



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

# **Uma proposta para controle da taxa de exposição de itens em testes adaptativos computadorizados**

por

**Lídia Almeida de Carvalho**

Orientador: Prof. Dr. Antônio Eduardo Gomes

**Novembro de 2022**

# **Uma proposta para controle da taxa de exposição de itens em testes adaptativos computadorizados**

**por**

**Lídia Almeida de Carvalho**

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Comissão Julgadora:

- Prof. Dr. Antonio Eduardo Gomes (orientador) - EST/UnB
- Prof. Dr. Raul Yukihiro Matsushita (membro interno) - EST/UnB
- Prof. Dr. Guilherme Souza Rodrigues (membro interno) - EST/UnB
- Prof. Dr. Caio Lucidius Naberezny Azevedo (membro externo) - IMECC/Unicamp

**Universidade de Brasília**

**Brasília, Novembro de 2022**

*"Sem dados você é apenas mais uma pessoa com uma opinião."  
William Edwards Deming*

# Agradecimentos

Agradeço a minha vó Marceonilia Almeida, in memoriam, meu maior exemplo de força, garra e conhecimento, quem me ensinou que o único caminho é a educação, bem como a minha mãe, que não mediu esforços para me proporcionar uma educação de qualidade e sempre esteve ao meu lado em todas as minhas decisões.

A minha família, que me deu suporte e compartilhou desta trajetória em tempos tão difíceis de pandemia. Em especial a Alice Amarante, por estar sempre do meu lado em todos os momentos e nunca soltar a minha mão. A todos os meus amigos que se fizeram presentes e entenderam meus momentos de ausência.

Agradeço ao meu orientador Professor Antônio Eduardo, por toda dedicação e disponibilidade durante este processo. Aos amigos e colegas de profissão que me deram suporte sempre que possível, especialmente Prof. Dr. Antônio dos Santos.

Esta página não é suficiente para agradecer à todos que compartilharam desta etapa junto comigo e contribuíram de forma direta e indireta para a conclusão deste trabalho. Sem jamais esquecer aqueles que se fizeram presente no início desta trajetória e partiram brevemente, sem a oportunidade de partilhar este momento. Minhas eternas saudades.

Agradeço a universidade pública pelo ensino gratuito e de qualidade, proporcionando a formação de diversos profissionais. Esta pesquisa foi realizada com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

# Sumário

<b>Introdução</b>	<b>12</b>
<b>1 Referencial Teórico</b>	<b>14</b>
1.1 Teoria da Resposta ao Item . . . . .	14
1.1.1 Curva Característica do Item . . . . .	15
1.1.2 Função de Informação . . . . .	16
1.1.3 Escala de Habilidade . . . . .	17
1.1.4 Suposições do Modelo . . . . .	17
1.1.5 Calibração dos Itens e Estimação do Traço Latente . . . . .	18
1.2 Testes Adaptativos Computadorizados . . . . .	18
1.2.1 Estimativa do Traço Latente . . . . .	20
1.2.2 Critério de Seleção de Itens . . . . .	22
1.2.3 Métodos de Seleção de Itens . . . . .	25
1.2.4 Critério de Parada do Teste . . . . .	27
1.3 Considerações . . . . .	27
<b>2 Metodologia Proposta</b>	<b>29</b>
2.1 Medidas de desempenho . . . . .	31
2.1.1 Fidelidade . . . . .	31
2.1.2 Vício . . . . .	31
2.1.3 Erro Quadrático Médio . . . . .	31
2.1.4 Número de itens superexpostos . . . . .	32
2.1.5 Número de itens subutilizados . . . . .	32
2.1.6 Taxa de sobreposição . . . . .	32
2.2 Simulações e Aplicações . . . . .	32
2.2.1 Banco de itens simulado . . . . .	33
2.2.2 Aplicação dos métodos a um banco de itens real . . . . .	44
<b>3 Conclusão e Propostas</b>	<b>55</b>
<b>Referências</b>	<b>56</b>
<b>Apêndice A</b>	<b>60</b>

# Lista de Figuras

1.1	Curva Característica do Item . . . . .	15
1.2	Fluxograma do CAT . . . . .	19
2.1	Taxa de exposição dos itens para todos os métodos . . . . .	41
2.2	Proficiência real vs Proficiência estimada . . . . .	42
2.3	Proporção de respondentes expostos a cada item vs Parâmetro de dificuldade . . . . .	42
2.4	Proporção de respondentes expostos a cada item vs Parâmetro de discriminação . . . . .	43
2.5	Diferença entre as proficiências vs Número de itens coincidentes . . . . .	44
2.6	Proficiência real vs Proficiência estimada . . . . .	52
2.7	Diferença entre as proficiências vs Número de itens coincidentes . . . . .	53

# Lista de Tabelas

2.1	Tabela resumo dos métodos comparados . . . . .	33
2.2	Resultados obtidos pelos métodos Sympson-Hetter, Máxima Informação e Randomesque . . . . .	36
2.3	Resultados obtidos pelo método da Máxima Informação Ponderada . . . . .	37
2.4	Resultados obtidos pelo método da Máxima Informação Ponderada . . . . .	38
2.5	Resultados obtidos pelo método da Informação Ponderada Aleatória . . . . .	39
2.6	Resultados obtidos pelo método da Máxima Informação Ponderada com $\alpha$ aleatório . . . . .	40
2.7	Resultados obtidos pelos métodos de Sympson-Hetter, Máxima Informação e Randomesque . . . . .	46
2.8	Frequência dos números de itens coincidentes na amostra de 10 mil pares de respondentes para o método de Máxima Informação . . . . .	46
2.9	Frequência dos números de itens coincidentes na amostra de 10 mil pares de respondentes para o método Randomesque . . . . .	47
2.10	Frequência dos números de itens coincidentes na amostra de 10 mil pares de respondentes para o método Sympson-Hetter . . . . .	47
2.11	Resultados obtidos pelo método da Máxima Informação Ponderada . . . . .	48
2.12	Frequência dos números de itens coincidentes na amostra de 10 mil pares de respondentes para o método de Máxima Informação Ponderada . . . . .	49
2.13	Resultados obtidos pelo método da Informação Ponderada Aleatória . . . . .	50
2.14	Resultados obtidos pelo Método da Método da Máxima Informação Ponderada com $\alpha$ aleatório . . . . .	51

# Lista de Abreviaturas

*ASVAB* Armed Services Vocational Aptitude Battery

*CAT* Teste Adaptativo Computadorizado

*CCI* Curva Característica do Item

*CEBRASPE* Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos

*EAP* Média a posterior

*ECDL* European Computer Driving Licence

*EMVP* Estimador de máxima verossimilhança ponderado

*EQM* Erro Quadrático Médio

*GMAT* Graduate Management Admission Test

*GRE* Graduate Record Examination

*IPA* Informação Ponderada Aleatória

*KL* Kullback-Leibler

*MAP* Moda a posteriori

*MI* Máxima informação

*MIP* Máxima Informação Ponderada

*ML3* Modelo Logístico de 3 Parâmetros

*MV* Máxima verossimilhança

*NAEP* National Assessment of Educational Progress

*PSD* Desvio padrão posterior

*RA* Randomesque

*SH* Sympson e Hetter

*TOEFL* Test of English as a Foreign Language

*TRI* Teoria de Resposta ao Item

*UnB* Universidade de Brasília

*USP* Universidade de São Paulo

## Resumo

O desenvolvimento dos testes adaptativos computadorizados só foi possível devido aos avanços tecnológicos das últimas décadas, permitindo que esta metodologia obtenha estimativas para a habilidade dos examinados com base em um número reduzido de itens selecionados especificamente para cada respondente a partir do seu traço latente estimado. Suas dificuldades surgem quando um pequeno grupo de itens é exposto com frequência, colocando em risco a segurança do teste. Assim, esta pesquisa tem como objetivo propor um método para a etapa de seleção de itens, com base no uso de informação ponderada por uma potência de ordem  $\alpha$  da proporção corrente de respondentes não expostos a cada item, a fim de reduzir a taxa de exposição dos itens, de modo que não tenham taxas de exposição muito altas ou itens que nunca tenham sido expostos mesmo com grau de dificuldade próximo da habilidade real  $\theta$  do respondente. Os resultados demonstram as vantagens da metodologia proposta com relação às já utilizadas, apresentando melhor desempenho na proporção de itens superexpostos com todos os valores de  $\alpha$  para informação ponderada aleatória e aumentando a proporção de itens expostos para maiores valores de  $\alpha$  no método de máxima informação ponderada, para o banco de itens simulado. O método de máxima informação ponderada com  $\alpha$  aleatório apresentou melhor desempenho dentre todos os métodos aqui abordados ao ser aplicado no banco de itens real. Outras vantagens relacionadas a escolha dos valores de  $\alpha$  também são citadas.

**Palavras-chave:** Teoria de Resposta ao Item (TRI); Teste Adaptativo Computadorizado(CAT); Informação ponderada.

# Abstract

The development of computerized adaptive tests was only possible due to the technological advances of the last decades, allowing this methodology to obtain estimates for the ability of the examinees based on a reduced number of items selected specifically for each respondent from their estimated latent trait. Its difficulties arise when a small group of items is exposed frequently, jeopardizing the security of the test. Thus, this research aims to propose a method for the item selection step, based on the use of information weighted by a power of order  $\alpha$  of the current proportion of respondents not exposed to each item, in order to reduce the exposure rate of the items, so that they do not have very high exposure rates or items that have never been exposed even with a degree of difficulty close to the respondent's real skill  $\theta$ . The results demonstrate the advantages of the proposed methodology in relation to those already used, presenting better performance in the proportion of overexposed items with all values of  $\alpha$  for random weighted information and increasing the proportion of exposed items for higher values of  $\alpha$  in the weighted maximum information method, for the simulated item bank. The weighted maximum information method with random  $\alpha$  presented the best performance among all the methods discussed here when applied to the real item bank. Other advantages related to the choice of  $\alpha$  values are also mentioned.

**Keywords:** Item Response Theory (IRT); Computerized Adaptive Test (CAT); Weighted information.

# Introdução

Provas realizadas nas mais diversas esferas têm como base a Teoria Clássica das Medidas e apresentam resultados baseados em seus escores brutos, ou seja, o número de acertos obtidos em uma determinada prova, o que inviabiliza a comparação entre indivíduos que não foram submetidos à mesma prova. Com o intuito de suprir esta e outras necessidades do processo avaliativo, foi desenvolvida a Teoria da Resposta ao Item (TRI) na década de 50.

A TRI é uma modelagem estatística que tem como objetivo mensurar características que não podem ser observadas diretamente. Segundo Andrade et al. (2000), sua metodologia sugere formas de representar a relação entre a probabilidade de um indivíduo dar certa resposta a um item e seus traços latentes, proficiências ou habilidades na área de conhecimento avaliada, tendo como elementos centrais os itens a serem respondidos e não mais as provas, de um modo geral. Tal modelagem possibilitou comparações no desempenho de diferentes populações, desde que tenham sido submetidas a provas com alguns itens em comum, ou ainda, a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente distintas, viabilizando também o acompanhamento anual de um determinado grupo. Desse modo, o interesse na aplicação de técnicas derivadas da TRI segue se expandindo.

Com o desenvolvimento das tecnologias, foi possível aplicar a TRI em novas formas de avaliação, desta vez, por meio de computadores, dando início aos testes adaptativos computadorizados (CAT), que vem ganhando cada vez mais espaço na área acadêmica, servindo como instrumento de diversos estudos com propostas de reduzir suas limitações quanto ao processo de aplicação. O CAT destaca-se por obter estimativas precisas do nível de habilidade dos examinados, com base em um pequeno subconjunto de itens selecionados especificamente para cada indivíduo, que varia de acordo com as respostas obtidas. Este tipo de teste tem como base a administração adaptativa dos itens, por fazer a sua seleção de acordo com a estimativa provisória do nível de habilidade e dos parâmetros dos itens, administrando combinações diferenciadas para cada respondente (Ruveta e Ponsoda, 1998; Magis e Barrada, 2017).

Desse modo, os itens subseqüentes fornecem informação máxima para estimação da habilidade do indivíduo. Logo, não há necessidade de que o examinado responda itens com grau de dificuldade diferente do que indica o seu traço latente, pois não fornecem informações relevantes para a estimação do mesmo, contornando assim, a grande quantidade de itens exigidos em testes convencionais (papel-e-caneta).

As dificuldades relacionadas ao CAT surgem a partir do momento em que vários respondentes tem acesso ao mesmo item em um curto período de tempo, sendo caracterizados como itens superexpostos, enquanto outros itens do mesmo banco não são administrados, gerando problemas relacionados ao custo do teste e a sua segurança. Itens superexpostos perdem a capacidade de efetuar medições válidas, pois se o indivíduo tem conhecimento prévio do item, antes mesmo de ser administrado, ocorre a superestimação do seu traço latente (Barrada, Olea e Ponsoda, 2007).

Com o objetivo de controlar a taxa de exposição dos itens, Sympson e Hetter (1985) criaram uma das primeiras abordagens para reduzir este efeito, sendo que posteriormente foram desenvolvidos outros métodos que tinham também o intuito de aumentar a taxa de uso de itens raramente selecionados.

Segundo Chang e Twu (1998), nas duas décadas que antecedem a sua pesquisa, dificuldades teóricas relacionadas à implementação de CATs em grande escala foram gradualmente superadas. Em seu artigo, classificam os cinco melhores métodos de controle de exposição disponíveis naquele período, entre eles o Método Sympson e Hetter, também citado em artigos mais recentes como Costa (2009) e Galvão, Neto e Borges (2013).

Posteriormente, outros estudos realizaram comparações entre diferentes métodos para controle da taxa de exposição de itens, como por exemplo, Costa (2009) que discute em sua pesquisa três métodos de seleção de itens que envolvem a TRI, entre eles, o critério de Máxima Informação, proposto na década de 80 e utilizado até hoje.

Em sua pesquisa, Revuelta e Ponsoda (1998) propõem o método aleatório, posteriormente citado por Leroux, Lopez e Hembry (2013) em comparação com outros dois métodos distintos.

Esta pesquisa teve como objetivo propor uma nova metodologia para controlar a taxa de exposição dos itens e compara-lá com os métodos adotados regularmente, que serão apresentados ao longo do desenvolvimento deste trabalho.

# Capítulo 1

## Referencial Teórico

### 1.1 Teoria da Resposta ao Item

A Teoria da Resposta ao Item é um conjunto de modelos matemáticos propostos por Lord (1952), que tiveram seu início na década de 50. Os primeiros modelos foram desenvolvidos na forma de uma função Ogiva Normal considerando apenas uma habilidade de um único grupo, sendo substituída posteriormente pela função Logística que é utilizada até os dias de hoje em modelos univariados e bivariados (Andrade, Tavares e Vale, 2000).

Para itens dicotômicos, a TRI estabelece três modelos baseados no número de parâmetros, sendo eles: o grau de dificuldade ( $b_i$ ), o poder de discriminação ( $a_i$ ) e a possibilidade de acerto ao acaso ( $c_i$ ), obtendo assim os modelos logísticos unidimensionais de 1, 2 e 3 parâmetros, respectivamente:

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-(\theta_j - b_i)}}, i = 1, \dots, I; j = 1, \dots, n, \quad (1.1)$$

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, i = 1, \dots, I; j = 1, \dots, n, \quad (1.2)$$

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \quad (1.3)$$

com  $i = 1, \dots, I; j = 1, \dots, n$ , onde:

- $U_{ij}$  assume valor 1, se o indivíduo  $j$  "acerta" o item  $i$ , e 0, caso contrário.
- $P(U_{ij} = 1|\theta_j)$  é a probabilidade de um indivíduo  $j$  com habilidade  $\theta_j$  responder corretamente o item  $i$ .
- $\theta_j$  é o traço latente da  $j$ -ésima unidade amostral.

- $I$  é a quantidade de itens.
- $n$  o número de unidades amostrais.
- $D$  é um fator de escala e assume o valor 1 (adota-se  $D=1,702$  caso se queira a curva próxima da ogiva da distribuição  $N(0,1)$ ).

O modelo logístico de 3 parâmetros (ML3) proposto por Birnbaum (1968) é o mais completo, pois considera a dificuldade, a discriminação e a probabilidade de resposta correta dada por indivíduos de baixa habilidade.

