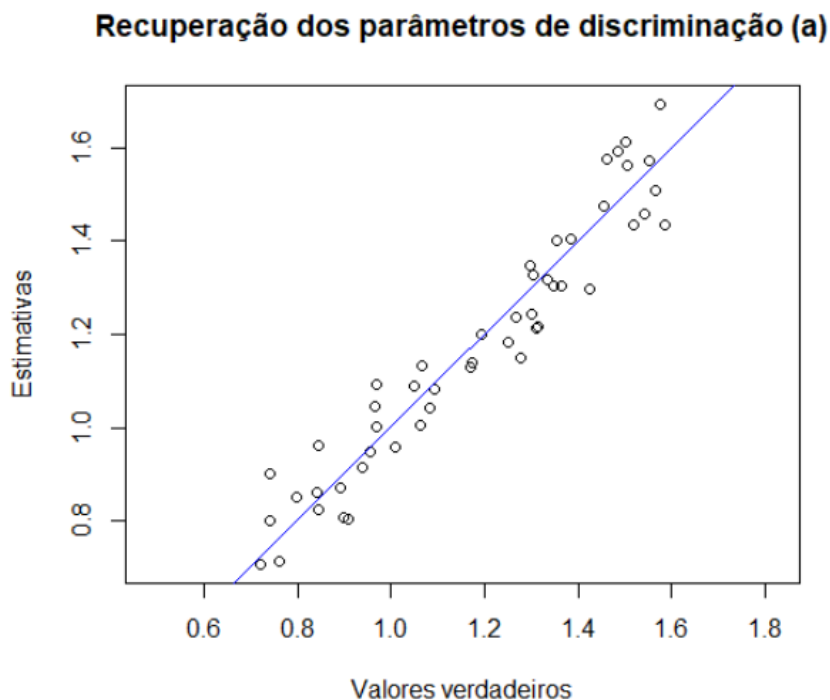


Figura A.7 - Recuperação de a - modelo longitudinal de grupos múltiplos

A recuperação dos parâmetros de dificuldade é confirmada pelo seguinte código e gráfico.

```
plot(b[1:50],par$G1[1][['items']][,2],
     main="Recuperação dos parâmetros de dificuldade (b)",
     xlab="Valores verdadeiros",
     ylab="Estimativas",
     asp = 1);
lines(c(-4,4),c(-4,4), col = "blue",)
```

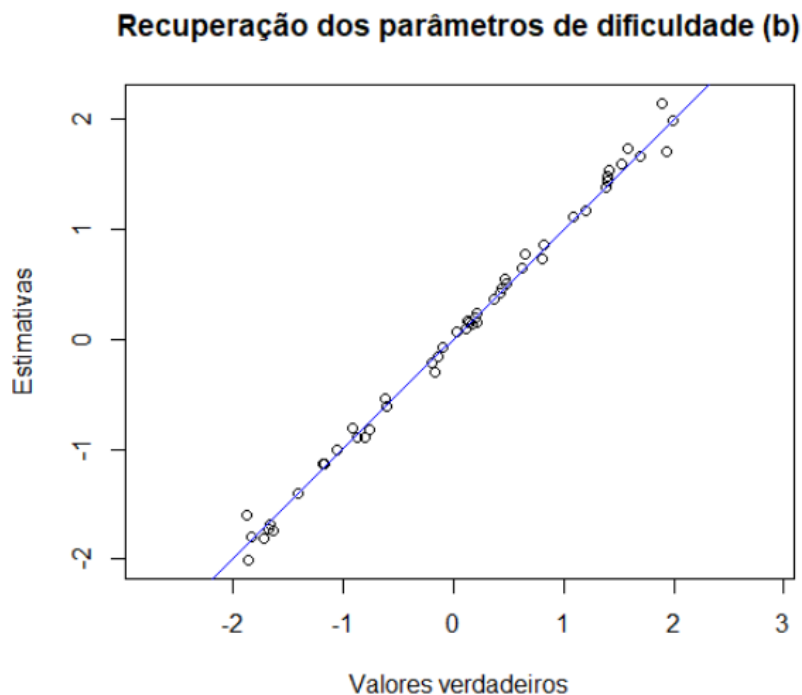


Figura A.8 - Recuperação de b - Modelo longitudinal de grupos múltiplos

As proficiências são obtidas com os seguintes comandos e pode se verificar que os valores obtidos são próximos dos simulados. Tabela A.1 mostra que a média é recuperada com elevada precisão e o mesmo não se pode dizer sobre a variância dado que está abordagem não a modela diretamente.

```
profic = fscores(mod) #estimativas das proficiências individuais

est.factors <- cbind(profic[1:N], profic[(N+1):(2*N)])
colMeans(est.factors)
cov(est.factors)
cor(est.factors)
```

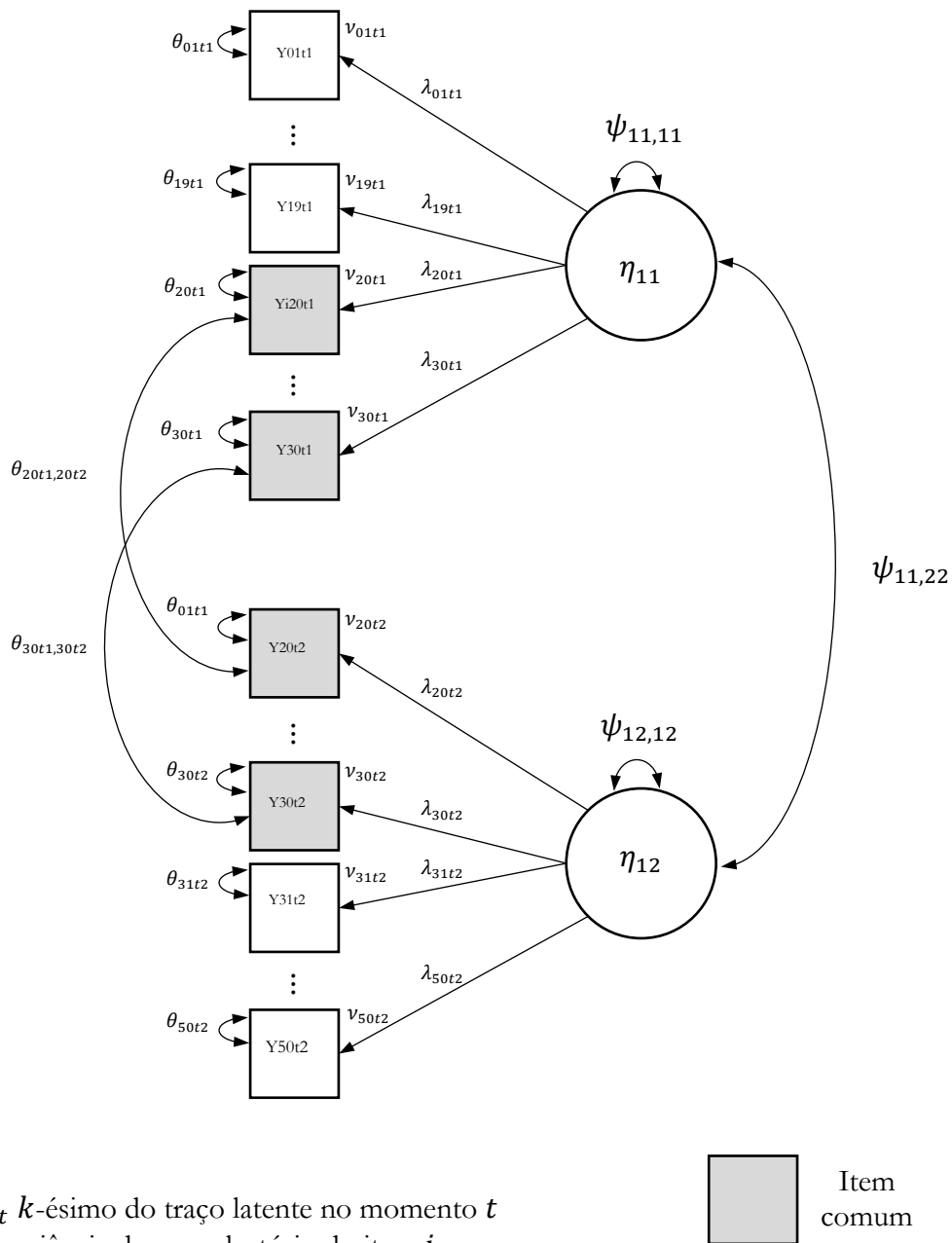
Tabela A.1 Recuperação de parâmetros populacionais.

Parâmetros populacionais	Valores simulados	Valores estimados
Média μ	$\begin{pmatrix} 0 \\ 0,5 \end{pmatrix}$	$\begin{pmatrix} -0,00047 \\ 0,552622 \end{pmatrix}$
Matriz de Covariância Σ	$\begin{pmatrix} 1 & 0,7 \\ 0,7 & 1 \end{pmatrix}$	$\begin{pmatrix} 0,874 & 0,541 \\ 0,541 & 0,938 \end{pmatrix}$
Matriz de Correlação	$\begin{pmatrix} 1 & 0,7 \\ 0,7 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0,598 \\ 0,598 & 1 \end{pmatrix}$

A.2.4 Utilizando a MEE

Os dados dicotômicos gerados na seção A.2.2 são ajustados em um modelo longitudinal conforme Figura A.9. As medidas tomadas nos diferentes momentos tem itens comuns para quais a invariância fatorial forte é assumida, ou seja, seus interceptos e cargas fatoriais não sofrem alteração através dos momentos. Assume-se também que exista uma covariância residual entre os itens comuns correspondentes.

A parametrizações utilizadas foram: `effects_coding`, `conditional_parameterization` e `free_threshold`.



η_{kt} k -ésimo do traço latente no momento t
 θ_i variância do erro aleatório do item i
 $\theta_{i,j}$ variância do erro aleatório entre itens i e j
 λ_i carga fatorial do item i
 v_i intercepto do item i
 α_{kt} média do traço latente η_{kt}
 $\psi_{.k_1t_1, k_2t_2}$ covariância entre traços latentes $\eta_{k_1t_1}$ e $\eta_{k_2t_2}$

Item comum

Figura A.9 Modelo longitudinal MEE de dois momentos.