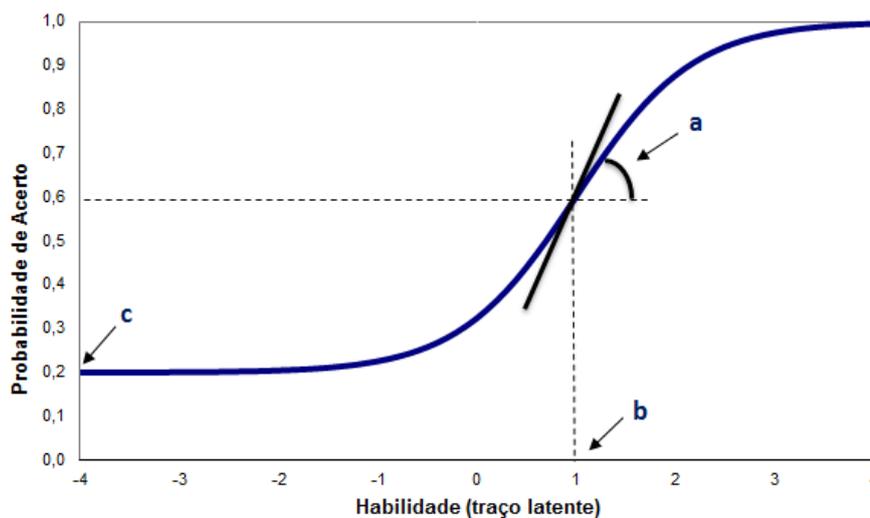
### 1.1.1 Curva Característica do Item

A relação entre a probabilidade de respostas corretas ao item  $i$  dentre todos os indivíduos da população com habilidade  $\theta_j$  e os parâmetros do modelo, é denominada Curva Característica do Item (CCI), fornecendo informações referentes aos parâmetros nela existentes, por meio da forma como as curvas se apresentam nos gráficos.

A Figura 1.1 mostra que esta relação não é linear, tendo como base o fato de que indivíduos com maior habilidade, possuem maior probabilidade de acertar o item. Caracteriza-se como monotônica crescente, sendo definida por meio dos parâmetros dos itens sua inclinação e seu deslocamento na escala de habilidade.

A probabilidade de acerto  $P(U_i = 1 | \theta)$  é representada no eixo das ordenadas e varia no intervalo de 0 a 1, enquanto a habilidade se encontra no eixo das abscissas e varia em uma escala de  $-\infty$  a  $+\infty$ .

Figura 1.1: Curva Característica do Item



O parâmetro de discriminação ( $a_i$ ) é determinado de acordo com a inclinação da curva característica do item em relação ao eixo da habilidade, podendo assumir qualquer valor em  $\mathbb{R}$ ,

sendo proporcional ao valor do ângulo formado por essa inclinação. Entretanto, não são esperados valores negativos para este parâmetro, pois indica que a curva está decrescendo em algum ponto, isto é, quanto maior a habilidade do indivíduo, menor a probabilidade de acerto do item. Valores de  $a$  muito pequenos, indicam itens com baixa discriminação, ou seja, respondentes com níveis de habilidades distintas têm aproximadamente a mesma probabilidade de responder corretamente ao item, não fornecendo informações relevantes para a estimação da habilidade.

O parâmetro  $b$  representa a abscissa do ponto de inflexão da curva característica do item que corresponde a uma probabilidade de acerto, medido na escala da habilidade ( $\theta_j$ ), definindo assim a habilidade necessária para uma probabilidade de acerto igual a  $(1 + c)/2$ . Itens mais difíceis tem um maior valor de  $b$ .

Outro parâmetro apresentado no gráfico é  $c$ , que representa a possibilidade de acerto ao acaso, em outras palavras, a probabilidade de um indivíduo com baixa habilidade, acertar o item e, portanto, encontra-se no intervalo entre 0 e 1. Quando temos o valor de  $c = 0$ , não existe probabilidade de acerto casual e trata-se de um modelo de dois parâmetros.

### 1.1.2 Função de Informação

A função de informação do item é dada por:

$$I_i(\theta) = \frac{\left[\frac{d}{d\theta} P_i(\theta)\right]^2}{P_i(\theta)Q_i(\theta)}, \quad (1.4)$$

Esta medida permite avaliar a quantidade de informação em um item para a estimação da habilidade, onde,

$I_i(\theta)$  é a “informação” fornecida pelo item  $i$  no nível de habilidade  $\theta$ ,

$$P_i(\theta) = P(U_{ij} = 1|\theta),$$

$$Q_i(\theta) = 1 - P_i(\theta).$$

Quando substituimos a probabilidade  $P_i(\theta)$  pela estabelecida no modelo logístico de 3 parâmetros, a equação pode ser escrita da seguinte forma:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[ \frac{P_i(\theta) - c_i}{1 - c_i} \right]^2 \quad (1.5)$$

Com base na função acima, pode-se observar o impacto que os parâmetros geram na quantidade de informação que pode ser obtida pelo item. Esta relação fica evidente pois  $a_i^2$  é diretamente proporcional à informação, logo, quanto maior o valor de  $a_i$ , maior a informação do item. O inverso é válido para  $c_i$ , pois quanto mais  $c_i$  se aproxima de zero, maior é a informação. Outra relação importante a ser destacada, é que a informação do item é maior quando a dificuldade se aproxima da habilidade do examinado (Andrade, Tavares e Vale 2000; Labarrere, da-Silva e

Costa, 2011).

A função de informação do teste é descrita por Andrade et al. (2000) de duas maneiras. A primeira é pela soma das informações fornecidas por cada item que compõe o teste.

$$I(\theta) = \sum_{i=1}^I I_i(\theta). \quad (1.6)$$

A segunda é através do erro-padrão de medida (erro padrão de estimação), que é obtido por meio da seguinte equação:

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (1.7)$$

### 1.1.3 Escala de Habilidade

A escala da habilidade pode, teoricamente, assumir valores reais no intervalo de  $-\infty$  a  $+\infty$ . Portanto, para definir a escala é necessário determinar uma origem e uma unidade de medida, a fim de representar, respectivamente, o valor médio e o desvio-padrão das habilidades dos respondentes.

Usualmente na TRI é considerada a escala (0,1), onde 0 é a média e 1 é o desvio-padrão, obtendo assim, a maioria dos valores entre -2 e 2 para o parâmetro de dificuldade do item e valores entre 0 e 2 para o poder de discriminação, classificando como itens com boa discriminação aqueles que apresentarem valores acima de 1 (Andrade, Tavares e Vale, 2000).

Entretanto, vale destacar que a métrica aplicada à escala da habilidade não interfere nos resultados obtidos por meio dela, sendo possível fazer a transformação de escalas, ou seja, a habilidade de um indivíduo é invariante à escala aplicada.

### 1.1.4 Suposições do Modelo

Para que o modelo possa ser ajustado é necessário que sejam atendidas as suposições de unidimensionalidade do teste e independência local.

A unidimensionalidade do teste afirma que deve haver apenas uma habilidade responsável pela realização de todos os itens da prova. Contudo, durante o processo de execução de um teste é difícil que haja apenas um traço latente influenciando na probabilidade de acerto do respondente, portanto a suposição de que exista um traço dominante é satisfatória para o modelo.

Outra suposição a ser feita é a de independência local ou independência condicional, a qual assume que os itens são respondidos de forma independente, por cada examinando, para uma dada proficiência, ou seja, a resposta dada a um determinado item, independe do que foi respondido a qualquer outro item do teste. Sendo fundamental para o processo de estimação dos parâmetros do modelo.

De modo geral pode-se considerar apenas uma suposição, pois a unidimensionalidade implica diretamente na independência local, portanto se a primeira estiver satisfeita, automaticamente os pressupostos são satisfeitos (Hambleton et al., 1991).

### 1.1.5 Calibração dos Itens e Estimação do Traço Latente

Um dos maiores desafios da TRI está diretamente relacionado ao processo de estimação dos parâmetros dos itens, também conhecido como calibração dos itens, e a estimação do traço latente. Em teoria, esta situação pode apresentar-se em três cenários distintos; tendo inicialmente itens calibrados, sendo necessário estimar apenas o traço latente : tendo conhecimento do traço latente e desejamos realizar a calibração dos itens e, por fim, situações em que tanto a estimação do traço latente, quanto a estimação dos parâmetros dos itens, são desconhecidas, sendo necessário estimá-las simultaneamente.

Em todas as situações descritas, a estimação pode ser realizada por meio do Método de Máxima Verossimilhança junto a um processo iterativo como *Newton-Raphson*, *Scoring de Fisher* ou procedimento bayesiano (Andrade et.al., 2000). Na estimação conjunta, temos quatro parâmetros para cada respondente, sendo três para cada item ( $a_i$ ,  $b_i$  e  $c_i$ ) e um traço latente. Portanto o processo torna-se mais denso, podendo ser desenvolvido através da estimação em uma única etapa, onde os parâmetros dos itens e habilidades são estimados de forma conjunta, ou em duas etapas, realizando primeiro a estimação dos parâmetros dos itens e posteriormente, das habilidades.

Andrade, Tavares e Valle (2000) apresentam em seu trabalho maiores detalhes sobre as vantagens e as desvantagens de cada uma das propostas de estimadores.

## 1.2 Testes Adaptativos Computadorizados

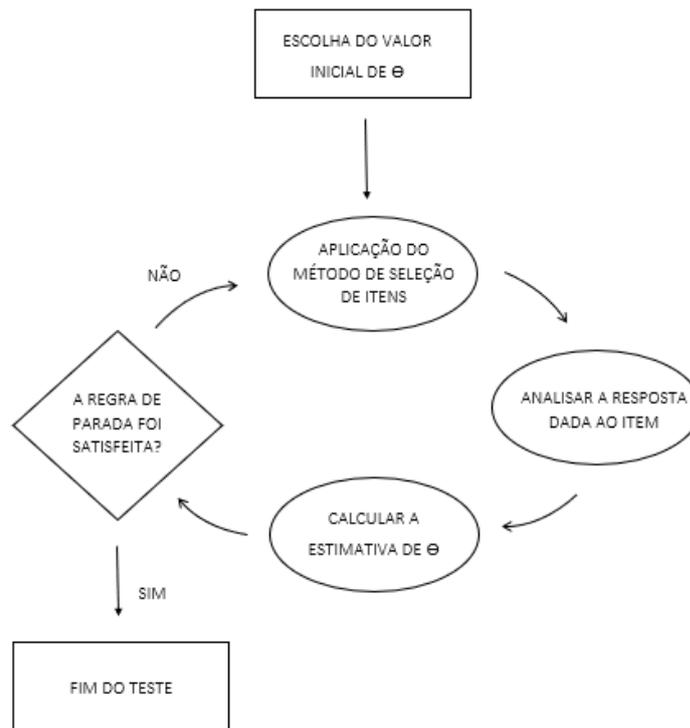
Com a informatização do processo educacional surgiram os testes adaptativos computadorizados como possibilidade de alcançar o mesmo nível de precisão de testes tradicionais, na estimação da habilidade do respondente com um número reduzido de itens, onde os resultados encontram-se disponíveis imediatamente após a realização do teste, aumentando assim a eficiência da avaliação (Salgado, 2018).

Para a aplicação do teste os itens são previamente calibrados e durante sua aplicação são selecionados especificamente para cada indivíduo, a proficiência do examinado é mensurada interativamente com base no grau de dificuldade do item, nível de informação fornecida, proficiência atual, resposta obtida anteriormente, entre outros, dependendo do método a ser aplicado, podendo apresentar uma quantidade diferente de itens para cada respondente (Labarrere, da-Silva e Costa, 2011).

A Figura 1.2 apresenta um fluxograma do CAT, que consiste em escolher o valor inicial da habilidade do indivíduo para aplicação do método de seleção do próximo item a ser aplicado. Após a escolha do item é analisada a resposta dada pelo examinado para aquele item para, enfim, ser calculada uma nova estimativa da sua habilidade. Se a regra de parada do teste for satisfeita,

o mesmo é finalizado, caso contrário é aplicado novamente o método de seleção de itens e o processo ocorre novamente, até que a condição seja atendida (Galvão, Neto e Borges, 2013).

Figura 1.2: Fluxograma do CAT



Fonte: Elaborado pelo autor (2022).

O valor inicial de  $\theta$  não causa impacto na precisão do valor da estimativa da proficiência obtida por este processo. Contudo, diferentes escolhas repercutem na quantidade de itens necessários para estimação do verdadeiro valor de  $\theta$  e da satisfação do critério de parada. Fonseca (2019) define o uso antecipado da média da distribuição de habilidades como valor inicial para todos os examinandos, sendo uma opção simples, destacando também a aleatoriedade nos valores como forma de evitar a atribuição de itens iniciais similares aos examinandos.

A etapa de aplicação do método de seleção de itens é fundamental para eficiência no processo de estimação do parâmetro de habilidade do examinando, balanceamento de conteúdo e controle à exposição de cada item, como destaca Zheng (2014, apud FONSECA, 2019). Assim, é de suma importância empregar um método adequado de escolha de itens para estimação das proficiências, uma vez que é por meio deste método que são definidos a proporção de vezes que um item pode ser utilizado, a atribuição de pesos ou aleatorização da seleção durante o processo e o equilíbrio entre o controle da taxa de exposição e a precisão da estimativa, entre outros.

Alguns exemplos de aplicações de testes adaptativos computadorizados são citados por Spennassato, Trierweiler e Andrade (2016), sendo eles, Graduate Record Examination (GRE), Graduate Management Admission Test (GMAT), Armed Services Vocational Aptitude Battery (AS-VAB), Test of English as a Foreign Language (TOEFL), European Computer Driving Licence (ECDL), National Assessment of Educational Progress (NAEP).

Em nível nacional temos a Proficiência em Inglês da Universidade de Brasília (UnB) e o Computer-Aided Learning of English for Academic Purposes da Universidade de São Paulo (USP) (Salgado, 2018).

### 1.2.1 Estimativa do Traço Latente

A estimação do traço latente em testes adaptativos computadorizados ocorre iterativamente após cada item respondido, sob a suposição de que as habilidades dos examinandos são estocasticamente independentes entre si, ou seja, os respondentes não possuem informação de outros examinandos, permitindo estimar cada habilidade separadamente.

O procedimento para se obter estimativas por meio do CAT leva em consideração a função de Verossimilhança, que relaciona a habilidade e as respostas do  $j$ -ésimo examinando aos primeiros  $p - 1$  itens, para o modelo de 3 parâmetros e é dada por:

$$L(\theta_j | u_{j1}, \dots, u_{j,p-1}) = \prod_{t=1}^{p-1} P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}} \quad (1.8)$$

Após maximizar esta verossimilhança em relação a todos os possíveis valores de  $\theta_j$ , obtém-se o Estimador de Máxima Verossimilhança (EMV):

$$\hat{\theta}_{j_{u_1, \dots, u_{p-1}}}^{EMV} \equiv \arg \max_{\theta_j} \{L(\theta_j | u_{j1}, \dots, u_{j,p-1}) : \theta_j \in (-\infty, \infty)\}. \quad (1.9)$$

O processo de maximização da verossimilhança acarreta problemas no processo de estimação das habilidades, como por exemplo, no momento em que o examinado responde todos os itens corretamente ou incorretamente. Foi proposto por Warm (1989) um estimador de máxima verossimilhança ponderado (EMVP), que consiste em considerar uma função ponderada, em seguida inseri-la na verossimilhança e maximizar a função obtida com relação as habilidades (Azevedo, 2003), de modo que:

$$\hat{\theta}_{j_{u_1, \dots, u_{p-1}}}^{EMVP} \equiv \arg \max_{\theta_j} \{w_{p-1}(\theta) L(\theta_j | u_{j1}, \dots, u_{j,p-1}) : \theta_j \in (-\infty, \infty)\}. \quad (1.10)$$

definido de tal forma que  $w_{p-1}(\theta)$  satisfaça:

$$\frac{\partial w_{p-1}(\theta)}{\partial \theta^2} = \frac{H_{p-1}(\theta)}{2I_{p-1}(\theta)} \quad (1.11)$$

com

$$H_{p-1}(\theta) = \sum_{t=1}^{p-1} \frac{[s'_{tj}(\theta)][s''_{tj}(\theta)]}{s_{tj}(\theta)[1 - s_{tj}(\theta)]},$$

$$s''_{tj}(\theta) = \frac{\partial s_{tj}(\theta)}{\partial \theta^2}$$

e  $I_{p-1}(\theta)$  é a informação do teste após  $(p - 1)$  itens respondidos.

Outras opções de estimadores para  $\theta$  são sugeridos com base nos métodos bayesianos. Nesta abordagem será combinada a função de verossimilhança e a distribuição *a priori* assumida para os valores desconhecidos da habilidade,  $g(\theta_j)$ , resultando na distribuição *a posteriori* para  $\theta_j$

$$g(\theta_j | u_1, \dots, u_{p-1}) = \frac{L(\theta_j; u_1, \dots, u_{p-1})g(\theta_j)}{\int_{\mathfrak{R}} L(\theta_j; u_1, \dots, u_{p-1})g(\theta_j)d\theta_j} \quad (1.12)$$

Em sua pesquisa, van der Linden e Pashley (2003, apud COSTA, 2009) destacam que é comum atribuir a Distribuição Uniforme em um intervalo fechado para a densidade  $g(\theta_j)$ .

Os estimadores pontuais para  $\theta_j$  pelo método Bayesiano tem como base a distribuição apresentada em (1.12). Os mais utilizados são a moda a posteriori (MAP), proposto por Lord (1986) e Mislevy (1986a); ou a média a posteriori (EAP) amplamente discutido por Bock e Mislevy (1982).

No método MAP, deseja-se maximizar a distribuição a posteriori de  $\theta_j$  e é dado pela equação a seguir:

$$\hat{\theta}_{j u_1, \dots, u_{p-1}}^{MAP} \equiv \arg \max_{\theta_j} \{g(\theta_j | u_{j1}, \dots, u_{j,p-1}) : \theta \in (-\infty, \infty)\}. \quad (1.13)$$

O procedimento de estimação MAP sempre converge, independentemente do padrão de resposta dos indivíduos, diferente do estimador MV, que não converge nos casos em que o respondente erra ou acerta todos os itens. Contudo, vale destacar que a equação de estimação para  $\theta$  é não-linear, por isso exige cálculos mais complexos do que o MV conforme citado por Azevedo (2003 apud COSTA, 2009).

A outra opção através do método bayesiano é a estimação pela Média a Posteriori (EAP), onde o estimador EAP é o valor esperado da distribuição a posteriori e pode ser representado por:

$$\hat{\theta}_{j u_1, \dots, u_{p-1}}^{EAP} \equiv \int_{\mathfrak{R}} \theta_j g(\theta_j | u_1, \dots, u_{p-1}) d\theta_j \quad (1.14)$$

Frequentemente, deseja-se resumir o grau de incerteza sobre o traço latente estimado. Essa mensuração pode ser feita sob a forma de variância da distribuição a posteriori de  $\theta$ , dado por:

$$Var(\theta_j | u_1, \dots, u_{p-1}) \equiv \int_{\mathfrak{R}} [\theta_j - E(\theta_j | u_1, \dots, u_{p-1})]^2 g(\theta_j | u_1, \dots, u_{p-1}) d\theta_j \quad (1.15)$$

Para a *priori* uniforme, o MAP e o EMV possuem propriedades semelhantes, no entanto, ao atribuir outras distribuições *a priori*, pode-se obter uma distribuição *a posteriori* multimodal, ou seja, deve-se tomar cuidado na busca do máximo global, uma vez que a maximização pode retornar algum máximo local.

Segundo Linden e Pashley (2010 apud SPENASSATO, 2017) a escolha da *priori* pode influenciar na seleção dos itens iniciais a serem aplicados gerando más estimativas iniciais, interferindo nas estimativas obtidas em testes curtos, por exemplo, com 10 itens. Testes com aproximadamente 20 a 30 itens, já são considerados testes longos, sendo assim, não terá impacto sob a estimação.

Para uma distribuição a *priori* própria, a estimativa EAP sempre vai existir, mesmo para aqueles padrões em que o MV não existe. Contudo, o estimador EAP apresenta-se mais estável para todos os tamanhos de testes adaptativos, incluindo o primeiro item administrado, além de ser de fácil obtenção, já que nenhum procedimento iterativo é exigido (Salgado, 2018; Costa 2009).

Em sua pesquisa, Salgado (2018) destaca os diferentes momentos em que o traço latente é estimado em testes adaptativos, sendo eles: a estimativa inicial, realizada para dar início ao processo de seleção dos itens; após cada item respondido é realizada uma nova estimação, com intuito de selecionar os itens que possuem maior informação para a habilidade estimada do respondente; e a estimativa final, após atender um dos critérios de parada e assim concluir o teste. Portanto a escolha da metodologia a ser aplicada deve levar em consideração as diversas etapas que ocorrem durante o processo, como por exemplo, a possibilidade de respostas totalmente corretas ou incorretas, o que é comum no estágio inicial do teste.

Após esta etapa é mais vantajoso fazer uso de uma metodologia que tenha convergência mais rápida, reduzindo assim o tempo de aplicação do exame e o número de itens expostos. Para a finalização do teste o método escolhido irá fornecer uma estimativa resumo da sua habilidade, sendo desejado a mais próxima possível do seu valor real. É possível realizar combinações entre o método de estimação e o critério de escolha do item, visando um melhor desempenho geral do teste. Dentre os pares mais utilizados temos o EMV para estimar o traço latente e a máxima informação para realizar a seleção do item, ou a seleção do item baseada na distribuição a posteriori, combinada com o processo de estimação bayesiana (Salgado, 2018).

### 1.2.2 Critério de Seleção de Itens

Uma das dificuldades enfrentada nos testes adaptativos é a falta de acurácia nos estágios iniciais da aplicação do teste e está diretamente relacionada com a escolha do critério de seleção de itens adequado (Chen, Ankenmann e Chang, 2000; Jatobá, 2018). Para um melhor desempenho do teste, o examinando deve ser exposto a itens não muito difíceis nem muito fáceis para este candidato, de modo a contornar o problema descrito na seção anterior. Entretanto, apenas a avaliação do nível de dificuldade dos itens não é suficiente para realizar a seleção mais adequada dos itens para a estimação das proficiências, por isso os métodos de seleção adaptativa buscam a combinação dos parâmetros dos itens e da estimativa da proficiência, denominada de medida de *Informação* (Lord, 1980; Costa, 2009).

Apesar da necessidade de um maior aprimoramento sobre o tema, já existem duas abordagens que podemos analisar: baseada em informação e bayesiana, que seguem estruturas distintas, selecionando o item que fornece maior informação com relação a estimativa de  $\theta$  e selecionando itens com base em *prioris* e *posterioris* das estimativas de  $\theta$ , respectivamente.

O uso da função de informação do teste para a seleção do item foi proposto por Birnbaum (1968) e consiste em maximizar (1.6) em  $\theta = \hat{\theta}_{u_{1j}, \dots, u_{(k-1)j}}$  como critério para seleção do  $k$ -ésimo item, tendo como vantagem a facilidade de incorporar a informação de cada item selecionado. Para escolha do próximo item, temos:

$$i_k \equiv \arg \max_i \{I_{U_i}(\hat{\theta}_{u_{1j}, \dots, u_{(k-1)j}}) : i \in R_k\}, \quad (1.16)$$

sendo:

- $i_k$   $k$ -ésimo item selecionado para o teste.
- $R_k$  conjunto de itens que ainda não foram utilizados no banco.

Ao utilizar a combinação do EMV para estimação de  $\theta$  em conjunto com a máxima informação, para um banco de itens grande de modo que todos os possíveis valores de  $a_i$  estejam entre 0 e  $\infty$  e valores de  $c_i$  limitado a 1, tem-se um estimador que converge para o valor real de  $\theta$  com variância amostral próxima ao limitante mínimo fornecido pela Informação de Fisher (Chang e Ying, 2009; Salgado, 2018), que é comumente utilizado para seleção do próximo item baseado em informação, por ser um dos conceitos mais conhecidos na literatura estatística, mensurando o grau de informação de uma variável, através da definição à seguir:

**Definição 1.2.2.1 (Informação de Fisher):** *Seja  $U$  um vetor de variáveis aleatórias com função de probabilidade  $P(u|\theta)$ . A Informação de Fisher esperada de  $\theta$  através de  $U$  é dada por:*

$$I_F(\theta) = -E_{U|\theta} \left[ \frac{\partial^2 \log P(u|\theta)}{\partial \theta^2} \right] \quad (1.17)$$

resultando no valor médio da curvatura da log-verossimilhança. Quanto maior esta curvatura, maior será a informação sumarizada na função de Verossimilhança e o valor de  $I_F(\theta)$ . Permitindo calcular individualmente para cada item quanto ele contém de informação para a medida de habilidade a partir dos seus parâmetros, possuindo maior eficiência e precisão na estimativa da habilidade quando  $\hat{\theta}$  está próximo da habilidade real do indivíduo. O  $k$ -item selecionado por esse método será aquele que possuir a máxima Informação de Fisher avaliada em  $\hat{\theta}_{k-1}$  (Costa, 2009).

Para o Modelo Logístico de três Parâmetros (ML3P) a Informação de Fisher do  $k$ -ésimo item é dada por:

$$I_{F,k}(\theta) = \frac{\left[ \frac{\partial P_k(\theta)}{\partial \theta} \right]^2}{P_k(\theta)[1 - P_k(\theta)]} = D^2 a_k^2 \left[ \frac{1 - P_k(\theta)}{P_k(\theta)} \right] \left[ \frac{P_k(\theta) - c_k}{1 - c_k} \right]^2 \quad (1.18)$$

Sabendo que a Informação de Fisher é inversamente proporcional ao erro-padrão do estimador de Máxima Verossimilhança, maximizar essa medida é obter uma estimação de  $\hat{\theta}_{k-1}$  mais eficiente ao minimizar a variância assintótica de  $\hat{\theta}_{k-1}$  (Costa, 2009).

Outra medida de informação que pode ser utilizada para substituir a Informação de Fisher na escolha do item é a informação de Kullback-Leibler, como foi proposto em Chang e Ying (1996). Sua principal justificativa é reduzir problemas relacionados aos erros iniciais do teste, quando a estimativa da proficiência estiver distante do valor real. É de extrema importância a etapa inicial da estimação para redução da quantidade de itens aplicados no decorrer do teste.

Autores como Chang e Ying (1996) definem a informação obtida pela medida de Kullback-Leibler como uma *Informação Global* e a Informação de Fisher como uma função de *Informação Local*. Em relação aos testes adaptativos tem-se que a informação global é ideal para seleção de itens quando não há informações suficientes sobre esse parâmetro, enquanto a informação local é preferida para seleção de itens quando existe conhecimento suficiente sobre a localização do valor verdadeiro de  $\theta$ .

A informação de Kullback-Leibler consiste em medir a distância entre duas distribuições, sendo que, quanto maior for a informação, mais fácil é discriminar entre duas distribuições.

**Definição 3.3.2 (Informação de Kullback-Leibler):** *Seja  $\theta_0$  o valor verdadeiro do parâmetro. Para qualquer valor de  $\theta$ , a Informação KL para o  $i$ -ésimo item (com resposta  $u_i$ ) é definido por:*

$$K_i(\theta|\theta_0) \equiv E_{\theta_0} \log \left[ \frac{L_i(\theta_0; u_i)}{L_i(\theta; u_i)} \right]. \quad (1.19)$$

A utilização de barras duplas verticais tem o intuito de diferenciar da notação utilizada para indicar condicionamento, sendo padrão para a Informação KL (Kullback, 1959).

Para o modelo logístico de três parâmetros, como  $U_i \sim \text{Bernoulli}(P_i(\theta_0))$  com o verdadeiro valor do traço latente,  $\theta_0$ , e a atual estimativa do traço latente,  $\theta$ , pode ser expressa por:

$$\begin{aligned} K_i(\theta|\theta_0) &= E_{\theta_0} \log \left[ \frac{P_i(\theta_0)^{u_i} [1 - P_i(\theta_0)]^{1-u_i}}{P_i(\theta)^{u_i} [1 - P_i(\theta)]^{1-u_i}} \right] \\ &= E_{\theta_0} \log \left[ \frac{P_i(\theta_0)}{P_i(\theta)} \right] + [1 - P_i(\theta_0)] \log \left[ \frac{1 - P_i(\theta_0)}{1 - P_i(\theta)} \right] \end{aligned} \quad (1.20)$$

em que:  $K_i(\theta|\theta_0) \geq 0$  e  $K_i$  não é simétrica, ou seja:  $K_i(\theta|\theta_0) \neq K_i(\theta_0|\theta)$ .

Para um teste com  $k - 1$  itens, a Informação do teste KL consiste na soma de todas as informações individuais dos itens e é dado por:

$$K(\theta|\theta_0) = \sum_{i=1}^{k-1} K_i(\theta|\theta_0) = E_{\theta_0} [\log L(\theta_0; u_1, \dots, u_{k-1}) - \log L(\theta; u_1, \dots, u_{k-1})].$$

$K$  representa o poder discriminatório do item em dois níveis,  $\theta$  e  $\theta_0$ , e resume a informação contida no item com respeito a um amplo espectro de  $\theta$ , o que não requer que  $\theta$  seja próximo de  $\theta_0$ . Ao contrário da Informação de Fisher que é uma função de  $\theta_0$  e representa o poder discriminatório como uma função de informação local, somente ao redor de  $\theta_0$  (Hambleton, Swaminathan e Rogers, 1991; Costa, 2009; Chang e Ying, 1996).

Critérios de seleção do item com abordagem bayesiana podem ser encontrados em Van der Linden e Pashley (2009).

### 1.2.3 Métodos de Seleção de Itens

Nesta sessão serão apresentados alguns métodos de seleção de itens já existentes que serão utilizados para realizar comparações com o método proposto.

#### **Método de Máxima Informação (MI)**

O método de Máxima Informação proposto por Lord (1980) é um dos mais utilizados para seleção de itens durante a aplicação de testes adaptativos computadorizados e tem como medida a Informação de Fisher, com intuito de mensurar o grau de informação de uma variável (Costa, 2009).

Este método tem como base a proficiência atual do respondente para a escolha do próximo item, administrando aquele que for mais informativo para última estimativa da habilidade, tendo maior informação no momento em que a dificuldade do item esta próxima da habilidade do examinado. Entretanto esta não é a verdadeira proficiência do indivíduo, podendo então agregar informações irrelevantes para o teste e apresentar diversas vezes o mesmo item para vários respondentes por tratar-se de itens que apresentam maior informação em determinados pontos da escala (Galvão, Neto e Borges, 2013).

#### **Método Randomesque (RA)**

Este método foi proposto por Kingsbury e Zara (1989), onde a seleção do item é feita de forma aleatória entre os  $n$  itens mais informativos, ao invés de selecionar o item mais informativo. Esta metodologia é semelhante a outras já existentes, como por exemplo o procedimento 5-4-3-2-1, entretanto nesta abordagem é selecionado o mesmo número de itens mais informativos, onde  $n$  pode receber valores como 5 ou 10, por exemplo, inserindo assim aleatoriedade no processo, empregada na escolha do item, viabilizando a exposição dos itens que são selecionados com menos frequência (Cheng e Behrens, 2017; Lee e Dodd, 2012; Stocking, 1992). Para escolha dos  $n$  itens a cada etapa, serão levados em consideração aqueles que tiverem maior proporção de informação para o teste.

Foi concluído em Revuelta e Ponsoda (1998) que este método apresentou bons resultados nos momentos iniciais da aplicação do teste, impedindo que grande parte dos examinados iniciem respondendo os mesmos itens. Este método tem capacidade de obter o menor erro padrão e raiz do erro quadrático médio, enviesamento mais próximo de zero e administrar o menor número de itens em comparação com outros métodos (Leroux, Lopez e Hembry, 2013).

#### **Método Sympson e Hetter (SH)**

O método Sympson e Hetter determina a frequência de uso como critério para controlar a taxa de exposição de itens, por meio da atribuição de um parâmetro de exposição,  $k$ , para cada item, variando de 0 a 1, de modo que seja inversamente proporcional ao numero de vezes em que esse item é selecionado para administração, com o intuito de controlar probabilisticamente

a frequência com a qual o item é escolhido (Galvão, Neto e Borges, 2013; Costa, 2009).

Para isto é necessário um processo iterativo, anterior à etapa de administração, composto por simulações que tem como objetivo calcular os parâmetros de exposição dos itens que são usados durante o teste, determinando a probabilidade do item ser administrado desde que tenha sido escolhido.

Após a seleção do item, é gerado um número aleatório a partir de uma distribuição uniforme entre 0 e 1 para ser comparado com o parâmetro de exposição do item. Se o número aleatório for menor ou igual ao parâmetro de exposição, o item é administrado e não mais será selecionado para aquele respondente. Caso contrário, o item é desconsiderado (Leroux et. al., 2013).