O código que define o modelo configurado é como abaixo.

```

lavaan.model.t12.free.CommonItems <- '

# Modelo de mensuração
eta1 =~ lbd1.t1*i1.t1 + lbd2.t1*i2.t1 + lbd3.t1*i3.t1 + lbd4.t1*i4.t1 + lbd5.t1*i5.t1 +
lbd6.t1*i6.t1 + lbd7.t1*i7.t1 + lbd8.t1*i8.t1 + lbd9.t1*i9.t1 + lbd10.t1*i10.t1 +
lbd11.t1*i11.t1 + lbd12.t1*i12.t1 + lbd13.t1*i13.t1 + lbd14.t1*i14.t1 + lbd15.t1*i15.t1 +
lbd16.t1*i16.t1 + lbd17.t1*i17.t1 + lbd18.t1*i18.t1 + lbd19.t1*i19.t1 + lbd20.t1*i20.t1 +
lbd21.t1*i21.t1 + lbd22.t1*i22.t1 + lbd23.t1*i23.t1 + lbd24.t1*i24.t1 + lbd25.t1*i25.t1 +
lbd26.t1*i26.t1 + lbd27.t1*i27.t1 + lbd28.t1*i28.t1 + lbd29.t1*i29.t1 + lbd30.t1*i30.t1
eta2 =~ lbd21.t2*i21.t2 + lbd22.t2*i22.t2 + lbd23.t2*i23.t2 + lbd24.t2*i24.t2 +
lbd25.t2*i25.t2 + lbd26.t2*i26.t2 + lbd27.t2*i27.t2 + lbd28.t2*i28.t2 + lbd29.t2*i29.t2 +
lbd30.t2*i30.t2 + lbd31.t2*i31.t2 + lbd32.t2*i32.t2 + lbd33.t2*i33.t2 + lbd34.t2*i34.t2 +
lbd35.t2*i35.t2 + lbd36.t2*i36.t2 + lbd37.t2*i37.t2 + lbd38.t2*i38.t2 + lbd39.t2*i39.t2 +
lbd40.t2*i40.t2 + lbd41.t2*i41.t2 + lbd42.t2*i42.t2 + lbd43.t2*i43.t2 + lbd44.t2*i44.t2 +
lbd45.t2*i45.t2 + lbd46.t2*i46.t2 + lbd47.t2*i47.t2 + lbd48.t2*i48.t2 + lbd49.t2*i49.t2 +
lbd50.t2*i50.t2

# Parametrização offset coding
lbd1.t1+lbd2.t1+lbd3.t1+lbd4.t1+lbd5.t1+lbd6.t1+lbd7.t1+lbd8.t1+lbd9.t1+lbd10.t1+lbd11.t1+
lbd12.t1+lbd13.t1+lbd14.t1+lbd15.t1+lbd16.t1+lbd17.t1+lbd18.t1+lbd19.t1+lbd20.t1+lbd21.t1+
lbd22.t1+lbd23.t1+lbd24.t1+lbd25.t1+lbd26.t1+lbd27.t1+lbd28.t1+lbd29.t1+lbd30.t1==30
lbd21.t2+lbd22.t2+lbd23.t2+lbd24.t2+lbd25.t2+lbd26.t2+lbd27.t2+lbd28.t2+lbd29.t2+lbd30.t2+
lbd31.t2+lbd32.t2+lbd33.t2+lbd34.t2+lbd35.t2+lbd36.t2+lbd37.t2+lbd38.t2+lbd39.t2+lbd40.t2+
lbd41.t2+lbd42.t2+lbd43.t2+lbd44.t2+lbd45.t2+lbd46.t2+lbd47.t2+lbd48.t2+lbd49.t2+lbd50.t2=
=30

#Covarianção
eta1 ~~ eta2

# Intercept da variável latente. O intercept to primeiro momento é fixado em zero.
# Outros momentos são estimados em referência ao momento 1.
eta1 ~ 0*1
eta2 ~ eta2m*1

#definição dos thresholds - momento 1
i1.t1 | thr1.t1*t1
i2.t1 | thr2.t1*t1
i3.t1 | thr3.t1*t1
i4.t1 | thr4.t1*t1
i5.t1 | thr5.t1*t1
i6.t1 | thr6.t1*t1
i7.t1 | thr7.t1*t1
i8.t1 | thr8.t1*t1
i9.t1 | thr9.t1*t1
i10.t1 | thr10.t1*t1
i11.t1 | thr11.t1*t1
i12.t1 | thr12.t1*t1
i13.t1 | thr13.t1*t1
i14.t1 | thr14.t1*t1
i15.t1 | thr15.t1*t1
i16.t1 | thr16.t1*t1
i17.t1 | thr17.t1*t1
i18.t1 | thr18.t1*t1
i19.t1 | thr19.t1*t1
i20.t1 | thr20.t1*t1
i21.t1 | thr21.t1*t1
i22.t1 | thr22.t1*t1
i23.t1 | thr23.t1*t1
i24.t1 | thr24.t1*t1
i25.t1 | thr25.t1*t1
i26.t1 | thr26.t1*t1
i27.t1 | thr27.t1*t1
i28.t1 | thr28.t1*t1
i29.t1 | thr29.t1*t1
i30.t1 | thr30.t1*t1

```

```

#definição dos thresholds - momento 2
i21.t2 | thr21.t2*t1
i22.t2 | thr22.t2*t1
i23.t2 | thr23.t2*t1
i24.t2 | thr24.t2*t1
i25.t2 | thr25.t2*t1
i26.t2 | thr26.t2*t1
i27.t2 | thr27.t2*t1
i28.t2 | thr28.t2*t1
i29.t2 | thr29.t2*t1
i30.t2 | thr30.t2*t1
i31.t2 | thr31.t2*t1
i32.t2 | thr32.t2*t1
i33.t2 | thr33.t2*t1
i34.t2 | thr34.t2*t1
i35.t2 | thr35.t2*t1
i36.t2 | thr36.t2*t1
i37.t2 | thr37.t2*t1
i38.t2 | thr38.t2*t1
i39.t2 | thr39.t2*t1
i40.t2 | thr40.t2*t1
i41.t2 | thr41.t2*t1
i42.t2 | thr42.t2*t1
i43.t2 | thr43.t2*t1
i44.t2 | thr44.t2*t1
i45.t2 | thr45.t2*t1
i46.t2 | thr46.t2*t1
i47.t2 | thr47.t2*t1
i48.t2 | thr48.t2*t1
i49.t2 | thr49.t2*t1
i50.t2 | thr50.t2*t1

#definição das variâncias e covariâncias - momento 1
i1.t1 ~~ i1.t1
i2.t1 ~~ i2.t1
i3.t1 ~~ i3.t1
i4.t1 ~~ i4.t1
i5.t1 ~~ i5.t1
i6.t1 ~~ i6.t1
i7.t1 ~~ i7.t1
i8.t1 ~~ i8.t1
i9.t1 ~~ i9.t1
i10.t1 ~~ i10.t1
i11.t1 ~~ i11.t1
i12.t1 ~~ i12.t1
i13.t1 ~~ i13.t1
i14.t1 ~~ i14.t1
i15.t1 ~~ i15.t1
i16.t1 ~~ i16.t1
i17.t1 ~~ i17.t1
i18.t1 ~~ i18.t1
i19.t1 ~~ i19.t1
i20.t1 ~~ i20.t1
#itens comuns covariam
i21.t1 ~~ i21.t1 + i21.t2
i22.t1 ~~ i22.t1 + i22.t2
i23.t1 ~~ i23.t1 + i23.t2
i24.t1 ~~ i24.t1 + i24.t2
i25.t1 ~~ i25.t1 + i25.t2
i26.t1 ~~ i26.t1 + i26.t2
i27.t1 ~~ i27.t1 + i27.t2
i28.t1 ~~ i28.t1 + i28.t2
i29.t1 ~~ i29.t1 + i29.t2
i30.t1 ~~ i30.t1 + i30.t2

```

```
#definição das variâncias e covariâncias - momento 2
i21.t2 ~~ i21.t2
i22.t2 ~~ i22.t2
i23.t2 ~~ i23.t2
i24.t2 ~~ i24.t2
i25.t2 ~~ i25.t2
i26.t2 ~~ i26.t2
i27.t2 ~~ i27.t2
i28.t2 ~~ i28.t2
i29.t2 ~~ i29.t2
i30.t2 ~~ i30.t2
i31.t2 ~~ i31.t2
i32.t2 ~~ i32.t2
i33.t2 ~~ i33.t2
i34.t2 ~~ i34.t2
i35.t2 ~~ i35.t2
i36.t2 ~~ i36.t2
i37.t2 ~~ i37.t2
i38.t2 ~~ i38.t2
i39.t2 ~~ i39.t2
i40.t2 ~~ i40.t2
i41.t2 ~~ i41.t2
i42.t2 ~~ i42.t2
i43.t2 ~~ i43.t2
i44.t2 ~~ i44.t2
i45.t2 ~~ i45.t2
i46.t2 ~~ i46.t2
i47.t2 ~~ i47.t2
i48.t2 ~~ i48.t2
i49.t2 ~~ i49.t2
i50.t2 ~~ i50.t2

# Assume-se que a invariância forte é mantida.
# cargas e interceptos dos itens comuns são iguais.
lbd21.t1 ==lbd21.t2
lbd22.t1 ==lbd22.t2
lbd23.t1 ==lbd23.t2
lbd24.t1 ==lbd24.t2
lbd25.t1 ==lbd25.t2
lbd26.t1 ==lbd26.t2
lbd27.t1 ==lbd27.t2
lbd28.t1 ==lbd28.t2
lbd29.t1 ==lbd29.t2
lbd30.t1 ==lbd30.t2

thr21.t1 ==thr21.t2
thr22.t1 ==thr22.t2
thr23.t1 ==thr23.t2
thr24.t1 ==thr24.t2
thr25.t1 ==thr25.t2
thr26.t1 ==thr26.t2
thr27.t1 ==thr27.t2
thr28.t1 ==thr28.t2
thr29.t1 ==thr29.t2
thr30.t1 ==thr30.t2
,
```

Após o ajuste do modelo os parâmetros de discriminação a e dificuldade b foram calculados a partir das cargas e os threshold estimados segundo as fórmulas descritas na Seção E.5. As estatísticas que mostram a qualidade do ajuste do modelo são mostradas na Tabela A.2.

Tabela A.2 Estatísticas de Ajuste de Modelo Longitudinal MEE

TLI	CFI	RMSEA
1.00114	1.000000	0.000000

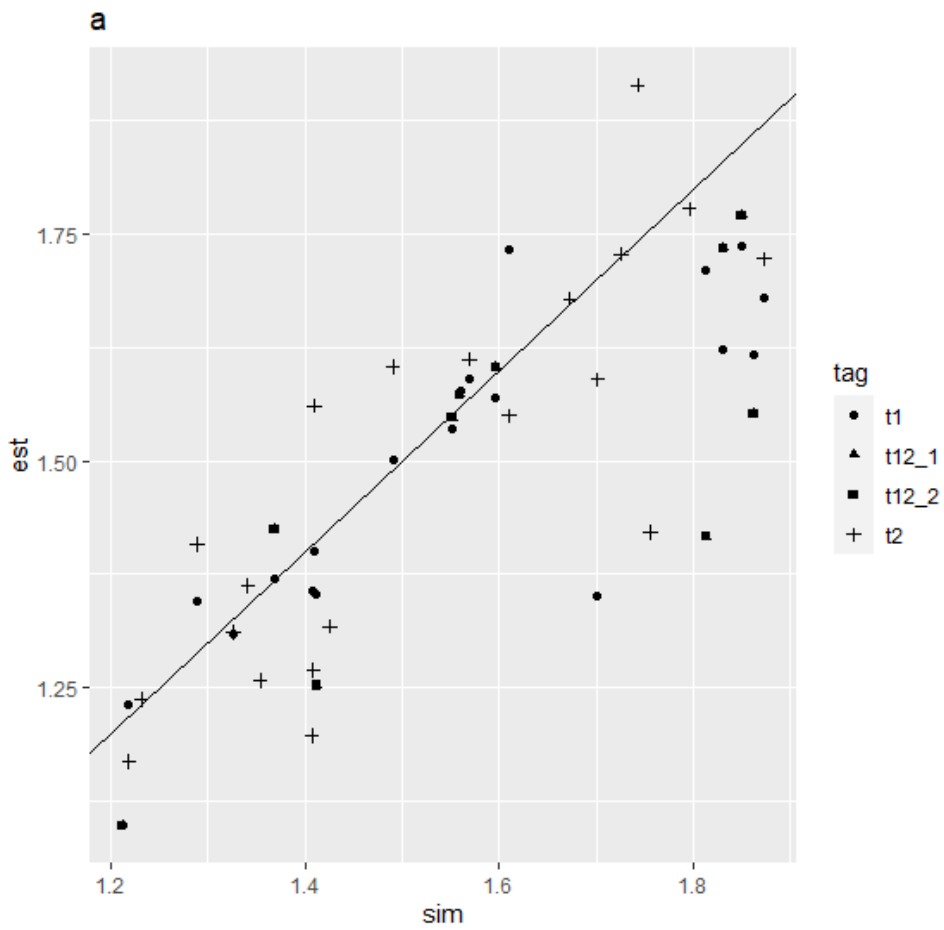


Figura A.10 Recuperação do parâmetro de dificuldade a . Modelo longitudinal MEE.