Este processo pode ser melhor descrito como em Labarrere, da-Silva e Costa (2011), no passo-a-passo a seguir:

- Passo 1: Definir o valor do  $r$ , taxa de exposição máxima esperada de um item para o teste.
- Passo 2: Elaborar uma tabela de informação, contendo as listas dos itens do banco ordenados do maior para o menor, de acordo com uma função de informação para cada habilidade.
- Passo 3: Gerar um vetor de  $1, \dots, I$ , para o banco de tamanho  $I$ , com todos os elementos iguais a 1, que corresponde ao parâmetro de exposição dos itens, sendo o conjunto de parâmetros iniciais dos itens, definido pela probabilidade do item ser administrado ( $A$ ), dado que foi selecionado ( $S$ ), e será representado por  $P(A|S)$ . Os passos a seguir serão aplicados iterativamente até que algum critério de parada seja satisfeito:
- Passo 4: Selecionar uma amostra aleatória de examinados para simulação de um teste adaptativo. Que consiste em identificar o item mais informativo da tabela elaborada anteriormente, sendo o mais próximo possível da proficiência estimada do examinando, para cada item. Com base na distribuição Uniforme (0,1) dev-se gerar um número  $x$  pseudo-aleatório e aplicar o item  $i$ , se  $x \leq P_i(A|S)$ . É importante destacar que mesmo que o item  $i$  tenha sido ou não administrado, não é possível aplicá-lo futuramente para o mesmo examinado. Para a primeira simulação, todos os itens selecionados serão administrados, pois inicialmente tem-se que  $P_i(A|S) = 1, \forall i$ .
- Passo 5: Observar a seleção e aplicação dos itens no total de amostras simuladas, dispondo do número de vezes que cada item é selecionado ( $NS$ ), da quantidade de vezes que ele é administrado ( $NA$ ) e do número total de examinandos ( $NE$ ), para realizar o cálculo da probabilidade  $P(S) = NS/NE$  do item ser selecionado e  $P(A) = NA/NE$  de ser administrado, para cada item, quando a amostra completa for testada.
- Passo 6: Calcular os novos valores de  $P_i(A|S)$ , considerando  $P(S)$  e o valor de  $r$  obtido no Passo 1, dado por:

$$P_i(A|S) = \begin{cases} \frac{r}{P(S)} & \text{se } P(S) > r, \\ 1 & \text{se } P(S) \leq r \end{cases} \quad (1.21)$$

- Passo 7: Assegurar que o banco não será exaurido antes que os examinandos possam completar o teste. Sendo necessário garantir que um CAT de tamanho  $n$ , tenha no mínimo  $n$  itens com  $P_i(A|S) = 1$ , para que esses itens possam ser administrados quando selecionados, visto que o número aleatório gerado segue uma distribuição Uniforme (0,1). Como alternativa, deve-se certificar que para  $n$  itens  $P_i(A|S)$  não ultrapasse 1.
- Passo 8: Refazem-se os Passos 4 a 7, após calculados os novos valores de  $P_i(A|S)$  e utilizando a mesma amostra de examinandos, até que o valor máximo de  $P(A)$ , calculado no Passo 5, chegue próximo do limite um pouco acima de  $r$  e, então, oscile em sucessivas simulações.

#### 1.2.4 Critério de Parada do Teste

O critério de parada de um teste determina quando a aplicação de itens deve ser finalizada, podendo ser baseado em um critério predeterminado de precisão de medida ou na quantidade de itens. Critérios que determinam a mesma quantidade de itens para todos os respondentes caracterizam os testes como teste de comprimento fixo, enquanto os que tem como base uma regra de parada sob uma estimativa, são caracterizados como testes de comprimento variável. A escolha do tipo de critério a ser adotado varia de acordo com o objetivo do teste (Costa, 2009).

Segundo Weiss e Kingsbury (1984 apud Costa, 2009) em situações em que um teste adaptativo informatizado não é usado para classificação, um nível pré-determinado do erro-padrão é fixado, finalizando o teste quando seu nível for satisfeito, com a finalidade de mensurar cada examinando para um nível de precisão fixo, o que resultará em um conjunto de medidas em que todos os examinandos terão escores com equivalentes precisões.

O desvio padrão posterior (PSD) é o erro-padrão associado à estimação das proficiências pelo método EAP (Costa, 2009).

### 1.3 Considerações

Os modelos selecionados para este estudo têm características similares ao método que será proposto. Algumas de suas abordagens já têm como princípio a ponderação, atribuição de pesos e máxima informação na escolha do item, mas ainda sendo necessário desenvolvimentos posteriores para melhor proporcionalidade no uso dos itens.

Ademais, o método Sympson e Hetter, tem à necessidade de simulações para o cálculo dos parâmetros de exposição dos itens e os mesmos devem ser refeitos a cada alteração no banco de itens ou na distribuição das proficiências, exigindo atenção e esforços computacionais. Além de ter limitações para com itens com baixa probabilidade de seleção, pois irão permanecer pouco expostos.

Vale destacar que em qualquer restrição aplicada, há perda de informação do teste ao limitar o algoritmo de seleção dos itens. Portanto, é de suma importância que o algoritmo tenha opções um pouco menos informativas, mas também interessantes para serem aplicadas nos casos em

que um item não puder ser administrado, para tal, os itens que compõe o banco devem ser elaborados com alta qualidade.

Para a estimação do traço latente, o método proposto utilizará o estimador bayesiano da média a posteriori (EAP). O método será apresentado com mais detalhes no Capítulo 2.

## Capítulo 2

# Metodologia Proposta

A metodologia proposta é aplicada na etapa de seleção do item e tem como principal conceito o uso de informação ponderada com base na resposta do examinado. Fazendo uso de um parâmetro de ajuste dos pesos juntamente com a proporção de respondentes anteriores que foram expostos a cada item no banco de itens para obtenção da probabilidade proporcional, que irá ser o critério para a seleção do próximo item, a fim de limitar o uso de itens superexpostos e aumentar o número de aplicações de itens pouco ou quase nunca utilizados.

Esta seleção pode ser feita de forma aleatória ou considerando a máxima informação ponderada, com critério de parada condicionado ao PSD e ao número de itens já aplicados. O uso da informação ponderada de forma aleatória tem uma perspectiva similar a metodologia utilizada no Randomesque, entretanto não limita a escolha a 5 itens, possibilitando uso de todos os itens do banco.

Para contornar os problemas obtidos por meio da utilização da máxima informação para a última estimativa da habilidade, como é considerado no método de máxima informação, será feito uso da máxima informação baseada na informação ponderada, concatenando assim duas abordagens distintas.

Ao utilizar o método é necessário que o usuário defina a habilidade inicial dos respondentes, o valor do parâmetro  $\alpha$  (definido mais abaixo), a quantidade de itens e seus respectivos parâmetros. Com base nestas informações, é feita a seleção do próximo item a ser aplicado pelo **Método de escolha aleatória pela informação ponderada pela taxa de exposição dos itens**, quanto pelo **Método da máxima informação ponderada pela taxa de exposição dos itens**.

O primeiro método proposto neste trabalho consiste em escolher o item seguinte com probabilidade proporcional a

$$[(1 - p_i)^\alpha] I_i(\hat{\theta}_j) \quad (2.1)$$

sendo:

- $p_i$  a proporção de respondentes anteriores que foram expostos ao  $i$ -ésimo item.
- $\alpha \geq 0$  um parâmetro de ajuste dos pesos.

- $I_i(\hat{\theta}_j)$  a informação do  $i$ -ésimo item no valor corrente  $\hat{\theta}_j$  da estimativa da proficiência do  $j$ -ésimo respondente.

Esta metodologia tem estrutura similar a apresentada anteriormente, sendo composta também pela proporção de respondentes, parâmetro de ajuste dos pesos e informação do item. Sua diferença está na seleção do item mais informativo para a proficiência estimada, i.e.,

$$i_k = \arg \max_{i \in R_k} [(1 - p_i)^\alpha] I_i(\hat{\theta}_j)$$

O  $\alpha$  é um parâmetro de ajuste dos pesos capaz de modificar o cenário da escolha dos itens, para que não haja muita dispersão ou concentração das probabilidades de escolha dos mesmos.

Cabe ao usuário especificar o valor de  $\alpha$  mais adequado a partir dos seus objetivos e da melhor ponderação para sua aplicação. Dado que se tem o banco de itens calibrado, pode-se realizar esse processo por meio de simulações para valores de  $\theta$  mais específicos, variando o  $\alpha$  para analisar a taxa de exposição dos itens e tomar uma decisão em relação ao melhor valor de  $\alpha$  para aquela situação específica.

Na seleção por meio da máxima informação ponderada o próximo item a ser administrado é o item mais informativo para a proficiência estimada. Para ambos os casos quando  $\alpha = 0$ , temos o método de escolha do próximo item com probabilidade diretamente proporcional a  $I_i(\hat{\theta}_j)$  e  $\arg \max I_i(\hat{\theta}_j)$  respectivamente, como caso particular dos métodos propostos. Quanto maior o valor de  $\alpha$ , maior o peso atribuído para a taxa de exposição.

A cada item administrado, novas probabilidades devem ser atribuídas aos itens restantes, cuja probabilidades dos itens somam 1, o que não impossibilita a continuação do teste caso a estrutura do banco de dados seja modificada. Caso haja identificação de um erro ou violação de algum item, ele pode ser removido do banco ou substituído por outro e a informação ponderada pela taxa de exposição para os itens do novo banco são recalculadas e se adaptam automaticamente à nova formatação.

A próxima etapa consiste na estimação de  $\theta$  com base na resposta obtida naquele item e pode ser realizada via estimador bayesiano da média a posteriori (EAP) ou a moda a Posteriori (MAP), definidos anteriormente em (1.13) e (1.14).

Adotamos como critério de parada o desvio padrão posterior (PSD) associado ao EAP, estabelecido considerando o novo valor de  $\hat{\theta}_j$ , finalizando o teste caso o  $psd_j$  seja igual ou menor que o valor estabelecido previamente, sendo calculado com base na seguinte equação:

$$psd_j = \sqrt{E(\hat{\theta}_j^2) - [E(\hat{\theta}_j)]^2} \quad (2.2)$$

Definimos, também, um número máximo de itens para cada respondente, de modo que não alongue excessivamente o teste, fazendo com que o mesmo seja interrompido caso o examinado tenha sido exposto à quantidade máxima de itens, mesmo se a estimativa da proficiência não tiver atingido o valor de parada para o PSD.

## 2.1 Medidas de desempenho

### 2.1.1 Fidelidade

Em estudos de simulação, a correlação entre o  $\theta$  real e a estimativa das habilidades individuais dos examinados é denominada de coeficiente de fidelidade, permitindo afirmar que o método com maior coeficiente é o mais confiável para inferência e tomada de decisão (Leung, Chang e Hau 2002, apud Urry, 1970; Vale Weiss, 1975).

### 2.1.2 Vício

O vício é capaz de analisar a precisão na estimativa da habilidade dos respondentes, sendo a diferença entre o valor esperado de  $\hat{\theta}_j$  e o próprio parâmetro  $\theta_j$ .

Obtido por meio da seguinte equação:

$$\frac{1}{n} \sum_{j=1}^n (\hat{\theta}_j - \theta_j) \quad (2.3)$$

Onde  $\theta_j$   $j = 1, \dots, n$  são as verdadeiras habilidades dos  $n$  examinandos e  $\hat{\theta}_j$  seus respectivos estimadores obtidos por meio do CAT.

O método ideal é aquele com maior redução dos vícios, obtendo assim o valor do vício igual a 0, indicando que a estimação realizada corresponde ao valor real da habilidade. Quando obtém-se um valor positivo para o vício, isto indica que a estimativa da habilidade obtida por meio do respectivo método foi maior que a estimativa real do respondente, o inverso se aplica a um valor negativo, pois tem-se que um vício negativo corresponde à uma habilidade estimada menor que a habilidade real do respondente.

### 2.1.3 Erro Quadrático Médio

O EQM é frequentemente usado para verificar a acurácia de estimadores de modelos e dá um maior peso aos maiores erros, pois para o seu cálculo, cada erro é elevado ao quadrado individualmente e posteriormente, a média desses erros quadráticos é calculada, como demonstrado na equação abaixo:

$$\frac{1}{n} \sum_{j=1}^n (\hat{\theta}_j - \theta_j)^2 \quad (2.4)$$

Sendo definido como a média da diferença entre o valor do estimador e do parâmetro ao quadrado. Onde:

- $\theta_j$  é a habilidade real dos examinandos
- $\hat{\theta}_j$  é a habilidade estimada
- $n$  é o número total de respondentes

Esta métrica reflete a precisão da medição, ou seja, ela indica o quanto alguns resultados passam a se afastar da habilidade real dos respondentes. Considera-se o melhor método aquele com menor EQM, pois tende a ter um menor afastamento do valor real das habilidades dos respondentes.

#### **2.1.4 Número de itens superexpostos**

Nesta pesquisa os itens são considerados superexpostos se tiverem uma taxa de exposição maior que 0,2. Esta taxa é obtida por meio da razão entre o número de vezes que o item é administrado sobre o número total de examinandos que fazem o teste. Itens com alta taxa de exposição são um problema para segurança e validade do teste, pois estes itens podem ser de conhecimento prévio do examinado. Portanto, o menor número de itens superexpostos, é uma das formas de análise do melhor método.

#### **2.1.5 Número de itens subutilizados**

Assim como os itens superexpostos, os itens subutilizados também trazem impacto negativo para o método. Os itens não utilizados durante a aplicação dos testes geram apenas custos e não fornecem informação sobre os respondentes, afetando diretamente a relação custo-benefício do conjunto de itens. Para um item ser considerado subutilizado sua taxa de exposição precisa ser inferior a 0,02.

#### **2.1.6 Taxa de sobreposição**

A taxa de sobreposição representa a proporção de itens que dois ou mais examinandos compartilham entre si em um teste. O melhor método é aquele com menor taxa de sobreposição, indicando que todos os itens do banco foram utilizados de maneira uniforme. Altos valores referentes a esta taxa apontam o compartilhamento de informações entre examinandos que realizam o teste em diferentes ocasiões.

### **2.2 Simulações e Aplicações**

Foi utilizado o *software R* na construção do *script* da nova metodologia proposta e comparação com os métodos MI, Randomesque e SH, que também serão implementados no mesmo ambiente. Tendo como referência a pesquisa de Leung, Chang e Hau (2002) serão utilizadas medidas de desempenho similares às do estudo citado, sendo elas: fidelidade, vício, erro quadrático médio, desvio padrão, proporção de itens expostos, número de itens superexpostos, número de itens subutilizados e taxa de sobreposição.

### 2.2.1 Banco de itens simulado

Na composição do banco de itens, foram simulados os valores dos parâmetros de 400 itens com base nas seguintes distribuições:

$$b_i \sim Normal(0, 1)$$

$$a_i \sim Uniforme(0.3, 2.5)$$

$$c_i = 0.2$$

para todo  $i = 1, \dots, 400$ .

Foram realizadas diferentes simulações para analisar o comportamento dos métodos propostos, através das duas formas de seleção e relacionar com as medidas de desempenho dos métodos de seleção de itens citados anteriormente. Para análise da máxima informação ponderada foram considerados valores de  $\alpha$  de 1 a 8, com objetivo de apresentar uma perspectiva mais abrangente e representar o impacto de cada valor de  $\alpha$ , sendo desconsiderados valores acima deste intervalo por apresentarem uma quantidade de itens expostos por respondente acima do desejado. Para o método de informação ponderada aleatória foi levantada uma perspectiva mais reduzida, considerando apenas valores de  $\alpha$  no intervalo de 1 a 3.

Neste estudo foram administrados 15 testes, sendo eles, três métodos de seleção de itens já existentes: Máxima Informação, Randomesque e Sympson e Hetter, oito para máxima informação ponderada com diferentes valores de  $\alpha$ , três para informação aleatória ponderada e um com máxima informação ponderada com  $\alpha$  aleatório. Para cada aplicação foram armazenadas as informações de 20 conjuntos de dados, constituídos por 400 itens e administradas a 3000 respondentes, com distribuição  $N(0, 1)$  para a proficiência real. Foram realizadas 20 simulações por método e foi calculado o erro padrão de cada medida apresentada.

Tabela 2.1: Tabela resumo dos métodos comparados

Nome do método	Sigla utilizada
Máxima Informação	MI
Randomesque	RA
Sympson e Hetter	SH
Máxima Informação Ponderada	MIP
Informação Ponderada Aleatória	IPA

Levando em consideração as vantagens destacadas no Capítulo 1, será utilizada a Média a Posteriori (EAP) para realizar as estimações do método proposto. Com uma estimativa inicial de  $\theta = 0$ , haja visto, que é o valor central na escala da distribuição  $N(0,1)$ , assumida como a distribuição de  $\theta$  na população. Para o método de máxima informação foi necessário aplicar a aleatorização dos 6 itens iniciais (3 com  $-1 < b < 0$ , e 3 com  $0 < b < 1$ , todos com  $a < 1$ ), para que não houvesse um item inicial com taxa de exposição igual a 1.

Os resultados aqui apresentados são desenvolvidos em termos de precisão do teste e uso do banco de itens e seguem estrutura similar aos estudos realizados por Leung, Chang e Hau (2002) tendo como critério de parada o número máximo de 60 itens disponíveis para cada respondente ou PSD inferior a 0.1 . Diversas figuras são apresentadas para análise do desempenho das medidas utilizadas.