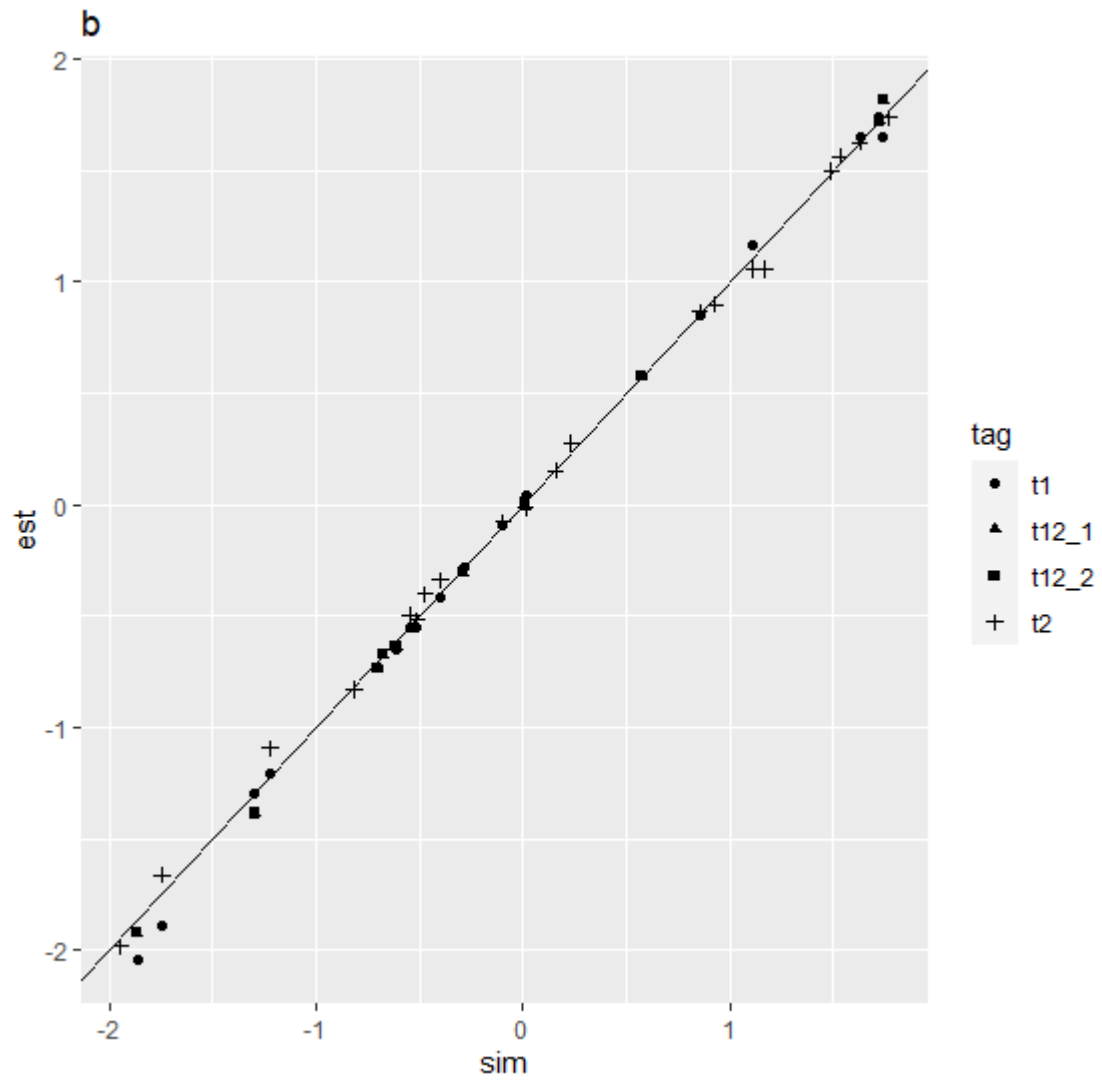


Figura A.11 Recuperação do parâmetro de dificuldade b . Modelo longitudinal MEE.

Tabela A.3 Recuperação de parâmetros populacionais.

Parâmetros Populacionais	Valores Simulados	Valores Estimados
Média μ	$\begin{pmatrix} 0 \\ 0,5 \end{pmatrix}$	$\begin{pmatrix} 0,000 \\ 0,474 \end{pmatrix}$
Matriz de Covariância Σ	$\begin{pmatrix} 1 & 0,7 \\ 0,7 & 1 \end{pmatrix}$	$\begin{pmatrix} 0,766 & 0,510 \\ 0,510 & 0,713 \end{pmatrix}$
Matriz de Correlação	$\begin{pmatrix} 1 & 0,7 \\ 0,7 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0,691 \\ 0,691 & 1 \end{pmatrix}$

Apêndice B Indeterminação fatorial

Este experimento visa explorar a dependência do número de itens e a correlação entre o traço latente estimado e simulado. Os valores apresentados foram obtidos simulando a resposta de 6000 indivíduos e são a média de trinta experimentos utilizando modelos da TRI ou da MEE. A Figura C.1 mostra que número expressivo de itens é necessário para se obter correlações acima de 0.95 entre os valores estimados e simulados. Isso evidencia a limitada quantidade de informação sobre o traço latente que pode ser obtida a partir de um único item e a importância de se maximizar a capacidade coletiva de coleta de informação de teste através da escolha cirúrgica não só dos melhores itens, mas também da melhor distribuição desses itens ao longo da escala. A figura evidencia também que o mesmo comportamento é observado independentemente do tipo de modelagem utilizado.

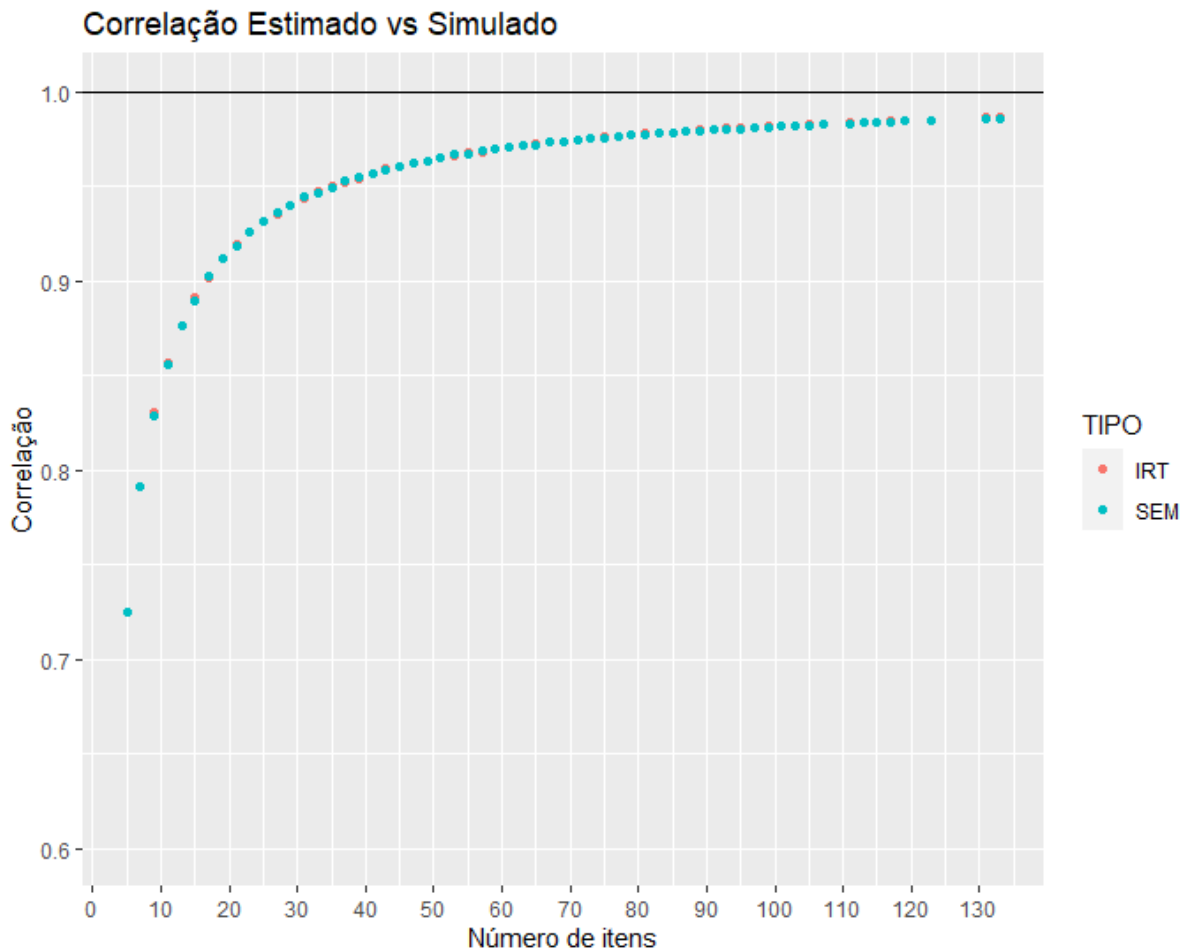


Figura B.1 – Número de itens versus a correlação entre traço latente estimado e simulado.

Abaixo segue o código utilizado na geração da Figura C.1.

```

N <- 6000
runs <- 30
I <- seq(5,10,by=2)
step <- 3/((I-1)/2)
data<- data.frame(ItemNumber= rep(NA,(runs)),cor.irt = rep(NA,(runs)), cor.sem =
rep(NA,(runs)))
log_data <- list()
for(i in 1:length(I)){
  for(j in 1:runs){
    a <- rep(1.7,I[i])
    b <- seq(-3.0, +3.0,by=step[i])
    d=-a*b # MIRT trabalha com o intercepto (d=-ab) e nao com a dificuldade (b)
    pars <- cbind(a,b,d)
    Eta <- mvrnorm(n=N, mu=0, 1 )
    dat <- simdata(a=a,d=d,N=N,
                  itemtype = '2PL',
                  Theta = matrix(Eta[,1],ncol=1,nrow = length(Eta[,1])))
    colnames(dat) <- paste("item",1:I[i],sep="")
    mod<-mirt(data=dat,model = 1,itemtype = "2PL")
    scores <- fscores(mod)
    data$ItemNumber[j] <- I[i]
    data$cor.irt[j] <- cor(Eta[,1], scores)[1,1]
  } #for j
  log_data[[i]]<- data
}#for(i)

```

Apêndice C Propriedades da esperança e da covariância

C.1 Esperança

C.1.1 Notação

Define-se $E(X)$ como a esperança de uma variável aleatória X .

C.1.2 Propriedades da esperança

A seguir segue uma lista não exaustiva das principais propriedades da esperança de variáveis aleatórias.

C.1.2.1 Esperança de uma constante

Seja $c \in \mathbb{R}$ uma constante, então

$$E(c) = c. \quad (\text{C.1})$$

C.1.2.2 Multiplicação por escalar

Seja X_1 uma variável aleatória e $c \in \mathbb{R}$ uma constante, então:

$$E(cX_1) = cE(X_1). \quad (\text{C.2})$$

C.1.2.3 Combinação linear de variáveis aleatórias

Seja X_1, \dots, X_k, K variáveis aleatórias e c_1, \dots, c_k, K constantes então

$$E(c_1X_1 + \dots + c_kX_k) = c_1E(X_1) + \dots + c_kE(X_k). \quad (\text{C.3})$$

C.1.2.4 Produto de variáveis aleatórias

Seja X_1 e X_2 são variáveis aleatórias **independentes**, então

$$E(X_1X_2) = E(X_1)E(X_2). \quad (\text{C.4})$$

Em termos gerais,

$$E(X_1X_2) \neq E(X_1)E(X_2). \quad (\text{C.5})$$

C.2 Covariância

A seguir segue uma lista não exaustiva das principais propriedades da covariância de variáveis aleatórias.

C.2.1 Notação

Define-se $COV(X_1, X_2)$, escrito em letras maiúsculas, como a covariância populacional entre as variáveis aleatórias X_1 e X_2 , tal que,

$$COV(X_1, X_2) = E\{ [X_1 - E(X_1)][X_2 - E(X_2)] \}, \quad (C.6)$$

e $VAR(X)$ como a variância populacional da variável aleatória X , tal que,

$$VAR(X) = COV(X, X). \quad (C.7)$$

Define-se também $cov(X_1, X_2)$, escrito em letras minúsculas, como a covariância amostral entre as amostras das variáveis aleatórias X_1 e X_2 , tal que, seu estimador não viesado é dado por,

$$cov(X_1, X_2) = \frac{\sum_{i=1}^N (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{N-1}, \quad (C.8)$$

em que x_{1i} e x_{2i} representam, respectivamente, as i -ésima amostras, e \bar{x}_1 e \bar{x}_2 , e as médias amostrais das variáveis aleatórias X_1 e X_2 . E, similarmente, a $var(X)$ é definida como a variância amostral da variável aleatória X , tal que,

$$var(X) = cov(X, X). \quad (C.9)$$

C.2.2 Definições

C.2.2.1 Covariância Populacional

Sejam X_1 e X_2 variáveis aleatórias, a covariância populacional entre elas é definida por;

$$\begin{aligned} COV(X_1, X_2) &= E\{ [X_1 - E(X_1)][X_2 - E(X_2)] \} \\ COV(X_1, X_2) &= E\{ X_1X_2 - X_1E(X_2) - E(X_1)X_2 + E(X_1)E(X_2) \} \end{aligned} \quad (C.10)$$

$$COV(X_1, X_2) = E(X_1 X_2) - E(X_1 E(X_2)) - E(E(X_1) X_2) + E(E(X_1) E(X_2))$$

$$COV(X_1, X_2) = E(X_1 X_2) - E(X_1) E(X_2) - E(X_1) E(X_2) + E(X_1) E(X_2)$$

$$COV(X_1, X_2) = E(X_1 X_2) - E(X_1) E(X_2).$$

De(D.10), obtém-se facilmente a equação para a variância populacional.

$$VAR(X) = COV(X, X) = E\{ [X - E(X)]^2 \} \tag{C.11}$$

$$VAR(X) = E(X^2) - E(X)^2$$

C.2.2.2 Covariância amostral

Sejam X_1 e X_2 variáveis aleatórias, o estimador não viesado da covariância entre N amostras dessas variáveis é dado por,

$$cov(X_1, X_2) = \frac{\sum_{i=1}^N (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{N - 1}, \tag{C.12}$$

em que x_{1i} e x_{2i} representam, respectivamente, as i -ésima amostras, e \bar{x}_1 e \bar{x}_2 , e as médias amostrais das variáveis aleatórias X_1 e X_2 .

C.2.3 Propriedades da covariância

C.2.3.1 Covariância com um escalar

Seja X_1 uma variável aleatória e $c \in \mathbb{R}$ uma constante, então:

$$COV(c, X_1) = 0. \tag{C.13}$$

C.2.3.2 Variância de um escalar

Fazendo $X_1 = c$ em (D.13) temos

$$VAR(c) = COV(c, c) = 0 \tag{C.14}$$

C.2.3.3 Multiplicação por escalar

Seja X_1, X_2 , variáveis aleatórias e $c \in \mathbb{R}$ uma constante, então:

$$COV(cX_1, X_2) = cCOV(X_1, X_2). \quad (C.15)$$

Da qual segue que

$$VAR(cX) = COV(cX, cX) = c^2COV(X, X) = c^2VAR(X). \quad (C.16)$$

C.2.3.4 Covariância com uma soma de variáveis aleatórias

Seja X_1, X_2 e X_3 variáveis aleatórias, então:

$$COV(X_1 + X_2, X_3) = COV(X_1, X_3) + COV(X_2, X_3). \quad (C.17)$$

C.2.3.5 Translação por escalar

Seja X_1, X_2 , variáveis aleatórias e $c \in \mathbb{R}$ uma constante, então:

$$COV(X_1 + c, X_2) = COV(X_1, X_2) + COV(c, X_2) = COV(X_1, X_2). \quad (C.18)$$

De qual segue que,

$$VAR(X + c) = COV(X + c, X + c) = COV(X, X) = VAR(X). \quad (C.19)$$

Apêndice D Parameterization effects on the Relation between Structural Equation and Item Response Models using binary indicators

Author Hugo Kenji Pereira Harada^{a*}, Lucia Pereira Barroso^a, Heliton Ribeiro Tavares^b, Julia Maria Pavan Soler^a

^aInstitute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil

^bFederal University of Pará, Pará, Brazil

Provide full correspondence details here including e-mail for the *corresponding author

Hugo Kenji Pereira Harada

Address: Rua Dr. Manoel de Paiva Ramos, 60, apto 83R,
 São Paulo – SP, CEP 05351-015, Brazil

E-mail: hkharada@ime.usp.br

This paper reviews known parameterization options available in the Structural Equation Model literature and extends previously proposed conversion formulas between Structural Equation and Item Response Model parameters. Simulation studies demonstrate the validity of the transformations.

Keywords: model equivalency, parameterization, identification, common factor, conversion formula

D.1 Introduction

In latent variable analysis, parameterization is done to set the origin and scale of latent variables and to guarantee that the system is identified. Numerous alternatives make parameterization one of the most confusing topics in Structure Equation Modeling (SEM) as terminology is inconsistent and tools are somewhat obscure about their default settings and assumptions. This poses a problem when relating Item Response Theory (IRT) and SEM parameters as the shape of the formulas change depending on parameterization choice. Kamata and Bauer (2008) successfully summarized previous work by Takane and Leeuw (1987), Mc Donald (1999), Millsap and Yun-Tein (2004), and Muthén and Asparouhov (2002), and introduced a general conversion formula to convert SEM parameters to IRT parameters under different parameterization methods. That work approached four different methods, and this article explains extra options which combined result in 18 parameterization schemes and proposes a minor adjustment to the general conversion formula presented by Kamata and Bauer to account for the updates. Simulation results show that these parameterizations lead to statistical equivalent models, and that conversions are accurate with the adjusted formula. The facilities of the `MIRT` and `laavan` R packages developed by, respectively, Rosseel (2012) and Chalmers (2012) were used to evaluate the performance of the simulation results.

D.2 Latent variable models - SEM and IRT approaches

Conditional probability models for latent variables can be obtained in two ways. First, by positing a conditional probability model based on cumulative logits, or, second, by deriving a conditional probability model from assuming a linear relation between the response variables and the common latent factor. The first approach is traditionally used in IRT literature with Rasch, two parameter and graded response models, while the second is used in Confirmatory Factor Analysis (CFA) and SEM model analysis.

D.3 SEM with binary indicators

As mentioned earlier, categorical indicators do not have a linear relationship with the common factor as do its continuous counterparts. For each categorical indicator y_{hi} , associated with item i and individual h , there is a continuous underlying latent response variable y_{hi}^* , such that, analogical to the continuous model, is linearly related to the common factor η_h of person h according to the following equation,

$$y_{hi}^* = v_i + \lambda_i \eta_h + \varepsilon_i, \quad (\text{D.1})$$

where v_i is the intercept, λ_i is the factor loading, and ε_i is the random error. Here, with no loss of generality, index h can be dropped from the common factor and η can be taken as random variable of factor scores. Equation (1), therefore, becomes

$$y_i^* = v_i + \lambda_i \eta + \varepsilon_i. \quad (\text{D.2})$$

Further, the following threshold model is necessary to establish the connection between y_i and y_i^*

$$y_i = \begin{cases} 0 & \text{if } -\infty < y_i^* \leq \tau_i \\ 1 & \text{if } \tau_i < y_i^* < +\infty \end{cases}, \quad (\text{D.3})$$

where τ_i denotes the threshold between categories along the latent response scale. Figure D.1 shows the case for a binary y_i in that the shaded area depicts $P[y_i = 1]$ for a given y_i^* distribution.

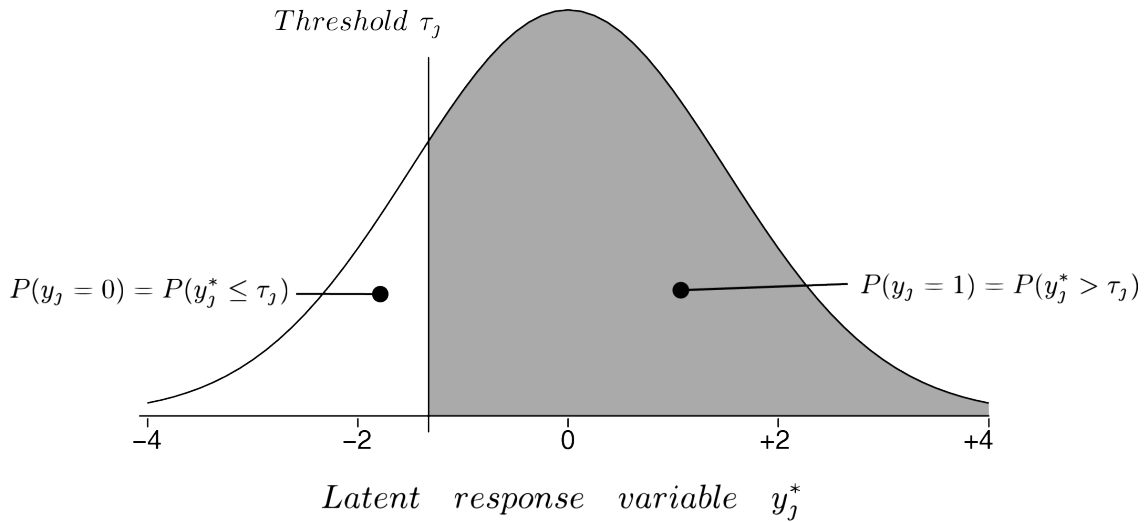


Figure D.1 Threshold model for a latent response variable with a binary indicator

When $y_i^* \sim N(\mu_i^*, \sigma_i^{*2})$, $\eta \sim N(\alpha, \psi)$, $\varepsilon_i \sim N(0, \theta_i)$, it is easy to show from equation (E.1) that:

$$\mu_i^* = v_i + \lambda_i \alpha, \quad (\text{D.4})$$

$$\sigma_i^{*2} = \lambda_i^2 \psi + \theta_i. \quad (\text{D.5})$$

Equation (D.4) relates the indicator means to the factor mean, while (D.5) relates the indicator variances to the factor variance. From these equations, it also follows that the conditional distribution of y_i^* given η is

$$y_i^* | \eta \sim N(\mu_c^*, \sigma_c^{*2}) = N(v_i + \lambda_i \eta, \theta_i). \quad (D.6)$$