O desempenho dos métodos comparativos são apresentados na Tabela 2.2, indicando vantagens em diferentes medidas de desempenho. Sympson e Hetter tem a menor taxa de sobreposição, indicando uma melhor uniformização na distribuição dos itens, maior proporção de itens expostos, utilizando maior quantidade de itens do banco e proporção de itens superexpostos menor que os demais (0.02), enquanto o método Randomesque (0.07) apresenta mais que o triplo do valor. Os três métodos de seleção forneceram estimativas altamente correlacionadas para habilidades individuais com fidelidade em torno de 0.9 e vício em torno de 0. Apesar do Randomesque fazer uso de uma quantidade maior de itens, isso não causou impacto no erro quadrático médio das estimativas das proficiências, mantendo valor similar aos demais métodos que utilizam uma quantidade menor de itens, entretanto seu desvio padrão médio do número de itens respondidos por respondente é menor.

Nas tabelas de resultados a taxa de sobreposição está representada em porcentagem, é o percentual médio de itens que dois ou mais examinandos compartilham entre si em um teste, analisado com base em diversos pares de respondentes, comumente respondentes que tem valores de  $\theta$  semelhantes tendem a ter mais itens em comum.

Nas Tabelas 2.3 e 2.4 encontram-se as medidas de desempenho do método da máxima informação ponderada para diferentes valores de  $\alpha$  com seu respectivo erro padrão. O método da informação ponderada aleatória tem suas medidas apresentadas na Tabela 2.5, seguindo a mesma estrutura das anteriores. Quando comparamos os resultados obtidos nas Tabelas 2.2 e 2.5, é observado que o método da informação ponderada aleatória tem em média o mesmo percentual de sobreposição que SH (0.28%). RA e MI tem 0.46% e 0.47% respectivamente, enquanto a máxima informação ponderada diminui seu percentual a medida que o  $\alpha$  aumenta, com valores entre 0.34% e 0.19%. Em testes adaptativos computadorizados é comum ter taxa de sobreposição inferior a 1%.

Pela Tabela 2.3 é possível notar que, assim como os demais, o método da máxima informação ponderada não apresentou bons resultados para proficiência abaixo de -1.5, obtendo estimativas melhores no intervalo de -1.5 a 1.5 e acima de 1.5, com melhor desempenho para proficiências mais altas. O erro quadrático médio das estimativas ficou em torno de 0, 10 para todos os métodos e valores de  $\alpha$  utilizados.

Houve uma melhora significativa na proporção de itens expostos a medida que os valores de  $\alpha$  foram aumentando para o método da máxima informação ponderada, esta análise fica mais evidente ao comparar  $\alpha = 1$ , presente na Tabela 2.3 com o valor de  $\alpha = 8$ , encontrado na Tabela 2.4. Ainda que o resultado para  $\alpha = 1$  tenha valor similar ao obtido por SH, considerando  $\alpha = 8$  o método tem melhor desempenho que os métodos comparativos, com o dobro do valor de exposição em relação a máxima informação e randomesque. Contudo, o método da informação ponderada aleatória apresentou melhor desempenho, com proporção igual a 1, fazendo uso de todos os itens que compõem o banco.

Como definido no tópico 2.1.5 itens subutilizados são aqueles aplicados a menos de 2% dos respondentes sendo alocados a eles. Neste campo os métodos de comparação (Tabela 2.2) apresentaram valores em torno de 0.8, enquanto os métodos propostos tem resultados variando entre 0.08 a 0.7 para diferentes valores de  $\alpha$ , o que demonstra ter menos itens subutilizados, ou seja, divide melhor a distribuição dos itens. No método de informação ponderada aleatória a maior proporção de itens subutilizados é de 0.1.

Tabela 2.2: Resultados obtidos pelos métodos Simpson-Hetter, Máxima Informação e Rando-  
mesque

	MI	R	SH
Taxa de sobreposição	0.4740%	0.4599%	0.2842%
	(0.00697)	(0.00715)	(0.00471)
Prop. de itens expostos	0.24825	0.29388	0.37350
	(0.01000)	(0.00829)	(0.01387)
Prop. de itens superexpostos	0.06475	0.07237	0.02000
	(0.00291)	(0.00172)	(0.00956)
Proporção de itens subutilizados	0.82337	0.79363	0.72500
	(0.00233)	(0.00319)	(0.00303)
Fidelidade	0.94984	0.95093	0.95024
	(0.00132)	(0.00197)	(0.00178)
Valor médio do vício	-0.00187	0.00081	-0.00081
	(0.00580)	(0.00566)	(0.00348)
Valor médio do vício para proficiências $\leq -1,5$	0.21251	0.21938	0.21327
	(0.02398)	(0.02096)	(0.01753)
Valor médio do vício para proficiências entre -1,5 e 1,5	-0.00302	-0.00165	-0.00191
	(0.00583)	(0.00545)	(0.00359)
Valor médio do vício para proficiências $>1,5$	-0.01678	-0.01482	-0.01602
	(0.00565)	(0.00572)	(0.00391)
EQM das estimativas das proficiências	0.09703	0.09552	0.09779
	(0.00208)	(0.00377)	(0.00286)
EQM das estimativas para as proficiências $\leq -1,5$	0.15961	0.16423	0.15456
	(0.02086)	(0.01989)	(0.01446)
EQM para as proficiências entre -1,5 e 1,5	0.08963	0.08750	0.09051
	(0.00183)	(0.00296)	(0.00243)
EQM para as proficiências $> 1,5$	0.09265	0.09057	0.09371
	(0.00218)	(0.00324)	(0.00245)
Prop. de respondentes expostos ao núm. máximo de itens	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)
Núm. médio de itens respondidos por respondente	19.73912	15.16923	16.32322
	(0.03802)	(0.05110)	(0.04356)
DP médio do núm. de itens respondidos por respondente	2.53591	3.04090	2.75032
	(0.06738)	(0.06313)	(0.10560)

O método da informação ponderada aleatória para todos os valores de  $\alpha$  e máxima informação ponderada com  $\alpha$  superior a 4 não apresentam taxa de superexposição, indicando que para estes casos menos de 20% dos respondentes tem acesso ao item. Já para os métodos de comparação seus valores variam entre 0.02 para SH, 0.06 para MI e 0.07 para RA.

A fidelidade para todos os métodos aqui citados ficaram em torno de 0.95, não sendo considerado um método para seleção do melhor método, por não apresentar nenhuma diferença significativa.

O número médio de itens respondidos por examinado deixou clara as diferenças entre os métodos. A informação ponderada aleatória exige a utilização de 29 itens para atingir o critério especificado, um número maior de itens por respondentes, quando comparado aos outros métodos citados nesta pesquisa, enquanto os demais produziram testes mais curtos, necessitando de 10 à 15 itens até a convergência. Em uma perspectiva de aplicação de testes em massa, a diferença entre a quantidade de itens por pessoa, gera consequências para o banco de dados, pois é imprescindível distribuí-los de forma homogênea a fim de manter esta taxa abaixo de 0.20 que é um valor clássico utilizado nas referências aqui citadas. Vale ressaltar que, ainda que a taxa não ultrapasse o valor estabelecido, mesmo gastando aproximadamente 15 itens a mais para convergir, o teste torna-se desgastante para o respondente, aumentando assim o tempo de prova.

Tabela 2.3: Resultados obtidos pelo método da Máxima Informação Ponderada

	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$
Taxa de sobreposição	0.3442% (0.00487)	0.2756% (0.00423)	0.2454% (0.00394)	0.2271% (0.00511)
Prop. de itens expostos	0.31825 (0.01133)	0.38512 (0.01510)	0.42288 (0.01168)	0.47400 (0.00720)
Prop. de itens superexpostos	0.07550 (0.00192)	0.06012 (0.00190)	0.03825 (0.00118)	0.01700 (0.00112)
Proporção de itens subutilizados	0.76262 (0.00309)	0.69812 (0.00179)	0.67063 (0.00362)	0.63050 (0.00411)
Fidelidade	0.94966 (0.00198)	0.94971 (0.00181)	0.95057 (0.00200)	0.95097 (0.00096)
Valor médio do vício	-0.00214 (0.00537)	0.00041 (0.00622)	-0.00017 (0.00406)	-0.00243 (0.00716)
Valor médio do vício para proficiências $\leq -1,5$	0.21396 (0.01787)	0.22609 (0.02307)	0.22161 (0.02463)	0.19908 (0.02838)
Valor médio do vício para proficiências entre -1,5 e 1,5	-0.00362 (0.00631)	-0.00196 (0.00601)	-0.00248 (0.00502)	-0.00424 (0.00752)
Valor médio do vício para proficiências $> 1,5$	-0.01729 (0.00581)	-0.01548 (0.00594)	-0.01607 (0.00445)	-0.01734 (0.00738)
EQM das estimativas das proficiências	0.09719 (0.00261)	0.09770 (0.00268)	0.09721 (0.00320)	0.09644 (0.00317)
EQM das estimativas para as proficiências $\leq -1,5$	0.16639 (0.02188)	0.17028 (0.01893)	0.16257 (0.02857)	0.15536 (0.02083)
EQM para as proficiências entre -1,5 e 1,5	0.08932 (0.00242)	0.08970 (0.00211)	0.08951 (0.00285)	0.08950 (0.00241)
EQM para as proficiências $> 1,5$	0.09234 (0.00243)	0.09258 (0.00245)	0.09253 (0.00296)	0.09206 (0.00173)
Prop. de respondentes expostos ao núm. máximo de itens	0.00000 (0.00000)	0.00000 (0.00000)	0.00000 (0.00000)	0.00000 (0.00000)
Núm. médio de itens respondidos por respondente	15.16567 (0.04128)	15.65465 (0.04696)	16.06202 (0.05584)	16.45367 (0.08664)
DP médio do núm. de itens respondidos por respondente	2.83984 (0.07843)	2.77374 (0.07934)	2.80263 (0.07896)	2.87673 (0.07523)

Tabela 2.4: Resultados obtidos pelo método da Máxima Informação Ponderada

	$\alpha = 5$	$\alpha = 6$	$\alpha = 7$	$\alpha = 8$
Taxa de sobreposição	0.2108%	0.2045%	0.1917%	0.1863%
	(0.00180)	(0.00370)	(0.00249)	(0.00125)
Prop. de itens expostos	0.52850	0.55900	0.58650	0.60300
	(0.00180)	(0.00945)	(0.00912)	(0.01052)
Prop. de itens superexpostos	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Proporção de itens subutilizados	0.58450	0.55100	0.51700	0.48950
	(0.00274)	(0.00285)	(0.00326)	(0.00209)
Fidelidade	0.95108	0.95028	0.94976	0.95063
	(0.00118)	(0.00170)	(0.00376)	(0.00235)
Valor médio do vício	-0.00081	-0.00182	-0.00174	-0.00250
	(0.00746)	(0.01072)	(0.01047)	(0.00740)
Valor médio do vício para proficiências $\leq -1,5$	0.20131	0.22913	0.20412	0.21109
	(0.02989)	(0.01632)	(0.01541)	(0.01704)
Valor médio do vício para proficiências entre -1,5 e 1,5	-0.00132	-0.00380	-0.00283	-0.00477
	(0.00706)	(0.01174)	(0.00950)	(0.00669)
Valor médio do vício para proficiências $> 1,5$	-0.01529	-0.01854	-0.01731	-0.01870
	(0.00708)	(0.01002)	(0.00989)	(0.00635)
EQM das estimativas das proficiências	0.09641	0.09820	0.09883	0.09701
	(0.00360)	(0.00080)	(0.00400)	(0.00220)
EQM das estimativas para as proficiências $\leq -1,5$	0.15320	0.17147	0.15706	0.15116
	(0.02500)	(0.01696)	(0.02352)	(0.01226)
EQM para as proficiências entre -1,5 e 1,5	0.08931	0.08968	0.09174	0.09000
	(0.00242)	(0.00269)	(0.00447)	(0.00290)
EQM para as proficiências $> 1,5$	0.09229	0.09291	0.09443	0.09287
	(0.00224)	(0.00181)	(0.00411)	(0.00194)
Prop. de respondentes expostos ao núm. máximo de itens	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Núm. médio de itens respondidos por respondente	16.81413	17.18953	17.52733	17.87053
	(0.05560)	(0.05851)	(0.04827)	(0.06059)
DP médio do núm. de itens respondidos por respondente	2.89211	2.96321	3.03958	3.09944
	(0.07405)	(0.12104)	(0.05578)	(0.08960)

Tabela 2.5: Resultados obtidos pelo método da Informação Ponderada Aleatória

	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
Taxa de sobreposição	0.2878%	0.2863%	0.2862%
	(0.00263)	(0.00371)	(0.00369)
Prop. de itens expostos	1.00000	1.00000	1.00000
	(0.00000)	(0.00000)	(0.00000)
Prop. de itens superexpostos	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)
Proporção de itens subutilizados	0.11000	0.09700	0.08388
	(0.00757)	(0.00759)	(0.00547)
Fidelidade	0.94909	0.94870	0.94899
	(0.00179)	(0.00161)	(0.00173)
Valor médio do vício	-0.00140	-0.00099	-0.00187
	(0.00533)	(0.00626)	(0.00533)
Valor médio do vício para proficiências $\leq -1,5$	0.21708	0.22084	0.21703
	(0.02176)	(0.02927)	(0.02220)
Valor médio do vício para proficiências entre -1,5 e 1,5	-0.00252	-0.00248	-0.00150
	(0.00509)	(0.00596)	(0.00567)
Valor médio do vício para proficiências $>1,5$	-0.01708	-0.01680	-0.01716
	(0.00591)	(0.00626)	(0.00573)
EQM das estimativas das proficiências	0.09888	0.09887	0.10008
	(0.00259)	(0.00332)	(0.00266)
EQM das estimativas para as proficiências $\leq -1,5$	0.16125	0.16527	0.16936
	(0.01846)	(0.02131)	(0.01880)
EQM para as proficiências entre -1,5 e 1,5	0.09039	0.09053	0.09112
	(0.00250)	(0.00252)	(0.00277)
EQM para as proficiências $> 1,5$	0.09441	0.09414	0.09523
	(0.00299)	(0.00254)	(0.00233)
Prop. de respondentes expostos ao núm. máximo de itens	0.00257	0.00255	0.00308
	(0.00077)	(0.00102)	(0.00122)
Núm. médio de itens respondidos por respondente	28.76288	29.09897	29.44593
	(0.11301)	(0.15030)	(0.12477)
DP médio do núm. de itens respondidos por respondente	6.78111	6.74987	6.78469
	(0.13963)	(0.13860)	(0.15741)

Como perspectiva alternativa ao método da máxima informação ponderada, foi selecionado um valor aleatório de  $\alpha$  com distribuição Uniforme no intervalo [0.5, 8.5] para cada item a ser escolhido, como possibilidade para cenários em que o pesquisador não tenha que determinar um valor para  $\alpha$ . Os resultados obtidos por meio deste método estão descritos na Tabela 2.6, com o respectivo erro padrão para cada medida de desempenho.

Este método optativo apresentou valores dentro do esperado, com comportamento similar ao obtido por SH, ainda com algumas vantagens sobre o mesmo, como maior proporção de itens expostos e menor taxa de sobreposição, sem alterar de forma significativa o número médio de itens respondidos.

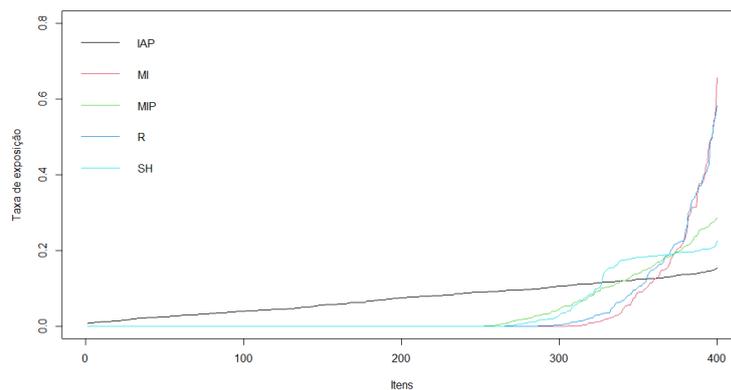
Tabela 2.6: Resultados obtidos pelo método da Máxima Informação Ponderada com  $\alpha$  aleatório

	$\alpha$ aleatório
Taxa de sobreposição	0.2128% (0.00431)
Prop. de itens expostos	0.56450 (0.01037)
Prop. de itens superexpostos	0.01950 (0.00209)
Prop. de itens subutilizados	0.54500 (0.00500)
Fidelidade	0.95039 (0.00166)
Valor médio do vício	-0.00102 (0.00776)
Valor médio do vício para proficiências $\leq -1,5$	0.20969 (0.02060)
Valor médio do vício para proficiências entre -1,5 e 1,5	-0.00315 (0.00636)
Valor médio do vício para proficiências $> 1,5$	-0.01611 (0.00639)
EQM das estimativas das proficiências	0.09693 (0.00222)
EQM das estimativas para as proficiências $\leq -1,5$	0.15327 (0.01605)
EQM para as proficiências entre -1,5 e 1,5	0.08971 (0.00217)
EQM para as proficiências $> 1,5$	0.09293 (0.00218)
Prop. de respondentes expostos ao núm. máximo de itens	0.00000 (0.00000)
Núm. médio de itens respondidos por respondente	17 (0.04269)
DP médio do núm. de itens respondidos por respondente	2.85871 (0.04343)

A Figura 2.1 ilustra o comportamento da taxa de exposição dos itens classificada em ordem crescente, para todos os métodos. Esta medida foi calculada através da relação da proporção

de itens respondidos sobre o número de respondentes. Entre os cinco métodos, três apresentam melhor comportamento, sendo eles IPA, MIP e SH, suas taxas tem valor inferior a 0.4 e apresentam comportamento crescente a partir de 260 itens. No entanto, como já foi visto anteriormente, o IPA necessita de mais itens para convergência. É importante destacar que para a construção do gráfico foram considerados valores de  $\alpha = 2$  para MIP e  $\alpha = 3$  para IPA.