Therefore, the conditional probability of $y_i = 1$ given η is

$$P[y_i = 1 | \eta] = \int_{z_\tau}^{\infty} \frac{1}{2\pi} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{-z_\tau} \frac{1}{2\pi} e^{-\frac{z^2}{2}} dz = \phi(-z_\tau)$$

$$P[y_i = 1 | \eta] = \phi\left(\frac{\mu_c^* - \tau_i}{\sigma_c^*}\right). \quad (D.7)$$

Combining (E.6) and (E.7),

$$P[y_i = 1 | \eta] = \phi\left(\frac{v_i + \lambda_i \eta - \tau_i}{\sqrt{\theta_i}}\right) = \phi\left(\frac{v_i - \tau_i}{\sqrt{\theta_i}} + \frac{\lambda_i}{\sqrt{\theta_i}} \eta\right). \quad (D.8)$$

D.4 IRT with binary indicators

The logistic regression model, a special case of a Generalized Linear Model when Y assumes a binomial distribution, is equivalent to the two-parameter logistic model in IRT, such that:

$$\frac{P(y_i = 1 | \eta)}{1 - P(y_i = 1 | \eta)} = a_i D \eta + D d_i = a_i D (\eta - b_i) \quad (D.9)$$

being a_i , the item discrimination, b_i , the item difficulty, d_i , the intercept, and D , a scaling constant. Solving (9) for $P(y_i = 1 | \eta)$,

$$P(y_i = 1 | \eta) = F(a_i D \eta + D d_i) = F(a_i D (\eta - b_i)) \quad (D.10)$$

where F is logistic cumulative distribution function (CDF) in its pure form when $D=1$, and a very good approximation of the normal CDF when $D=1.702$ as showed by Ayala (2009).

D.5 Conversion formula

Observing that the distribution $x \sim N(\mu_x, \sigma_x^2)$ with threshold τ is equivalent to the distribution $(K_1x + K_2) \sim N(K_1\mu_x + K_2, (K_1\sigma_x)^2)$ with threshold $K_1\tau + K_2$, K_1 and K_2 constants, Kamata and Bauer (2008) conveniently chose $K_1 = \frac{1}{q}$ and $K_2 = -\frac{\lambda_i r}{q}$ to rewrite equation (6) as below. Note the factor term is scaled by simultaneously multiplying and dividing it by s , as to allow its normalization given r and s .

$$\left(\frac{y_i^* - \lambda_i r}{q} \mid \frac{\eta - r}{s}\right) \sim N\left(\frac{v_i}{q} + \frac{\lambda_i s}{q} \left(\frac{\eta - r}{s}\right), \frac{\theta_i}{q^2}\right). \quad (\text{D.11})$$

Accordingly, the threshold that dichotomizes y_i^* must also be adjusted,

$$y_i = \begin{cases} 0 & \text{if } \frac{y_i^* - \lambda_i r}{q} \leq \frac{\tau_i - \lambda_i r}{q} \\ 1 & \text{if } \frac{y_i^* - \lambda_i r}{q} > \frac{\tau_i - \lambda_i r}{q} \end{cases} \quad (\text{D.12})$$

And therefore, the probability of observing $\frac{y_i^* - \lambda_i r}{q} = 1$ given this new scaled common factor is, by analogy with Equation (D.8):

$$P\left(y_i = 1 \mid \frac{\eta - r}{s}\right) = P\left(\frac{y_i^* - \lambda_i r}{q} > \frac{\tau_i - \lambda_i r}{q} \mid \frac{\eta - r}{s}\right) = \phi\left(\frac{v_i - \tau_i + \lambda_i r}{\sqrt{\theta_i}} + \frac{\lambda_i s}{\sqrt{\theta_i}} \left(\frac{\eta - r}{s}\right)\right) \quad (\text{D.13})$$

Making $r = E(\eta) = \alpha$, $s = \text{Var}(\eta)^{\frac{1}{2}} = \sqrt{\psi}$ and $q = \text{Var}(\varepsilon_i)^{\frac{1}{2}} = \sqrt{\theta_i}$, and grouping correspondent terms in equations (D.10) and (D.13), the conversion formula now becomes:

$$d_i = \frac{v_i - \tau_i + \lambda_i \alpha}{D\sqrt{\theta_i}} \quad (\text{D.14})$$

$$a_i = \frac{\lambda_i \sqrt{\psi}}{D \sqrt{\theta_i}} \quad (\text{D.15})$$

D.6 Possible parameterizations

Parameterization of equations (D.4) e (D.5) of the measurement model is necessary to (a) set the origin and scale for the common factor and latent response variables, and (b) to guarantee the system is identified. Note that the problem boils down to decide what needs to be fixed in order to solve a system that relates y_i and $y_i^* \sim N(\mu_i^*, \sigma_i^*)$ with threshold τ_i given the proportions of observed categories 0 and 1.

D.6.1 Scaling the common factor

There are three possible parameterizations to scale the common factor: (i) standardize the factor variance by making $\psi = 1$ in equation (5), (ii) define indicator i as a reference by making $\lambda_i = 1$ in equation (D.4), and (iii) make the factor assume the average scale of its indicators by stating $\sum_1^I \lambda_i = I$ as suggested by Little and Slegers (2006), where I is the number of indicators. Hereafter, these parameterizations are, respectively, named: *factor parameterization*¹⁴, *reference indicator*¹⁵, and *effects coding*.

D.6.2 Setting the scale of the latent response variable¹⁶

As per equation (D.5), the scale of σ^* depends on the factor variance ψ and on its residual variance θ_i . With the *marginal parameterization*, σ_i^* is set to unit, and θ_i is calculated, not estimated, from

$$\theta_j = 1 - \lambda_j^2 \psi \quad (\text{D.16})$$

With the *conditional parameterization*, θ_i is set to unit and σ_i^* is calculated as

¹⁴ Kline (2016) calls it unit variance identification (UVI) constraint, Kamata and Bauer (2008) call it *Standardized Factor* and Little (2013) calls it *fixed factor* method.

¹⁵ Kline (2016) calls it unit loading identification (ULI) constraint, and Little (2013) calls it *marker variable* method.

¹⁶ Marginal and Conditional parameterizations are used in scaling the reference group in the Delta and Theta approaches proposed by Muthén and Asparouhov (2002).

$$\sigma_i^{*2} = \lambda_i^2 \psi + 1 \quad (\text{D.17})$$

This parameterization is particularly useful in longitudinal models where one is interested in research changes on the variances of indicators.

D.6.3 Identification of the mean equation

The origin/mean of the latent variables in the system is defined by the intercept in equation (D.4). First, note that the equation is unidentified. If y_i is continuous and there are I indicators, equation (D.4) introduces I equations with $I + 1$ unknown parameters: I intercepts and the factor mean. One can identify the system by setting the factor mean or one of the intercepts to zero. However, when y_i is categorical, equation (E.4) introduces I equations with $2I + 1$ unknown parameters: I intercepts, I thresholds, and the factor mean. Traditionally, there are three ways to drop $I + 1$ unknowns. First, ν_i and α are fixed to zero as this allows the estimation of I thresholds τ_i given σ_i^{*2} . MPLUS and lavaan adopt this option. Second, LISREL uses the approach of fixing τ_i and α to zero to estimate I intercepts given σ_i^{*2} . And, lastly, ν_i and one of the thresholds are fixed to zero, and the factor mean and $I - 1$ thresholds are estimated given σ_i^{*2} . Instead of setting one of the thresholds to zero, a possible alternative and equivalent constraint is $\sum_1^I \tau_i = 0$ which might be preferable when using the *effects coding* parameterization to determine the common factor scale. In the following experiments, these three parameterizations are called, respectively, *free threshold*, *free intercept*, and *free mean*.

D.6.4 Eighteen parameterization options

TABLE D.1 summarizes identification and scaling choices discussed thus far. Next section contains simulation results showing these parameterizations lead to equivalent statistical models, and estimated parameters with SEM methods will be checked against parameters estimated with IRT ones.

TABLE D.1 Assumptions for all parameterization combinations. k is any integer, usually zero.

		Marginal (MP)	Conditional (CP)
Free Threshold (FT)	Factor Parameterization (FP)	$v_i = 0, \text{all } i, \alpha = 0$ $\psi = 1$ $\sigma_i^* = 1$	$v_i = 0, \text{all } i, \alpha = 0$ $\psi = 1$ $\theta_i = 1$
	Reference Indicator (RI)	$v_i = 0, \text{all } i, \alpha = 0$ $\lambda_i = 1, \text{any single } i$ $\sigma_i^* = 1$	$v_i = 0, \text{all } i, \alpha = 0$ $\lambda_i = 1, \text{any single } i$ $\theta_i = 1$
	Effects Coding (EF)	$v_i = 0, \text{all } i, \alpha = 0$ $\sum_1^I \lambda_i = I$ $\sigma_i^* = 1$	$v_i = 0, \text{all } i, \alpha = 0$ $\sum_1^I \lambda_i = I$ $\theta_i = 1$
Free Intercepts (FI)	Factor Parameterization (FP)	$\tau_i = 0, \text{all } i, \alpha = 0$ $\psi = 1$ $\sigma_i^* = 1$	$\tau_i = 0, \text{all } i, \alpha = 0$ $\psi = 1$ $\theta_i = 1$
	Reference Indicator (RI)	$\tau_i = 0, \text{all } i, \alpha = 0$ $\lambda_i = 1, \text{any single } i$ $\sigma_i^* = 1$	$\tau_i = 0, \text{all } i, \alpha = 0$ $\lambda_i = 1, \text{any single } i$ $\theta_i = 1$
	Effects Coding (EF)	$\tau_i = 0, \text{all } i, \alpha = 0$ $\sum_1^I \lambda_i = I$ $\sigma_i^* = 1$	$\tau_i = 0, \text{all } i, \alpha = 0$ $\sum_1^I \lambda_i = I$ $\theta_i = 1$

Free Mean (FM)	Factor Parameterization (FP)	$v_i = 0, \text{all } i$ $\tau_i = 0, \text{any single } i$ $\psi = 1$ $\sigma_i^* = 1$	$v_i = 0, \text{all } i$ $\tau_i = 0, \text{any single } i$ $\psi = 1$ $\theta_i = 1$
	Reference Indicator (RI)	$v_i = 0, \text{all } i$ $\tau_i = k, \text{any single } j$ $\lambda_i = 1, \text{any single } i$ $\sigma_i^* = 1$	$v_i = 0, \text{all } i$ $\tau_i = k, \text{any single } j$ $\lambda_i = 1, \text{any single } i$ $\theta_i = 1$
	Effects Coding (EF)	$v_i = 0, \text{all } i$ $\sum_1^I \tau_i = k, \sum_1^I \lambda_i = I$ $\sigma_i^* = 1$	$v_i = 0, \text{all } i$ $\sum_1^I \tau_i = k, \sum_1^I \lambda_i = I$ $\theta_i = 1$

D.7 Experiments

A unidimensional latent variable model with 13 indicators was simulated to evaluate the appropriateness of the eighteen parameterizations presented. The IRT parameters of the simulated items is summarized as follows. All items were set to have the same discrimination parameter of 0.7 and difficulty parameters with values ranging from -3 to $+3$, is in such way that this cumulatively covers most of the scale and allows for evaluation of the estimation method limitations at scale extremes – see TABLE D.2 Simulated item parameters. Dichotomous input values were generated assuming a normal factor distribution with mean fixed to zero and fixed unit variance. The parameterization used is indicated by a combination of the abbreviations noted in TABLE D.1 Assumptions for all parameterization combinations. k is any integer, usually zero.. For example, FP_MP_FT indicates that the Factor, Marginal and Free Thresholds parameterizations were used. MIRT package was configured to use the standard expectation-maximization (EM) algorithm with fixed quadrature to estimated item parameters, and the expected a-posteriori (EAP) method to estimate factors. The lavaan package used the diagonally weighted least squares (DWLS) algorithm to estimated indicator parameters and the Empirical Bayes Model (EBM) and for the maximum likelihood (ML) approaches to estimate factors. A sample of 10,000 individuals

were simulated and the results presented are the means of 500 runs. Estimated SEM parameters were used as inputs to equations (D.14) and (D.15) to calculate IRT parameters, a and b , for comparison. Lastly, to promote the reproduction of results and enhance audience reach, all simulations were done using freely distributed and open source MIRT and lavaan R packages both with growing popularity among SEM users. Simulation scripts and results are available from the author upon demand.

TABLE D.2 Simulated item parameters

item	1	2	3	4	5	6	7	8	9	10	11	12	13
a	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
b	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	2.0	2.5	3.0

D.7.1 Simulation results

TABLE D.3 shows global fit statistics for the simulated models. Values suggest appropriate fit and confirm the theory that such values should not be affected by the parameterization used. TABLE D.5 also confirms the theory that, except for the case when residual variances are fixed to unit, they should not vary with the chosen parameterization.

TABLE D.6 and TABLE D.7 contain the calculated values for parameters a and b . The best way to navigate through the results it to first focus on the results of item 7 at the center of the scale where most of the test information is concentrated and optimal results are expected and, then, compare its results with items located in positions further away from the mean. In TABLE D.6, the simulated discrimination parameter was fixed to 0.7. First evident result is that observing item 7, the EM algorithm returns an estimation closer to 0.7 than any other SEM model using DWLS. Further, this difference is exacerbated as the item location moves away from the factor mean, probably caused because there less information to calculate the covariances associated with items as they move in the direction of the scale extremes. The same trend can be observed when estimating difficulty parameters in TABLE D.7 with standard error becoming in the order of the decimals for measurements done with over two standard deviations from the mean.

TABLE D.4 lists the estimated values for α , and ψ for all parameterization options, note that α values are non-zero, under the “FREE MEAN” parameterization, when

“REFERENCE INDICATOR” is used, or when “FACTOR PARAMETERIZATION” or “EFFECTS CODING” is used with $k \neq 0$ – not shown. With $k = 0$, α is always zero.

D.7.2 Conclusions

This paper summarizes all known parameterization options available in the SEM literature and revisits the conversion formula proposed by Kamata and Bauer so that it can be used to transform SEM parameters into IRT ones irrespective of the chosen parameterization. It is our hope that this paper eases comparison between models using different parameterizations and increases awareness about the pros and cons of such choices.

D.8 Acknowledgement

This work was conducted at the Institute of Mathematics and Statistics, University of Sao Paulo (IME-USP) with financial support from Adaptativa Inteligência Educacional Ltda. and Ânima Educação S.A.

D.9 References

Ayala, R.J. (2009). *The Theory and Practice of Item Response Theory*. New York. The Guilford Press.

Chalmers RP (2012). *mirt: A Multidimensional Item Response Theory Package for the R Environment*. *Journal of Statistical Software*, 48(6), 1–29.

Kamata, A. Bauer, D.J. (2008). A Note on the Relation Between Factor Analytic and Item Response Theory Models. *Structural Equation Modeling*, 15, 136-153.

Kline, R.B. (2016). *Principles and Practice of Structural Equation Modeling*. New York. The Guilford Press.

Little, T.D. (2013). *Longitudinal Structural Equation Modeling*. New York. The Guilford Press.

Little, T.D., Slegers, D.W., Card, N.A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling*, 13, 59-72.

McDonald, R. P. (1999). *Test theory: Unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, 39, 479–515.

Muthén, B. O., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus (Mplus Web Note No. 4).

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52, 393–408.

Rosseel Y (2012). “lavaan: An R Package for Structural Equation Modeling.” *Journal of Statistical Software*, 48 (2), 1–36

TABLE D.3 Model fit statistics across multiple parameterization options when $\eta \sim N(0,1)$ and $k = 0$.

PARAMETERIZATION	TLI	CFI	RMSEA
MIRT	1.000	0.999	0.002
FP_MP_FT	1.002	1.000	0.001
FP_MP_FI	1.002	1.000	0.001
FP_MP_FM	1.002	1.000	0.001
FP_CP_FT	1.002	1.000	0.001
FP_CP_FI	1.002	1.000	0.001
FP_CP_FM	1.002	1.000	0.001
RI_MP_FT	1.002	1.000	0.001
RI_MP_FI	1.002	1.000	0.001
RI_MP_FM	1.002	1.000	0.001
RI_CP_FT	1.002	1.000	0.001
RI_CP_FI	1.002	1.000	0.001
RI_CP_FM	1.002	1.000	0.001
EC_MP_FT	1.002	1.000	0.001
EC_MP_FI	1.002	1.000	0.001
EC_MP_FM	1.002	1.000	0.001
EC_CP_FT	1.002	1.000	0.001
EC_CP_FI	1.002	1.000	0.001
EC_CP_FM	1.002	1.000	0.000

TABLE D.4 Simulation results for factor mean α and ψ . $\eta \sim N(0,1)$. Mean of 500 repetitions with standard deviations shown in parenthesis. $k = 0$.

	ψ	α
FP_MP_FT	1.000 (0.000)	0.000 (0.000)
FP_MP_FI	1.000 (0.000)	0.000 (0.000)
FP_MP_FM	1.000 (0.000)	0.001 (0.031)
FP_CP_FT	1.000 (0.000)	0.000 (0.000)
FP_CP_FI	1.000 (0.000)	0.000 (0.000)
FP_CP_FM	1.000 (0.000)	0.000 (0.031)
RI_MP_FT	0.147 (0.013)	0.000 (0.000)
RI_MP_FI	0.146 (0.013)	0.000 (0.000)
RI_MP_FM	0.147 (0.013)	0.587 (0.014)
RI_CP_FT	0.172 (0.019)	0.000 (0.000)
RI_CP_FI	0.172 (0.018)	0.000 (0.000)
RI_CP_FM	0.172 (0.018)	0.635 (0.015)
EC_MP_FT	0.143 (0.004)	0.000 (0.000)
EC_MP_FI	0.142 (0.004)	0.000 (0.000)
EC_MP_FM	0.143 (0.004)	0.000 (0.005)

Apêndice D - Parameterization effects on the Relation between Structural Equation and Item Response Models using binary indicators - References

EC_CP_FT	0.167 (0.005)	0.000 (0.000)
EC_CP_FI	0.167 (0.005)	0.000 (0.000)
EC_CP_FM	0.167 (0.005)	-0.001 (0.006)

TABLE D.5 Item random error variance (θ_i) across multiple parameterization options when $\eta \sim N(0,1)$. $k = 0$.

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}
FP_MP_FT	0.877 (0.015)	0.867 (0.014)	0.861 (0.014)	0.854 (0.014)	0.848 (0.013)	0.844 (0.014)	0.843 (0.013)	0.845 (0.013)	0.847 (0.013)	0.854 (0.013)	0.860 (0.014)	0.869 (0.014)	0.875 (0.014)
FP_MP_FI	0.875 (0.016)	0.868 (0.014)	0.859 (0.014)	0.853 (0.014)	0.848 (0.014)	0.845 (0.013)	0.843 (0.013)	0.845 (0.013)	0.848 (0.013)	0.852 (0.013)	0.860 (0.014)	0.868 (0.015)	0.876 (0.015)
FP_MP_FM	0.877 (0.015)	0.867 (0.014)	0.861 (0.014)	0.854 (0.014)	0.848 (0.013)	0.844 (0.014)	0.843 (0.013)	0.845 (0.013)	0.847 (0.013)	0.854 (0.013)	0.860 (0.014)	0.869 (0.014)	0.875 (0.014)
FP_CP_FT	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
FP_CP_FI	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
FP_CP_FM	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)

TABLE D.5 Item random error variance (θ_i) across multiple parameterization options when $\eta \sim N(0,1)$. $k = 0$ – continued.

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}
RI_MP_FT	0.875 (0.015)	0.867 (0.015)	0.860 (0.013)	0.853 (0.013)	0.848 (0.013)	0.845 (0.012)	0.844 (0.013)	0.845 (0.012)	0.847 (0.013)	0.853 (0.013)	0.860 (0.013)	0.867 (0.014)	0.876 (0.015)
RI_MP_FI	0.878 (0.015)	0.868 (0.014)	0.859 (0.013)	0.854 (0.013)	0.849 (0.013)	0.844 (0.013)	0.845 (0.014)	0.844 (0.013)	0.847 (0.013)	0.853 (0.014)	0.860 (0.014)	0.867 (0.014)	0.875 (0.015)
RI_MP_FM	0.876 (0.015)	0.868 (0.014)	0.858 (0.014)	0.853 (0.013)	0.848 (0.013)	0.844 (0.014)	0.844 (0.013)	0.844 (0.013)	0.848 (0.014)	0.853 (0.014)	0.859 (0.014)	0.866 (0.015)	0.876 (0.015)
RI_CP_FT	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
RI_CP_FI	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
RI_CP_FM	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
EC_MP_FT	0.875 (0.015)	0.866 (0.015)	0.859 (0.014)	0.853 (0.013)	0.848 (0.014)	0.845 (0.013)	0.843 (0.012)	0.845 (0.013)	0.848 (0.013)	0.853 (0.014)	0.859 (0.014)	0.868 (0.014)	0.875 (0.016)
EC_MP_FI	0.876 (0.015)	0.868 (0.014)	0.859 (0.014)	0.854 (0.013)	0.847 (0.014)	0.845 (0.013)	0.844 (0.012)	0.845 (0.013)	0.848 (0.013)	0.853 (0.013)	0.860 (0.015)	0.868 (0.014)	0.876 (0.016)
EC_MP_FM	0.876 (0.015)	0.867 (0.015)	0.859 (0.013)	0.853 (0.013)	0.847 (0.013)	0.844 (0.013)	0.844 (0.013)	0.844 (0.013)	0.848 (0.013)	0.852 (0.014)	0.859 (0.014)	0.868 (0.014)	0.876 (0.014)
EC_CP_FT	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)

Apêndice D - Parameterization effects on the Relation between Structural Equation and Item Response Models using binary indicators - References

EC_CP_FI	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
EC_CP_FM	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)