Figura 2.1: Taxa de exposição dos itens para todos os métodos



Fonte: Elaborado pelo autor (2022).

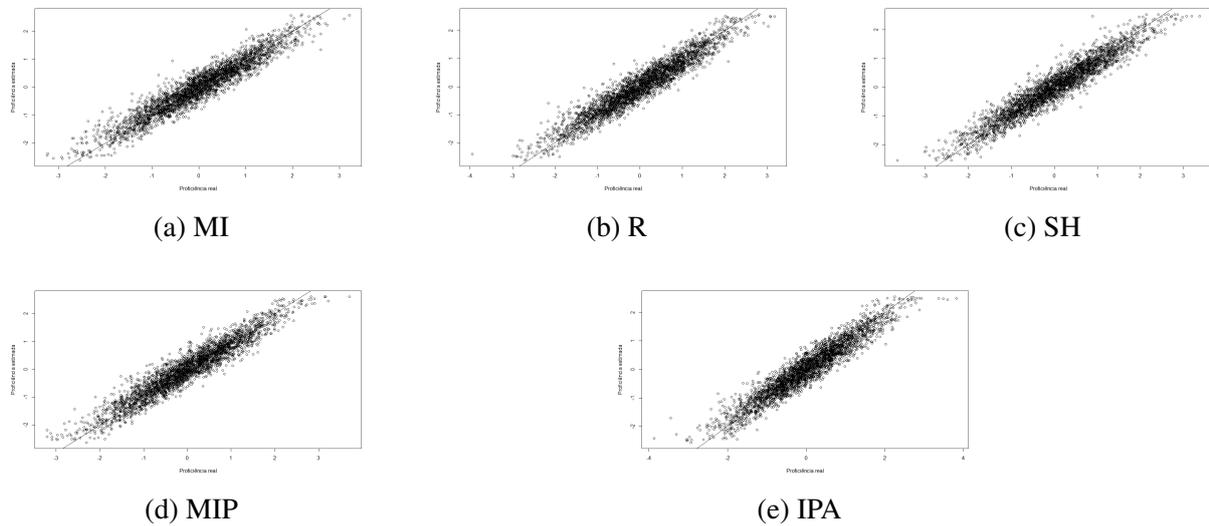
Todos os métodos aqui apresentados forneceram estimativas adequadas para as habilidades dos respondentes como mostra a Figura 2.2, onde é possível notar que os pontos se encontram em torno da reta da proficiência rela dos respondentes, especialmente em torno de  $-1,5$  a  $1,5$ , havendo leves discrepâncias nas caudas das distribuições, com maior ênfase em indivíduos com habilidades inferiores.

Por meio da Figura 2.3 é possível observar que os parâmetros de dificuldade se distribuem normalmente entre a proporção de respondentes expostos a cada item, ou seja, o parâmetro de dificuldade se distribui dentro da população de forma similar a curva normal, indicando que alguns respondentes foram expostos a itens com baixo nível de dificuldade, outros a um nível de dificuldade alto, e a maioria encontra-se em um nível intermediário. Os modelos citados têm maior proporção de respondentes expostos a cada item no mesmo intervalo (de  $-1.5$  a  $1.5$ ) em que as habilidades têm estimativas mais precisas.

A influência do parâmetro de discriminação na proporção de respondentes expostos a cada item tem comportamento distinto para os métodos aplicados. Esse comportamento se dá devido a itens com maior discriminação serem utilizados com frequência durante a aplicação dos métodos MI, R, SH e MIP. A aleatorização presente na seleção do item por meio da IPA faz com que a proporção de respondentes expostos a cada item tenha valor diferente de 0, para todos os parâmetros de discriminação e valor máximo igual a 0,15.

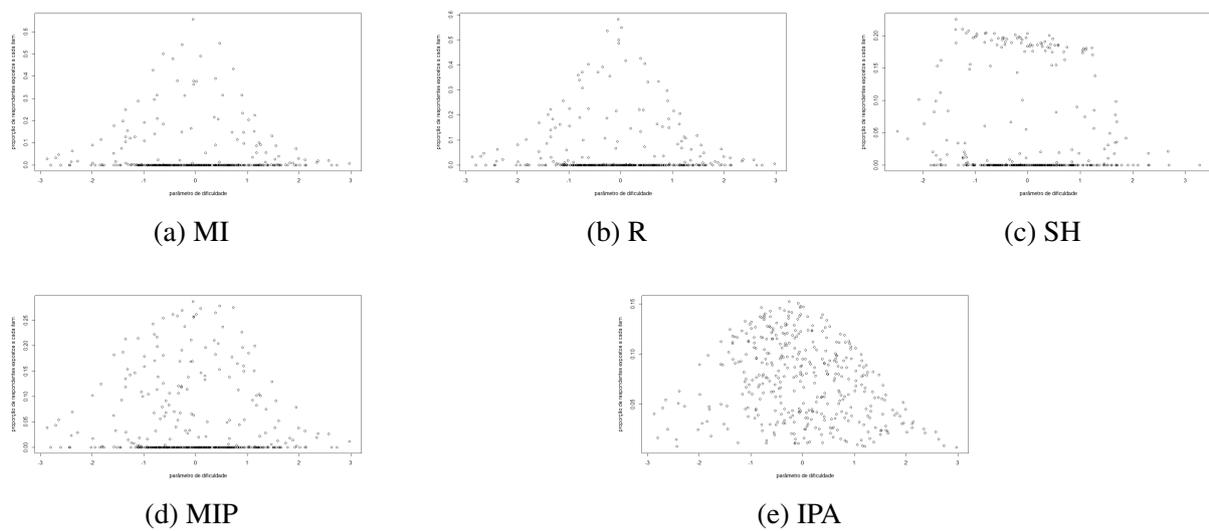
Entretanto se for levado em consideração a quantidade de itens utilizados, como mostra a Figura 2.4, os melhores desempenhos são apresentados pela MIP e SH que apresentam proporções de respondentes expostos a cada item variando entre 0 e 0,3, em sua grande maioria para

Figura 2.2: Proficiência real vs Proficiência estimada



Fonte: Elaborado pelo autor (2022).

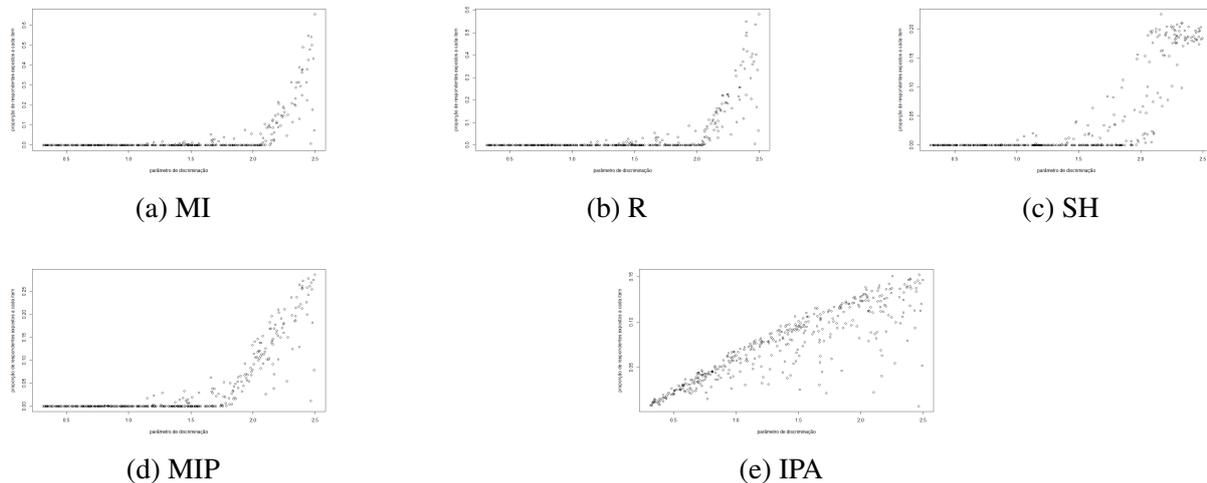
Figura 2.3: Proporção de respondentes expostos a cada item vs Parâmetro de dificuldade



Fonte: Elaborado pelo autor (2022).

discriminação  $\geq 1,5$ . Os demais métodos têm proporção máxima em torno de 0.6 e usam com maior frequência itens com discriminação  $\geq 2,0$ .

Figura 2.4: Proporção de respondentes expostos a cada item vs Parâmetro de discriminação



Fonte: Elaborado pelo autor (2022).

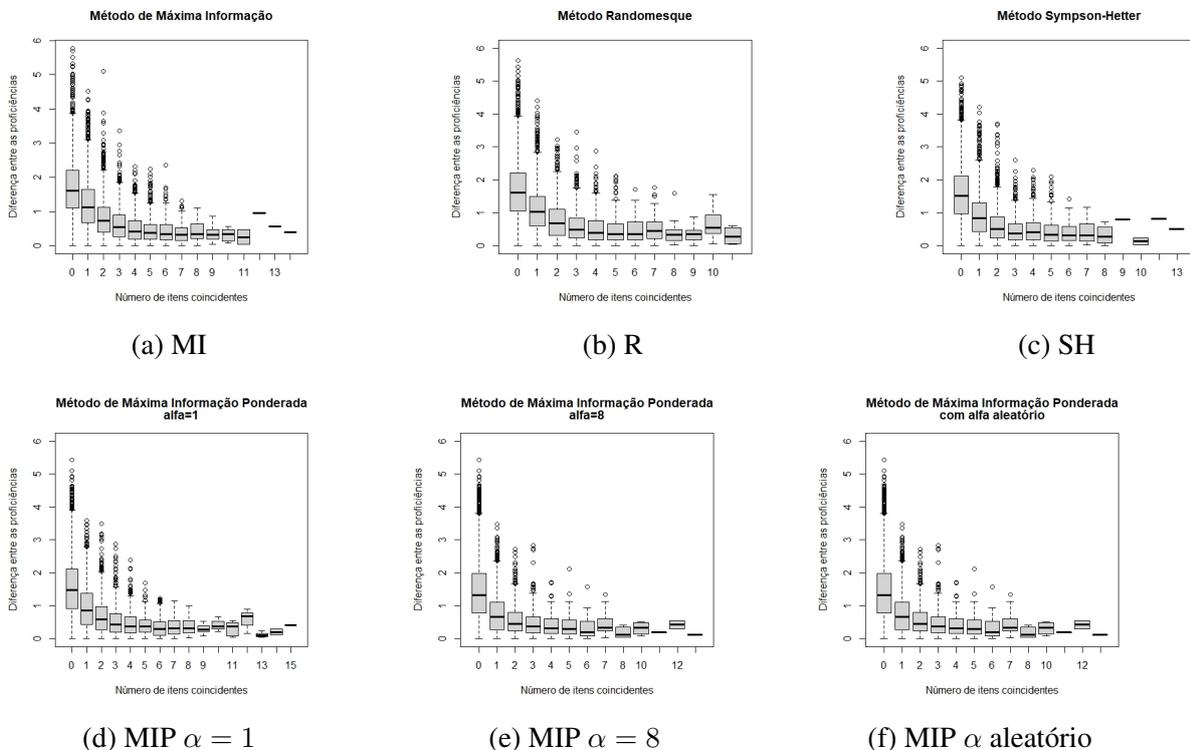
Para a construção de boxplots da diferença entre as proficiências de pares de respondentes plotados em função do número de itens em comum entre dois respondentes, foi selecionado um conjunto de 3000 respondentes e uma amostra de 10000 combinações aleatórias para verificar a coincidência entre os itens. Na Figura 2.5, analisou-se a diferença entre as proficiências a cerca dos pares de respondentes, com variação de 0 até próximo de 6, ou seja, temos variação máxima quando um respondente está em torno de  $-3$  e o outro em torno de  $3$ .

Considerando a diferença das proficiências no intervalo  $0-1$  o número de itens coincidentes pode chegar até 15 para o MIP com  $\alpha = 1$  e a medida que a diferença da proficiência vai aumentando o número de itens coincidentes é sempre baixo, muitas vezes igual 0.

Quando comparamos o método de máxima informação ponderada para diferentes valores de  $\alpha$  e o método de máxima informação, as diferenças entre as proficiências dos respondentes tendem a se concentrar em valores um pouco menores para a MIP. Isto é uma vantagem do método de máxima informação ponderada pois é necessário ter as proficiências dos respondentes mais próximas entre si para ter o mesmo número de itens em comum nos dois métodos. Pelos gráficos, o método de máxima informação ponderada tem desempenho levemente superior aos métodos de máxima informação e randomesque, e comportamento similar ao método de Simpson-Hetter.

Pares de respondentes que não apresentam itens coincidentes tem maior variabilidade e valores discrepantes, alcançando a diferença máxima entre as proficiências. Consequentemente, pares com menor diferença ou com a mesma proficiência, tendem a responder uma maior quantidade de itens em comum.

Figura 2.5: Diferença entre as proficiências vs Número de itens coincidentes



Fonte: Elaborado pelo autor (2022).

## 2.2.2 Aplicação dos métodos a um banco de itens real

Adicionalmente, foram realizadas simulações utilizando o banco de itens da prova de Proficiência em Inglês da Universidade de Brasília (UnB), composto por nove provas anteriormente aplicadas pela universidade, por intermédio do CEBRASPE – Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos que é responsável pela realização de seleções, avaliações e certificações no Brasil. Esta prova é oferecida semestralmente a seus alunos, com intuito de atribuir os créditos referentes às disciplinas Inglês Instrumental 1 e/ou Inglês Instrumental 2, caso sejam aprovados. Sua aplicação por meio do CAT teve início no primeiro semestre de 2010 (Labarrere, da-Silva e Costa, 2011).

Nesta etapa o banco será composto por 422 itens com seus respectivos parâmetros, calibrados de forma conjunta, pois haviam itens em comum entre as provas, os três parâmetros do modelo foram estimados pelo modelo Normal. As respostas utilizadas serão simuladas, assim como, a proficiência dos 3000 examinados, com distribuição Normal (0,1), sendo a mesma utilizada anteriormente, permanecendo com valor de  $\theta = 0$  para estimativa inicial. Com critério de parada definido pelo desvio padrão a posteriori = 0.1 e restrição de 40 itens disponíveis

para cada respondente, ou seja, caso o critério de parada não fosse satisfeito, respondia-se, no máximo, 40 itens.

Para análise da máxima informação ponderada foram considerados  $\alpha = 1, 3, 5$  e  $7$ , já para o método de informação ponderada aleatória utilizou-se  $\alpha$  no intervalo de  $1$  a  $3$ , utilizando o estimador bayesiano da média a posteriori (EAP) em ambos os métodos. Totalizando 11 testes, sendo eles, quatro para máxima informação ponderada com diferentes valores de  $\alpha$ , três para informação aleatória ponderada, um com máxima informação ponderada com  $\alpha$  aleatório e três métodos comparativos (Máxima Informação, Randomesque e Sympson e Hetter). Por ter sido realizada apenas uma simulação para cada teste, não é aplicável o calculo do erro padrão para as medidas de desempenho.

A tabela 2.7 apresenta o resultado da aplicação dos métodos comparativos a um banco de itens real, permitindo afirmar que seu desempenho foi inferior aos resultados obtidos anteriormente, apresentando  $0.57\%$  como a menor taxa de sobreposição, ou seja, o método SH segue tendo melhor uniformização na distribuição dos itens, entretanto com o dobro do valor apresentado anteriormente, com diferença de  $0.03\%$  e  $0.10\%$  aproximadamente quando comparado aos outros. O desempenho deste método segue se destacando ao apresentar menor proporção de itens superexpostos e subutilizados, além de ter maior proporção de itens expostos, fazendo uso de uma quantidade maior de itens do banco. No entanto a média de itens respondidos por respondente fica em torno de  $34$ , valor este, próximo do número máximo de itens disponíveis por respondente, impactando na proporção de respondentes expostos ao número máximo de itens ( $0.2$ ), sendo o maior entre os três.

Todos os métodos apresentaram fidelidade em torno de  $0.9$ , fornecendo estimativas altamente correlacionadas com o valor real das habilidades dos respondentes e vício em torno de zero para proficiências  $> -1, 5$ . O EQM para as três aplicações se manteve em torno de  $0.1$  para os diferentes intervalos analisados, com desvio padrão médio do número de itens respondido por respondente variando entre  $5$  e  $7$  aproximadamente.

Tabela 2.7: Resultados obtidos pelos métodos de Sympson-Hetter, Máxima Informação e Randomesque

	Max. Info.	Random.	SH
Taxa de sobreposição	0.60474%	0.67085%	0.57457%
Prop. de itens expostos	0.29384	0.32227	0.55924
Prop. de itens superexpostos	0.07820	0.08531	0.03081
Prop. de itens subutilizados	0.76066	0.74171	0.50237
Fidelidade	0.93048	0.93459	0.94697
Valor médio do vício	0.00212	0.00737	-0.00162
Valor médio do vício para proficiências $\leq -1,5$	0.22247	0.24215	0.22042
Valor médio do vício para proficiências entre $-1,5$ e $1,5$	0.00346	0.00415	0.00155
Valor médio do vício para proficiências $> 1,5$	-0.01319	-0.00895	-0.01496
EQM das estimativas das proficiências	0.13702	0.12923	0.09953
EQM das estimativas para as proficiências $\leq -1,5$	0.19015	0.18911	0.14910
EQM para as proficiências entre $-1,5$ e $1,5$	0.12813	0.12044	0.09276
EQM para as proficiências $> 1,5$	0.13333	0.12506	0.09655
Prop. de respondentes expostos ao núm. máximo de itens	0.01600	0.00900	0.22833
Núm. médio de itens respondidos por respondente	22.54333	18.63100	34.87900
DP médio do núm. de itens respondidos por respondente	6.76008	7.09710	5.07066

A Tabela 2.8 representa a frequência dos números de itens coincidentes para uma amostra de 10 mil pares de respondentes utilizando o método da Máxima informação, indicando que aproximadamente 29% dos pares de respondentes não apresentaram itens em comum, 56% da amostra tem menos de 4 itens coincidentes e pares de respondentes com 6 itens ou menos em comum representam cerca de 93% da amostra.

Tabela 2.8: Frequência dos números de itens coincidentes na amostra de 10 mil pares de respondentes para o método de Máxima Informação

Itens	Frequência								
0	2921	6	625	12	11	18	3	25	1
1	1419	7	353	13	17	19	2	28	1
2	1301	8	152	14	11	20	3	29	1
3	1135	9	62	15	10	21	1	33	1
4	1094	10	22	16	4	22	1	34	1
5	829	11	8	17	8	23	3		

Para o método Randomesque a quantidade de itens coincidentes teve variação entre 0 e 35, conforme Tabela 2.9, entretanto 56% foram expostos a no máximo 2 itens em comum, metade da quantidade de itens que o mesmo percentual da amostra respondeu no método anterior e apenas 1% dos pares tem mais de 10 itens coincidentes.

Pode-se observar na Tabela 2.10 que a menor amplitude foi obtida para o método Sympson e Hetter com no máximo 21 itens coincidentes para apenas um par de respondentes. 90% da sua amostra tem no máximo 5 itens em comum, onde 23% destes não compartilham de nenhum item e apenas 1% da amostra tem mais de 10 itens coincidentes. Com isso SH tem o melhor

Tabela 2.9: Frequência dos números de itens coincidentes na amostra de 10 mil pares de respondentes para o método Randomesque

Itens	Frequência								
0	2239	6	704	12	11	18	3	26	1
1	1550	7	391	13	17	19	4	27	1
2	1294	8	171	14	11	20	5	28	1
3	1307	9	46	15	10	22	1	30	2
4	1129	10	23	16	4	23	3	33	1
5	1053	11	17	17	8	24	4	35	2

desempenho relacionado a itens coincidentes por pares de respondentes.

Tabela 2.10: Frequência dos números de itens coincidentes na amostra de 10 mil pares de respondentes para o método Sympton-Hetter

Itens	Frequência	Itens	Frequência	Itens	Frequência	Itens	Frequência
0	2319	6	366	12	25	20	2
1	1932	7	205	13	18	21	1
2	1728	8	131	14	18		
3	1397	9	68	15	9		
4	1025	10	41	16	8		
5	678	11	21	17	8		

A Tabela 2.11 apresenta os resultados para o método da máxima informação ponderada, nela é possível observar que de modo geral o  $\alpha = 7$  apresentou melhor desempenho. Com menor taxa de sobreposição, vício e EQM. A fidelidade se manteve constante em torno de 0.93 para todos os valores de  $\alpha$  aplicados, enquanto o desvio padrão aumentou discretamente para os maiores valores de  $\alpha$ . Entretanto  $\alpha = 7$  necessita de uma quantidade maior de itens para estimação do  $\theta$ .

Tabela 2.11: Resultados obtidos pelo método da Máxima Informação Ponderada

	$\alpha = 1$	$\alpha = 3$	$\alpha = 5$	$\alpha = 7$
Taxa de sobreposição	0.5093%	0.4148%	0.3669%	0.35578
Prop. de itens expostos	0.36729	0.55213	0.67062	0.74645
Prop. de itens superexpostos	0.09715	0.09242	0.06872	0.04502
Prop. de itens subutilizados	0.69668	0.53791	0.41232	0.33886
Fidelidade	0.93192	0.93642	0.93417	0.93829
Valor médio do vício	-0.00229	0.00130	0.00334	-0.00524
Valor médio do vício para proficiências $\leq -1,5$	0.20255	0.22941	0.22357	0.20563
Valor médio do vício para proficiências entre -1,5 e 1,5	0.00260	-8.53e-06	0.00337	-0.00604
Valor médio do vício para proficiências $> 1,5$	-0.01653	-0.01455	0.01197	-0.01990
EQM das estimativas das proficiências	0.13432	0.12563	0.13044	0.12171
EQM das estimativas para as proficiências $\leq -1,5$	0.17073	0.15032	0.18706	0.17989
EQM para as proficiências entre -1,5 e 1,5	0.12386	0.11748	0.12272	0.11383
EQM para as proficiências $> 1,5$	0.13179	0.12392	0.12650	0.11768
Prop. de respondentes expostos ao núm. máximo de itens	0.01367	0.02167	0.04767	0.06400
Núm. médio de itens respondidos por respondente	19.11033	22.29600	25.01800	27.14100
DP médio do núm. de itens respondidos por respondente	7.27276	7.66172	8.22221	8.27937

A frequência dos números de itens coincidentes em uma amostra de 10 mil pares de respondentes para o método de máxima informação ponderada considerando os quatro valores de  $\theta$  será apresentada na Tabela 2.12, com a quantidade de itens coincidentes no intervalo de 0 a 33 para cada valor de  $\alpha$ . A quantidade de itens em comum diminui a medida que o valor de  $\alpha$  aumenta, na qual 90% dos respondentes tem acesso a no máximo 6 itens em comum quando utilizado  $\alpha = 1$ , para  $\alpha = 3$  tem-se no máximo 5 itens coincidentes considerando o mesmo percentual da amostra e no máximo 4 para  $\alpha = 5$  e  $\alpha = 7$ . Vale ressaltar que para  $\alpha = 5$  e  $\alpha = 7$ , mais da metade da amostra tem no máximo 1 item em comum. Apenas 1% dos pares de respondentes aproximadamente recebem testes com no mínimo 9 itens em comum utilizando  $\alpha = 1$ , para o restante dos valores de  $\alpha$  tem-se 10 como valor mínimo. Podendo concluir que maiores valores de  $\alpha$  seguem apresentando melhor desempenho para esse método.

A Tabela 2.13 contém os resultados para o método da informação ponderada aleatória, onde é possível observar valores muito similares para os três  $\alpha$ , com relação a taxa de exposição, obtendo valores em torno de 0.26 para todos os casos e utilização todos os itens do banco, representado pelo valor máximo (1.0) na proporção de itens expostos. Fazendo uso desta metodologia, nenhum item foi super exposto e obtiveram estimativas da habilidade do respondente altamente correlacionadas com a habilidade real do respondente para todos os  $\alpha$  utilizados, com EQM das estimativas das proficiências em torno de 0.1. Mantendo o valor médio do vício e a

Tabela 2.12: Frequência dos números de itens coincidentes na amostra de 10 mil pares de respondentes para o método de Máxima Informação Ponderada

	$\alpha = 1$	$\alpha = 3$	$\alpha = 5$	$\alpha = 7$
0	3750	4494	4837	4890
1	1598	1579	1668	1653
2	1069	1069	1135	1143
3	954	894	811	766
4	849	730	593	613
5	709	511	352	376
6	466	337	239	239
7	310	167	134	123
8	157	79	77	76
9	50	38	53	40
10	21	28	19	22
11	14	12	17	13
12	9	14	11	4
13	6	7	6	5
14	5	6	8	3
15	4	4	1	6
16	4	8	7	5
17	3	2	5	1
18	3	4	4	4
19	5	2	4	2
20	2	4	2	3
21	3	2	1	4
22	1	0	3	2
23	1	1	1	4
24	0	0	7	0
25	1	0	1	1
26	3	1	0	0
27	0	1	1	0
28	1	1	1	0
29	1	1	0	0
30	0	1	1	1
31	0	1	1	0
32	1	1	0	1
33	0	1	0	0

proporção de respondentes expostos ao número máximo de itens em torno de 0, apesar de fazer uso de 29 itens em média por respondente, com desvio padrão médio do número de itens respondidos em torno de 7.

Tabela 2.13: Resultados obtidos pelo método da Informação Ponderada Aleatória

	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
Taxa de sobreposição	0.2633%	0.2609%	0.2602%
Prop. de itens expostos	1.00000	1.00000	1.00000
Prop. de itens superexpostos	0.00000	0.00000	0.00000
Prop. de itens subutilizados	0.10900	0.09479	0.08531
Fidelidade	0.94821	0.94874	0.94849
Valor médio do vício	-0.00377	-0.00128	0.00112
Valor médio do vício para proficiências $\leq -1,5$	0.21151	0.21703	0.24398
Valor médio do vício para proficiências entre -1,5 e 1,5	-0.00535	0.00182	-0.00386
Valor médio do vício para proficiências $> 1,5$	-0.01874	-0.01646	-0.01577
EQM das estimativas das proficiências	0.10252	0.10152	0.10200
EQM das estimativas para as proficiências $\leq -1,5$	0.16286	0.17558	0.20987
EQM para as proficiências entre -1,5 e 1,5	0.09470	0.09027	0.09010
EQM para as proficiências $> 1,5$	0.09832	0.09637	0.09450
Prop. de respondentes expostos ao núm. máximo de itens	0.00300	0.00367	0.00333
Núm. médio de itens respondidos por respondente	29.14267	29.44167	29.59767
DP médio do núm. de itens respondidos por respondente	7.17222	7.23183	7.05273

Considerando um valor de  $\alpha$  aleatório, escolhido a partir da distribuição Uniforme no intervalo de  $[0.5, 7.5]$  para cada item a ser escolhido pela máxima informação ponderada temos um desempenho satisfatório do método, como mostra a Tabela 2.14, com baixa taxa de sobreposição (0.19%) e alta correlação entre a habilidade estimada e a habilidade real do respondente, com menor erro quadrático médio para proficiências entre  $[-1.5, 1.5]$ .

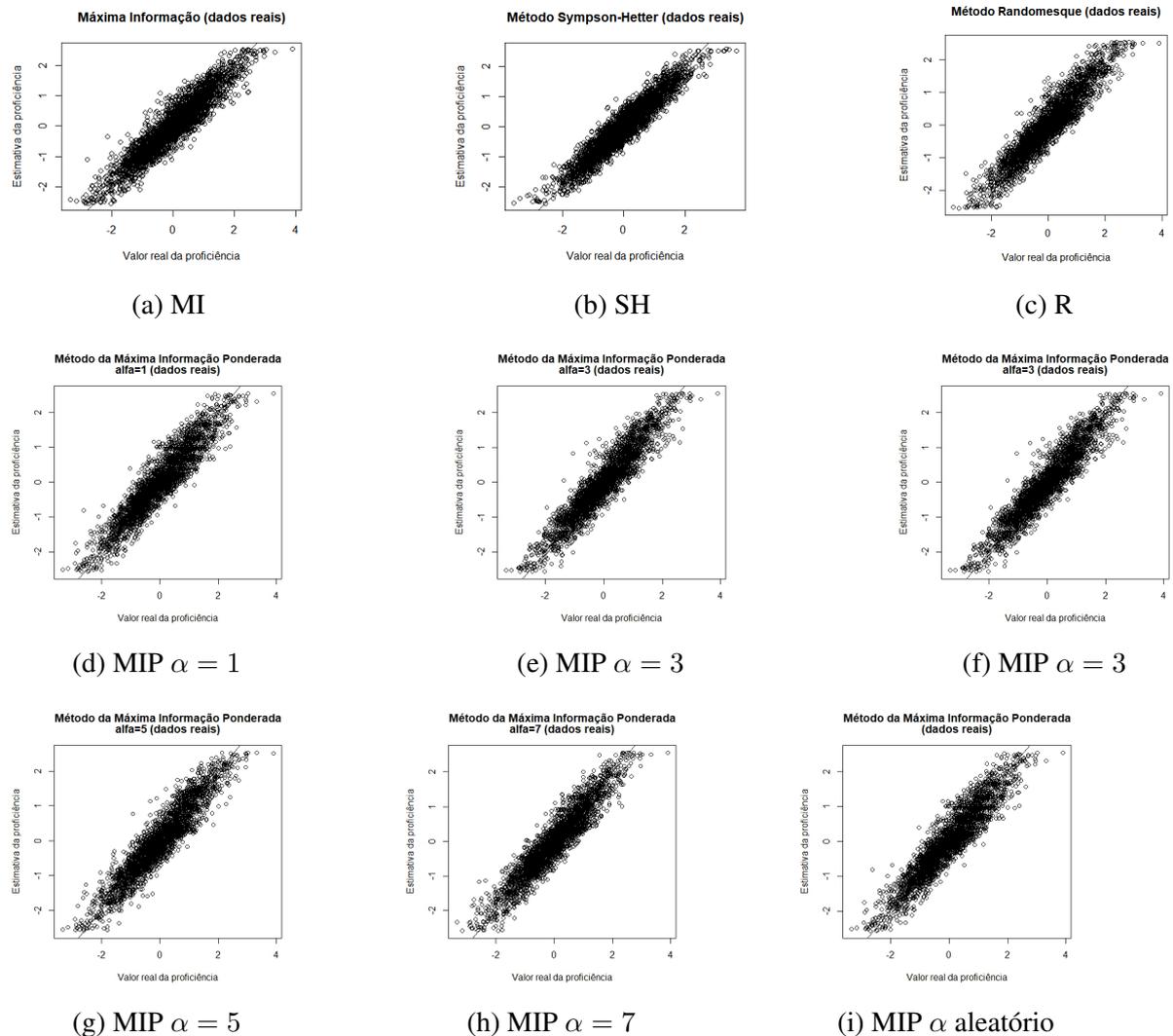
Pouco mais da metade do banco de itens foi utilizado, com número médio de 16 itens por respondente, desvio padrão em torno de 3 e valor médio do vício menor para proficiências acima de  $-1.5$ . Nenhum dos respondentes utilizou o número máximo de itens.

Tabela 2.14: Resultados obtidos pelo Método da Máxima Informação Ponderada com  $\alpha$  aleatório

	$\alpha$ aleatório
Taxa de sobreposição	0.1965%
Prop. de itens expostos	0.54028
Prop. de itens superexpostos	0.01895
Prop. de itens subutilizados	0.55924
Fidelidade	0.94656
Valor médio do vício	0.00685
Valor médio do vício para proficiências $\leq -1,5$	0.25314
Valor médio do vício para proficiências entre -1,5 e 1,5	0.00468
Valor médio do vício para proficiências $> 1,5$	-0.01027
EQM das estimativas das proficiências	0.10575
EQM das estimativas para as proficiências $\leq -1,5$	0.17644
EQM para as proficiências entre -1,5 e 1,5	0.09829
EQM para as proficiências $> 1,5$	0.10083
Prop. de respondentes expostos ao núm. máximo de itens	0.00000
Núm. médio de itens respondidos por respondente	16.49433
DP médio do núm. de itens respondidos por respondente	2.96674

A Figura 2.6 apresenta diagrama de dispersão entre o valor real da proficiência e a estimativa dos respondentes para todos os métodos abordados nesta pesquisa. Pelos gráficos é possível observar que os pontos das habilidades estimadas estão bem ajustadas a reta da proficiência real dos respondentes, principalmente no intervalo de  $[-1.5, 1.5]$ , indicando a qualidade das estimativas realizadas, com maior dispersão apenas nas suas extremidades e maior ênfase em indivíduos com habilidades inferiores. Se não houvesse vício todos os pontos estariam em cima da linha, ou seja, quanto mais perto o ponto estiver da reta, mais acurado é o método.