TABLE D.6 Discrimination parameter a_j across multiple parameterization options. $\eta \sim N(0,1)$. Mean of 500 repetitions with standard deviations shown in parenthesis. SIM shows simulated results. IRT indicates IRT software estimations. $k = 0$.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}
SIM	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
IRT	0.698 (0.046)	0.702 (0.043)	0.701 (0.041)	0.701 (0.038)	0.701 (0.039)	0.703 (0.036)	0.701 (0.034)	0.700 (0.037)	0.699 (0.037)	0.700 (0.039)	0.699 (0.043)	0.700 (0.045)	0.700 (0.050)
FP_MP_FT	0.636 (0.044)	0.665 (0.042)	0.683 (0.041)	0.703 (0.038)	0.720 (0.037)	0.731 (0.038)	0.732 (0.036)	0.728 (0.037)	0.721 (0.038)	0.702 (0.037)	0.685 (0.041)	0.660 (0.042)	0.640 (0.042)
FP_MP_FI	0.642 (0.046)	0.663 (0.041)	0.687 (0.039)	0.705 (0.039)	0.720 (0.038)	0.727 (0.036)	0.733 (0.037)	0.727 (0.037)	0.719 (0.036)	0.707 (0.038)	0.686 (0.040)	0.663 (0.043)	0.640 (0.044)
FP_MP_FM	0.636 (0.044)	0.665 (0.042)	0.683 (0.041)	0.703 (0.038)	0.720 (0.037)	0.731 (0.038)	0.732 (0.036)	0.728 (0.037)	0.721 (0.038)	0.702 (0.037)	0.685 (0.041)	0.660 (0.042)	0.640 (0.042)
FP_CP_FT	0.638 (0.045)	0.666 (0.045)	0.686 (0.040)	0.708 (0.038)	0.721 (0.035)	0.726 (0.039)	0.733 (0.037)	0.728 (0.036)	0.721 (0.037)	0.704 (0.037)	0.687 (0.040)	0.666 (0.041)	0.642 (0.044)
FP_CP_FI	0.643 (0.048)	0.667 (0.044)	0.687 (0.040)	0.703 (0.039)	0.720 (0.038)	0.730 (0.037)	0.733 (0.037)	0.728 (0.036)	0.724 (0.037)	0.702 (0.038)	0.686 (0.039)	0.665 (0.042)	0.642 (0.045)
FP_CP_FM	0.642 (0.046)	0.663 (0.041)	0.687 (0.039)	0.705 (0.039)	0.720 (0.038)	0.727 (0.036)	0.733 (0.037)	0.727 (0.037)	0.719 (0.036)	0.707 (0.038)	0.686 (0.040)	0.663 (0.043)	0.640 (0.044)

TABLE D.6 Discrimination parameter a_i across multiple parameterization options. $\eta \sim N(0,1)$. Mean of 500 repetitions with standard deviations shown in parenthesis. SIM shows simulated results. IRT indicates IRT software estimations. $k = 0$ – continued.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}
SIM	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
RI_MP_FT	0.641 (0.043)	0.665 (0.044)	0.686 (0.038)	0.704 (0.038)	0.719 (0.037)	0.729 (0.034)	0.732 (0.036)	0.727 (0.035)	0.721 (0.037)	0.705 (0.036)	0.686 (0.037)	0.664 (0.041)	0.639 (0.043)
RI_MP_FI	0.634 (0.045)	0.663 (0.040)	0.687 (0.038)	0.704 (0.037)	0.717 (0.038)	0.729 (0.037)	0.729 (0.038)	0.732 (0.037)	0.722 (0.036)	0.706 (0.038)	0.687 (0.041)	0.666 (0.041)	0.642 (0.046)
RI_MP_FM	0.639 (0.044)	0.663 (0.041)	0.691 (0.039)	0.706 (0.037)	0.720 (0.036)	0.731 (0.038)	0.730 (0.037)	0.730 (0.037)	0.721 (0.038)	0.706 (0.039)	0.687 (0.039)	0.667 (0.042)	0.639 (0.045)
RI_CP_FT	0.644 (0.046)	0.664 (0.041)	0.689 (0.041)	0.705 (0.039)	0.718 (0.037)	0.732 (0.036)	0.732 (0.037)	0.729 (0.036)	0.718 (0.036)	0.704 (0.039)	0.686 (0.039)	0.667 (0.042)	0.642 (0.044)
RI_CP_FI	0.640 (0.044)	0.662 (0.043)	0.687 (0.041)	0.704 (0.037)	0.719 (0.036)	0.730 (0.036)	0.735 (0.037)	0.730 (0.037)	0.719 (0.039)	0.709 (0.038)	0.688 (0.040)	0.663 (0.042)	0.643 (0.046)
RI_CP_FM	0.639 (0.041)	0.663 (0.045)	0.687 (0.040)	0.705 (0.037)	0.723 (0.037)	0.730 (0.034)	0.733 (0.035)	0.731 (0.036)	0.720 (0.037)	0.707 (0.039)	0.685 (0.039)	0.666 (0.042)	0.641 (0.043)
EC_MP_FT	0.642 (0.043)	0.669 (0.042)	0.687 (0.041)	0.705 (0.036)	0.720 (0.038)	0.728 (0.037)	0.732 (0.035)	0.729 (0.037)	0.719 (0.037)	0.705 (0.039)	0.688 (0.039)	0.663 (0.042)	0.641 (0.046)
EC_MP_FI	0.640 (0.045)	0.662 (0.041)	0.689 (0.039)	0.703 (0.037)	0.722 (0.038)	0.728 (0.037)	0.731 (0.034)	0.729 (0.035)	0.720 (0.037)	0.705 (0.037)	0.685 (0.042)	0.663 (0.041)	0.640 (0.046)
EC_MP_FM	0.638 (0.045)	0.665 (0.043)	0.688 (0.037)	0.704 (0.037)	0.723 (0.036)	0.731 (0.035)	0.730 (0.036)	0.730 (0.036)	0.720 (0.037)	0.707 (0.038)	0.688 (0.040)	0.663 (0.040)	0.639 (0.042)

Apêndice D - Parameterization effects on the Relation between Structural Equation and Item Response Models using binary indicators - References

EC_CP_FT	0.637 (0.044)	0.663 (0.041)	0.684 (0.040)	0.704 (0.039)	0.722 (0.036)	0.728 (0.037)	0.731 (0.036)	0.731 (0.035)	0.718 (0.038)	0.708 (0.041)	0.685 (0.040)	0.668 (0.041)	0.639 (0.045)
EC_CP_FI	0.640 (0.042)	0.663 (0.040)	0.687 (0.040)	0.707 (0.040)	0.719 (0.037)	0.728 (0.037)	0.735 (0.037)	0.728 (0.037)	0.718 (0.036)	0.706 (0.036)	0.687 (0.039)	0.664 (0.041)	0.640 (0.046)
EC_CP_FM	0.638 (0.046)	0.667 (0.045)	0.686 (0.039)	0.706 (0.037)	0.721 (0.039)	0.726 (0.036)	0.730 (0.036)	0.732 (0.037)	0.722 (0.039)	0.708 (0.038)	0.687 (0.039)	0.667 (0.042)	0.639 (0.044)

Item	b	MIRT	FP_MP_FI	FP_MP_FI	FP_CP_FT	FP_CP_FI	RI_MP_FT	RI_MP_FI	RI_CP_FT	RI_CP_FI	EC_MP_FI	EC_MP_FI	EC_CP_FT	EC_CP_FI	FP_MP_FI	FP_CP_FI	RI_MP_FI	RI_CP_FI	EC_MP_FI	EC_CP_FI
1	-3	0.0463	0.0437	0.0457	0.0448	0.0478	0.0427	0.0450	0.0462	0.0443	0.0430	0.0451	0.0442	0.0421	0.0437	0.0457	0.0444	0.0415	0.0446	0.0456
2	-2.5	0.0431	0.0417	0.0412	0.0446	0.0440	0.0444	0.0401	0.0408	0.0430	0.0424	0.0413	0.0406	0.0402	0.0417	0.0412	0.0411	0.0453	0.0432	0.0455
3	-2	0.0412	0.0413	0.0392	0.0397	0.0399	0.0380	0.0384	0.0412	0.0410	0.0408	0.0395	0.0396	0.0401	0.0413	0.0392	0.0385	0.0396	0.0365	0.0392
4	-1.5	0.0378	0.0384	0.0393	0.0381	0.0388	0.0381	0.0367	0.0386	0.0373	0.0363	0.0369	0.0393	0.0400	0.0384	0.0393	0.0374	0.0373	0.0367	0.0373
5	-1	0.0388	0.0366	0.0381	0.0354	0.0383	0.0370	0.0377	0.0366	0.0356	0.0382	0.0383	0.0363	0.0371	0.0366	0.0381	0.0355	0.0367	0.0364	0.0391
6	-0.5	0.0355	0.0381	0.0364	0.0392	0.0368	0.0340	0.0372	0.0360	0.0359	0.0373	0.0370	0.0374	0.0368	0.0381	0.0364	0.0380	0.0341	0.0351	0.0361
7	0	0.0344	0.0355	0.0374	0.0368	0.0370	0.0364	0.0377	0.0373	0.0368	0.0347	0.0343	0.0359	0.0373	0.0355	0.0374	0.0373	0.0349	0.0365	0.0365
8	0.5	0.0370	0.0374	0.0368	0.0357	0.0364	0.0346	0.0367	0.0364	0.0370	0.0371	0.0354	0.0354	0.0365	0.0374	0.0368	0.0371	0.0359	0.0355	0.0374
9	1	0.0374	0.0375	0.0356	0.0368	0.0375	0.0369	0.0358	0.0360	0.0393	0.0375	0.0370	0.0378	0.0360	0.0375	0.0356	0.0379	0.0372	0.0373	0.0387
10	1.5	0.0385	0.0372	0.0380	0.0375	0.0378	0.0359	0.0380	0.0393	0.0382	0.0390	0.0374	0.0409	0.0364	0.0372	0.0380	0.0389	0.0389	0.0381	0.0382
11	2	0.0433	0.0409	0.0397	0.0401	0.0391	0.0370	0.0408	0.0395	0.0404	0.0391	0.0416	0.0399	0.0386	0.0409	0.0397	0.0388	0.0393	0.0396	0.0392
12	2.5	0.0449	0.0419	0.0432	0.0406	0.0420	0.0408	0.0410	0.0419	0.0422	0.0418	0.0410	0.0414	0.0414	0.0419	0.0432	0.0421	0.0420	0.0402	0.0419
13	3	0.0497	0.0424	0.0437	0.0436	0.0451	0.0434	0.0456	0.0436	0.0464	0.0456	0.0460	0.0448	0.0463	0.0424	0.0437	0.0449	0.0431	0.0415	0.0439

Figure D-2 Heat map chart of the standard errors for discrimination parameter a_j across multiple parameterization options.

TABLE D.7 Difficulty parameter b_i across multiple parameterization options. $\eta \sim N(0,1)$. Mean of 500 repetitions with standard deviations shown in parenthesis. SIM shows simulated results. IRT indicates IRT software estimations. $k = 0$.

	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}
SIM	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3
IRT	-3.024 (0.178)	-2.501 (0.136)	-2.004 (0.107)	-1.503 (0.078)	-1.003 (0.057)	-0.499 (0.037)	-0.002 (0.031)	0.504 (0.040)	1.004 (0.056)	1.507 (0.079)	2.009 (0.114)	2.510 (0.141)	3.012 (0.188)
FP_MP_FT	-3.266 (0.205)	-2.645 (0.158)	-2.089 (0.117)	-1.540 (0.080)	-1.011 (0.056)	-0.500 (0.041)	-0.001 (0.031)	0.504 (0.040)	1.008 (0.057)	1.538 (0.078)	2.084 (0.116)	2.664 (0.158)	3.247 (0.196)
FP_MP_FI	-3.245 (0.211)	-2.656 (0.152)	-2.076 (0.111)	-1.534 (0.083)	-1.011 (0.055)	-0.503 (0.038)	0.000 (0.031)	0.501 (0.038)	1.014 (0.057)	1.532 (0.080)	2.078 (0.111)	2.656 (0.155)	3.250 (0.203)
FP_MP_FM	-3.266 (0.209)	-2.645 (0.157)	-2.088 (0.119)	-1.540 (0.084)	-1.010 (0.062)	-0.499 (0.050)	0.000 (0.000)	0.505 (0.047)	1.009 (0.062)	1.538 (0.083)	2.085 (0.120)	2.664 (0.162)	3.247 (0.199)
FP_CP_FT	-3.257 (0.205)	-2.643 (0.162)	-2.080 (0.115)	-1.530 (0.080)	-1.007 (0.054)	-0.504 (0.041)	0.000 (0.031)	0.503 (0.038)	1.009 (0.057)	1.537 (0.079)	2.079 (0.116)	2.642 (0.150)	3.244 (0.207)
FP_CP_FI	-3.239 (0.219)	-2.641 (0.163)	-2.077 (0.116)	-1.541 (0.084)	-1.009 (0.060)	-0.502 (0.039)	-0.002 (0.031)	0.503 (0.038)	1.009 (0.056)	1.541 (0.079)	2.080 (0.114)	2.645 (0.153)	3.243 (0.211)
FP_CP_FM	-3.245 (0.213)	-2.655 (0.152)	-2.076 (0.112)	-1.534 (0.087)	-1.011 (0.062)	-0.503 (0.046)	0.000 (0.000)	0.501 (0.045)	1.014 (0.062)	1.533 (0.085)	2.078 (0.113)	2.656 (0.157)	3.250 (0.205)

TABLE D.7 Difficulty parameter b_i across multiple parameterization options. $\eta \sim N(0,1)$. Mean of 500 repetitions with standard deviations shown in parenthesis. SIM shows simulated results. IRT indicates IRT software estimations. $k = 0$ – continued.

	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}
SIM	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3
RI_MP_FT	-3.246 (0.199)	-2.651 (0.163)	-2.082 (0.108)	-1.537 (0.083)	-1.015 (0.058)	-0.502 (0.038)	0.001 (0.030)	0.500 (0.036)	1.008 (0.057)	1.535 (0.077)	2.078 (0.108)	2.649 (0.149)	3.257 (0.208)
RI_MP_FI	-3.281 (0.210)	-2.653 (0.146)	-2.078 (0.109)	-1.538 (0.079)	-1.014 (0.061)	-0.500 (0.037)	-0.001 (0.031)	0.499 (0.038)	1.009 (0.055)	1.535 (0.079)	2.078 (0.114)	2.644 (0.150)	3.243 (0.212)
RI_MP_FM	-3.250 (0.211)	-2.653 (0.152)	-2.069 (0.107)	-1.535 (0.078)	-1.010 (0.055)	-0.502 (0.037)	0.001 (0.033)	0.501 (0.038)	1.011 (0.058)	1.532 (0.082)	2.083 (0.114)	2.641 (0.153)	3.260 (0.208)
RI_CP_FT	-3.235 (0.210)	-2.648 (0.148)	-2.073 (0.117)	-1.536 (0.079)	-1.015 (0.059)	-0.497 (0.038)	-0.001 (0.032)	0.504 (0.041)	1.015 (0.053)	1.535 (0.083)	2.081 (0.112)	2.643 (0.145)	3.241 (0.204)
RI_CP_FI	-3.247 (0.208)	-2.657 (0.155)	-2.076 (0.114)	-1.535 (0.082)	-1.014 (0.054)	-0.500 (0.037)	-0.001 (0.030)	0.502 (0.039)	1.012 (0.056)	1.529 (0.080)	2.076 (0.113)	2.656 (0.156)	3.242 (0.211)
RI_CP_FM	-3.253 (0.192)	-2.657 (0.167)	-2.077 (0.109)	-1.536 (0.079)	-1.008 (0.053)	-0.501 (0.038)	0.001 (0.031)	0.498 (0.038)	1.012 (0.056)	1.529 (0.083)	2.085 (0.114)	2.641 (0.152)	3.245 (0.201)
EC_MP_FT	-3.239 (0.197)	-2.636 (0.154)	-2.076 (0.116)	-1.534 (0.077)	-1.009 (0.057)	-0.503 (0.038)	0.001 (0.033)	0.498 (0.039)	1.010 (0.057)	1.537 (0.082)	2.075 (0.109)	2.654 (0.155)	3.249 (0.215)
EC_MP_FI	-3.253 (0.211)	-2.657 (0.151)	-2.072 (0.111)	-1.536 (0.076)	-1.006 (0.055)	-0.500 (0.039)	0.001 (0.033)	0.501 (0.038)	1.011 (0.057)	1.538 (0.078)	2.083 (0.117)	2.656 (0.151)	3.253 (0.222)
EC_MP_FM	-3.262 (0.212)	-2.650 (0.161)	-2.070 (0.103)	-1.534 (0.078)	-1.006 (0.054)	-0.500 (0.038)	0.001 (0.031)	0.503 (0.036)	1.013 (0.058)	1.531 (0.082)	2.077 (0.111)	2.653 (0.148)	3.255 (0.196)

EC_CP_FT	-3.265 (0.210)	-2.652 (0.147)	-2.088 (0.112)	-1.538 (0.081)	-1.009 (0.054)	-0.505 (0.039)	-0.002 (0.033)	0.499 (0.038)	1.011 (0.055)	1.531 (0.086)	2.084 (0.116)	2.633 (0.155)	3.251 (0.214)
EC_CP_FI	-3.251 (0.192)	-2.652 (0.150)	-2.079 (0.115)	-1.533 (0.083)	-1.013 (0.055)	-0.503 (0.037)	-0.001 (0.032)	0.504 (0.039)	1.014 (0.055)	1.531 (0.078)	2.078 (0.114)	2.649 (0.155)	3.251 (0.217)
EC_CP_FM	-3.257 (0.212)	-2.641 (0.168)	-2.080 (0.110)	-1.530 (0.081)	-1.008 (0.061)	-0.503 (0.040)	0.002 (0.032)	0.500 (0.039)	1.007 (0.060)	1.534 (0.081)	2.077 (0.112)	2.642 (0.155)	3.256 (0.206)

IND	SIM	MIRT	FP_MP_FI	FP_MP_FI	FP_CP_FT	FP_CP_FI	RI_MP_FT	RI_MP_FI	RI_CP_FT	RI_CP_FI	EC_MP_FI	EC_MP_FI	EC_CP_FT	EC_CP_FI	FP_MP_FI	FP_CP_FM	RI_MP_FM	RI_CP_FM	EC_MP_FI	EC_CP_FM
1	-3	0.177502	0.205321	0.211137	0.205276	0.219223	0.199175	0.210225	0.210001	0.208486	0.196967	0.210836	0.209577	0.192327	0.208786	0.212878	0.210969	0.192018	0.211873	0.212315
2	-2.5	0.135787	0.157717	0.151514	0.161858	0.163384	0.162877	0.146212	0.14839	0.155198	0.154438	0.151368	0.146667	0.150377	0.156804	0.152146	0.15225	0.16724	0.160635	0.167899
3	-2	0.10713	0.117257	0.110561	0.114727	0.115888	0.108042	0.108509	0.117014	0.113561	0.116017	0.110505	0.112163	0.114727	0.119347	0.112412	0.106591	0.109466	0.103058	0.109755
4	-1.5	0.078395	0.080345	0.08274	0.080243	0.083935	0.083184	0.07866	0.079323	0.081555	0.076958	0.07629	0.080616	0.083325	0.084309	0.087096	0.078098	0.078692	0.078104	0.080902
5	-1	0.057047	0.055696	0.05547	0.053905	0.059547	0.05761	0.060599	0.058522	0.05414	0.056565	0.054929	0.054176	0.055428	0.061989	0.061821	0.055018	0.052747	0.054048	0.061132
6	-0.5	0.037127	0.040755	0.038401	0.040854	0.038893	0.037964	0.036843	0.038014	0.037359	0.037823	0.039005	0.038604	0.037079	0.050106	0.046498	0.037322	0.037565	0.037734	0.039968
7	0	0.031498	0.030523	0.030507	0.031135	0.030847	0.029806	0.03062	0.032293	0.029988	0.032522	0.032771	0.032718	0.03191	2.54E-08	0	0.033368	0.031124	0.030996	0.032337
8	0.5	0.040238	0.040062	0.037884	0.037627	0.037837	0.036197	0.037617	0.040704	0.038863	0.039172	0.037754	0.038043	0.038982	0.046954	0.045116	0.038374	0.038223	0.036399	0.038595
9	1	0.055868	0.056574	0.057096	0.056835	0.056445	0.056942	0.054777	0.053094	0.055673	0.057233	0.057189	0.054533	0.054689	0.062489	0.061992	0.05767	0.05553	0.057561	0.060175
10	1.5	0.07927	0.078154	0.080092	0.078627	0.078535	0.076943	0.079435	0.083452	0.079588	0.082066	0.077936	0.086306	0.077554	0.08304	0.084791	0.082017	0.083347	0.082264	0.081427
11	2	0.113776	0.115547	0.111003	0.116179	0.113941	0.108035	0.113823	0.112424	0.113094	0.109262	0.116963	0.116321	0.114011	0.119578	0.113071	0.11441	0.113615	0.110899	0.111622
12	2.5	0.140966	0.158359	0.155085	0.150066	0.15325	0.149093	0.150393	0.145038	0.155857	0.155246	0.150584	0.155425	0.155365	0.162095	0.156817	0.153248	0.152356	0.148241	0.154996
13	3	0.187582	0.195662	0.203176	0.206635	0.210535	0.207815	0.212323	0.204034	0.211054	0.214733	0.222443	0.214334	0.217418	0.199101	0.204903	0.208445	0.200881	0.196299	0.205799

Figure D-3 Heat map chart of the standard errors for difficulty parameter b_i across multiple parameterization options.

Apêndice E Autorização para uso de dados

O documento a seguir contém a autorização, pelo Secretário Municipal de Educação José Roberto Martins Aguiar, para o uso neste trabalho dos dados coletados durante as avaliações educacionais efetuadas no **Programa Especial de Aumento de Rendimento Escolar** no âmbito do **Projeto Escola +** executado no município de Cariacica em 2019.

Prefeitura Municipal de Cariacica
Secretaria Municipal de Educação de Cariacica

Requerimento

Caro responsável,

Sou Hugo Harada, estudante de doutorado do Instituto de Matemática e Estatística da Universidade de São Paulo (IME/USP), e venho por meio deste solicitar o compartilhamento de dados das avaliações educacionais efetuadas no **Programa Especial de Aumento de Rendimento Escolar** no âmbito do **Projeto Escola+**, executado no município de Cariacica em 2019, para utilizá-los em um exemplo de aplicação dos modelos estatísticos pesquisados em minha tese intitulada: "*Equivalência da Análise Longitudinal de Dados Multidimensionais através da Teoria da Resposta ao Item e de Modelos de Equações Estruturais*".

O uso do material terá fim estritamente acadêmico e a identidade das partes envolvidas nas avaliações será preservada devido a um processo de anonimização dos dados. Não será possível identificar indivíduos nem a instituição de ensino a qual eles foram associados.

Estou disponível para quaisquer esclarecimentos e aguardo ansioso sua resposta.

Atenciosamente,



Hugo Kenji Pereira Harada
CPF:429.541.932-04

*Autorizada
Em 30/11/22*



José Roberto Martins Aguiar
Secretário Municipal
de Educação
Mat 195657 Port. CP 9912894

Figura E.1 – Autorização para o uso de dados do **Programa Especial de Aumento de Rendimento Escolar / Programa Escola +** – Município de Cariacica/ES.