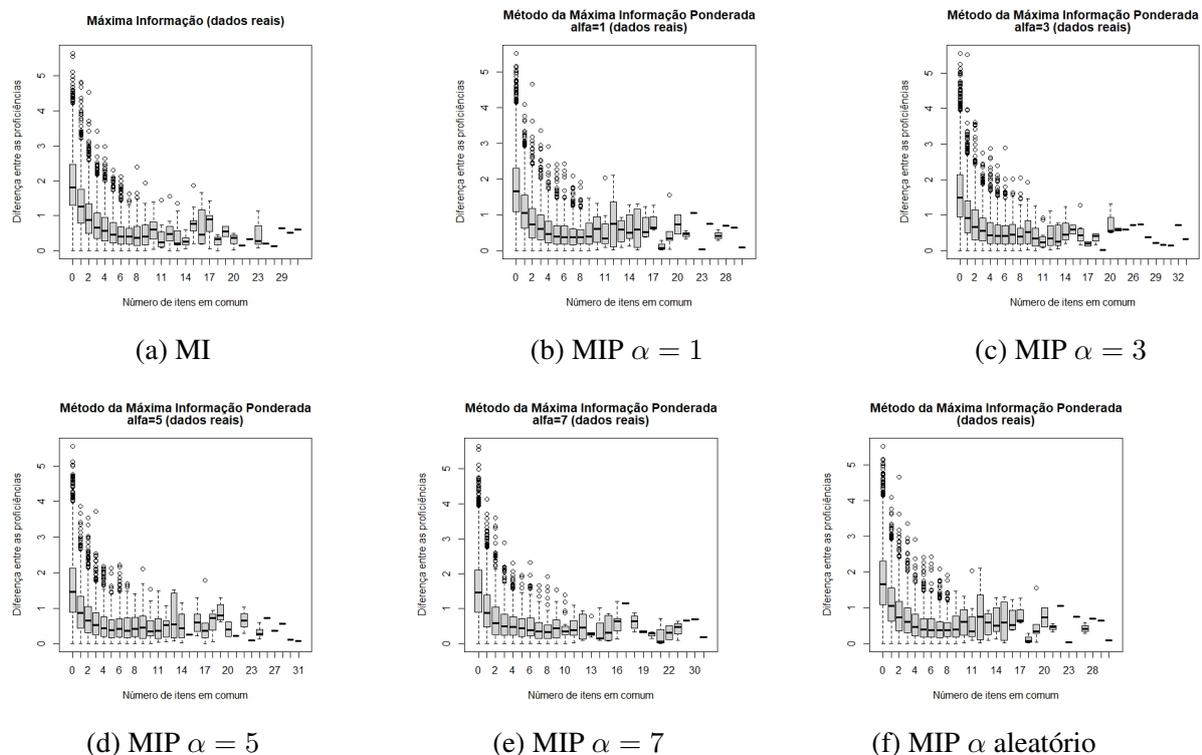
Figura 2.6: Proficiência real vs Proficiência estimada



Fonte: Elaborado pelo autor (2022).

Os gráficos presentes na Figura 2.7 apresentam a relação inversamente proporcional que ocorre entre o número de itens respondidos em comum por pares de respondentes e a diferença entre as suas proficiências. Quanto maior a diferença entre as proficiências menor é o número de itens coincidentes. Testes com mais de 15 itens em comum em sua grande maioria são aplicados a respondentes com baixa diferença entre as proficiências ( $\leq 2$ ). É possível observar uma alta variação para testes com mais itens em comum, este comportamento se dá devido a utilização de poucas observações para a construção do gráfico, apesar disso, o box plot apresentou resultados condizentes com a interpretação das tabelas 2.8 e 2.12 e correlacionou tal comportamento com a diferença da proficiência dos pares de respondentes.

Figura 2.7: Diferença entre as proficiências vs Número de itens coincidentes



Fonte: Elaborado pelo autor (2022).

Dentre os critérios propostos, o que demonstrou melhor desempenho foi o método da máxima informação ponderada com  $\alpha$  aleatório, quando avaliada a taxa de sobreposição, o número médio de itens respondidos por respondente e seu desvio padrão. Com relação a distribuição dos itens que compõe o método de informação ponderada aleatória consegue expor os itens em sua totalidade, sem gerar uma superexposição de nenhum dos itens, resultado este, obtido apenas por este método. A fidelidade se mantém constante para todas as abordagens propostas nesta pesquisa, podendo afirmar que os métodos conseguem realizar a estimação da habilidade do respondente de forma satisfatória, com valores altamente correlacionados com sua habilidade real.

Levando em consideração todos os métodos para controle da taxa de exposição dos itens abordados nesta pesquisa, o método da máxima informação ponderada com  $\alpha$  aleatório para cada item a ser escolhido apresenta vantagem quanto ao número de itens utilizado para estimação da habilidade do respondente, fazendo uso de uma menor quantidade de itens, com menor desvio padrão médio, possibilitando que nenhum examinado seja submetido ao número máximo de itens, promovendo um teste menos desgastante para o examinado e que pode ser realizado em um período de tempo mais curto. Contudo, sua proporção de itens expostos é inferior aos métodos SH, IPA e MIP com  $\alpha = 3, 5$  e  $7$ , ou seja, este método faz uso de um percentual menor de itens do banco.

A fidelidade dos métodos utilizados atualmente é a mesma alcançada pelos métodos propostos, não sendo uma medida viável para comparação. As dificuldades apresentadas na MI, R e SH com relação ao vício médio para proficiências  $\leq -1.5$ , não foi solucionada pelos métodos aqui propostos, portanto, os mesmos também apresentam maior viés neste intervalo da distribuição.

## Capítulo 3

### Conclusão e Propostas

Mediante as simulações realizadas nesta pesquisa, podemos concluir que a metodologia proposta e suas diversas abordagens tem vantagens em alguns aspectos sobre os métodos já utilizados, mas ainda não são eficazes o suficiente para substituir tais metodologias.

Em relação à aplicação no banco de itens simulados, a escolha do valor de  $\alpha$  para a máxima informação ponderada, é indicado  $\alpha = 5$  caso o principal objetivo da aplicação seja não apresentar itens superexpostos, obtendo também um número de itens subexpostos menor, fazendo uso de mais itens e com uma diferença de menos de dois itens a mais em média por respondente. Para o critério do número médio de itens respondidos, é indicado a atribuição de  $\alpha = 1$ , tendo visto que o número médio de itens aumenta de forma discreta conforme aumenta o valor de  $\alpha$ . O  $\alpha$  máximo considerado (8) aumenta a taxa de exposição para 60%, não tem itens superexpostos e não deixa itens com pouca exposição, entretanto tem 3 itens a mais em média por respondente.

Como solução para escolha do valor de  $\alpha$  pode-se utilizar um valor aleatório com base em uma distribuição uniforme para cada item a ser respondido, sem que haja perda das vantagens citadas anteriormente.

A grande vantagem da IPA é a utilização total do banco, pois ter itens que não são utilizados, significa que foi um gasto desnecessário em ter aquele item no banco. Em contra partida, necessita de uma quantidade maior de itens para convergir, sendo exigidos aproximadamente 15 itens a mais em média.

Avaliando conjuntamente a taxa de exposição e o número médio de itens utilizados pra convergência, a máxima informação ponderada é classificada como um dos melhores métodos, juntamente com Sympton-Hetter. Os mesmos produziram resultados muito similares em diversas condições

Sendo assim, o método da máxima informação ponderada (com  $\alpha = 1$ ) tem desempenho similar ao Randomesque, levemente superior ao Sympton-Hetter e superior ao de informação ponderada aleatória e de máxima informação ao utilizar a ponderação na escolha dos itens, por possuir menor porcentagem média de itens coincidentes entre respondentes, menor proporção de itens superexpostos, e proporção de itens expostos similar aos demais métodos.

Ao realizar a aplicação na prova de Proficiência de Inglês da UnB a máxima informação ponderada com  $\alpha$  aleatório se destacou de forma positiva das demais metodologias consideradas

nos critérios aqui avaliados, com exceção da proporção de itens expostos, fazendo uso de um percentual menor de itens do banco.

A possibilidade de seguimento na aplicação do teste caso haja necessidade de modificações urgentes no banco de itens, por diversos motivos, é mais uma das vantagens sob o método Sympton-Hetter, pelo fato do mesmo ter a necessidade de recalcular as taxas de exposição sempre que houver a exclusão ou inclusão de algum item no banco de itens. Em casos excepcionais como identificação de um erro ou violação de algum item, modificações no banco podem ser aplicadas e não trarão grandes impactos para os métodos aqui propostos.

Por fim, nenhuma das abordagens para a informação ponderada pode ser considerada como solução final para o problema da taxa de exposição dos itens. Como proposta para trabalho futuros sugere-se a ampliação das simulações e a utilização de outros critérios de seleção como os citados em 1.2.2.

# Referências

- [1] Andrade, D. F., Tavares, H. R., e Valle, R. C. (2000). *Teoria da resposta ao item: conceitos e aplicações*. ABE, Sao Paulo.
- [2] Assis, J. P. D., Sousa, R. D., Dias, C. T. D. S. (2019). *Glossário de estatística*. Mossoró: EdUFERSA.
- [3] Azevedo, C. L. N. (2003). *Métodos de estimação na teoria de resposta ao item*. Tese de Doutorado. Universidade de São Paulo.
- [4] Barrada, J. R., Olea, J., Ponsoda, V. (2007). *Methods for restricting maximum exposure rate in computerized adaptive testing*. *Methodology*, 3(1), 14-23.
- [5] Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*. *Statistical theories of mental test scores*, pages 395–479.
- [6] Chang, H. H., Ying, Z. (1996). *A global information approach to computerized adaptive testing*. *Applied Psychological Measurement*, 20(3), 213-229.
- [7] Chang, H. H., Ying, Z. (2009). *Nonlinear sequential designs for logistic item response theory models with applications to computerized adaptive tests*. *The Annals of Statistics*, 1466-1488.
- [8] Chang, S-W, Twu B-Y.(1998). *A comparative study of item exposure control methods in computerized adaptive testing*. The University of Iowa.
- [9] Chen, S. Y., Ankenmann, R. D., Chang, H. H. (2000). *A comparison of item selection rules at the early stages of computerized adaptive testing*. *Applied Psychological Measurement*, 24(3), 241-255.
- [10] Cheng, P. E., Liou, M. (2003). *Computerized adaptive testing using the nearest-neighbors criterion*. *Applied Psychological Measurement*, 27(3), 204-216.
- [11] Cheng, Y., Diao, Q., Behrens, J. T. (2017). *A simplified version of the maximum information per time unit method in computerized adaptive testing*. *Behavior research methods*, 49(2), 502-512.

- [12] Costa, D. R. (2009). *Métodos estatísticos em testes adaptativos informatizados*. Master's thesis, UFRJ-Universidade Federal do Rio de Janeiro, 9.
- [13] Fonseca, J. C. G. (2019). *Uma aplicação do teste adaptativo computadorizado via Filtro de Kalman não-linear*.
- [14] Galvao, A. F., Neto, R. F., Borges, C. C. H. (2013). *Um Modelo Inteligente para Seleção de Itens em Testes Adaptativos Computadorizados*.
- [15] Hambleton, R. K., Swaminathan, H., Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.
- [16] Jatobá, V. M. G. (2018). *Uma abordagem personalizada no processo de seleção de itens em Testes Adaptativos Computadorizados*. Tese de Doutorado. Universidade de São Paulo.
- [17] Kullback, S. (1959). *Information theory and statistics*. John Riley and Sons, Inc. New York.
- [18] Kingsbury, G. G., Zara, A. R. (1989). *Procedures for selecting items for computerized adaptive tests*. *Applied measurement in education*, 2(4), 359-375.
- [19] Labarrere, J. G., Da-Silva, C. Q., Costa, D. R. (2011). *Testes adaptativos computadorizados*. *Revista Brasileira de Biometria*, São Paulo, 29(2), 229-261.
- [20] Lee, H., Dodd, B. G. (2012). *Comparison of exposure controls, item pool characteristics, and population distributions for CAT using the partial credit model*. *Educational and Psychological Measurement*, 72(1), 159-175.
- [21] Leroux, A. J., Lopez, M., Hembry, I., Dodd, B. G. (2013). *A comparison of exposure control procedures in CATs using the 3PL model*. *Educational and Psychological Measurement*, 73(5), 857-874.
- [22] Leung, C. K., Chang, H. H., Hau, K. T. (2002). *Item selection in computerized adaptive testing: Improving the a-stratified design with the Sympton-Hetter algorithm*. *Applied Psychological Measurement*, 26(4), 376-392.
- [23] Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- [24] Magis, D., Barrada, J. R. (2017). *Computerized adaptive testing with R: Recent updates of the package catR*. *Journal of Statistical Software*, 76(1), 1-19.
- [25] Parshall, C., Harmes, J. C., Kromrey, J. D. (2000). *Item exposure control in computer-adaptive testing: The use of freezing to augment stratification*. *Florida Journal of Educational Research*, 40(1), 28-52.
- [26] Revuelta, J., Ponsoda, V. (1998). *A comparison of item exposure control methods in computerized adaptive testing*. *Journal of Educational Measurement*, 35(4), 311-327.

- [27] Salgado, A. M. (2018). *Uso de regressão isotônica na escolha de itens em testes adaptativos computadorizados*. Dissertação de Mestrado em Estatística, Universidade de Brasília.
- [28] Spenassato, D., Trierweiller, A. C., de Andrade, D. F., Bornia, A. C. (2016). *Testes adaptativos computadorizados aplicados em avaliações educacionais*. Revista Brasileira de Informática na Educação, 24(02), 1.
- [29] Spenassato, D. (2017). *Manutenção do banco de itens para testes adaptativos computadorizados aplicados em avaliações de alto impacto*.
- [30] Stocking, M. L. (1993). *Controlling item exposure rates in a realistic adaptive testing paradigm*. ETS Research Report Series, 1993(1), i-31.
- [31] Stocking, M. L., Lewis, C. (1998). *Controlling item exposure conditional on ability in computerized adaptive testing*. Journal of Educational and Behavioral Statistics, 23(1), 57-75.
- [32] Linden, W. J., Pashley, P. J. (2009). *Item selection and ability estimation in adaptive testing*. In Elements of adaptive testing (pp. 3-30). Springer, New York, NY.
- [33] Warm, T. A. (1989). *Weighted Likelihood Estimation of Ability in Item Response Theory*. Psychometrika, 54, 427-450.

# Apêndice A

## Código em R para implementação do método de Máxima Informação

```
1
2 #####
3 ### METODO DE MAXIMA INFORMACAO
4 ### ESCOLHA DE ITENS POR MAXIMA INFORMACAO
5 ### NUMERO ALEATORIO DE ITENS POR TESTE
6 ### NUMERO MAXIMO DE ITENS RESPONDIDOS = 40 (n.max)
7 ### NUMERO DE ITENS NO BANCO DE ITENS = 400 (n.itens)
8 ### NUMERO DE RESPONDENTES = 3000 (n.resp)
9 #####
10
11 #####
12 ### CAT
13 ### - Estimacao da proficiencia via EAP.
14 ### - Escolha do item com maior valor para a informacao ponderada em theta
    estimado.
15 ### - Criterio de parada: PSD.
16 ### - Integracao via retangulos
17 #####
18
19 #####
20 ### Funcao "veross"          ###
21 ### calcula o valor da verossimilhanca ###
22 #####
23 veross <- function(par) {
24   prod <- 1
25   for (k in 1:length(resp.aux)) {
26     p.aux <- c.aux[k] + (1-c.aux[k]) / (1+exp(-a.aux[k]*(par-b.aux[k])))
27     prod <- prod*(p.aux^resp.aux[k]) * ((1-p.aux)^(1-resp.aux[k]))
28   }
29   return(prod)
30 }
31
32 #####
33 ### Funcao "numerador"      ###
34 ### numerador a ser integrado em theta no calculo de EAP ###
35 #####
36 numerador <- function(par) {
```

```

37 par*veross(par)*dnorm(par,0,1)
38 }
39
40 #####
41 ### Funcao "numerador2" ###
42 ### numerador a ser integrado em theta no calculo de EAP ###
43 #####
44 numerador2 <- function(par) {
45   (par^2)*veross(par)*dnorm(par,0,1)
46 }
47
48 #####
49 ### Funcao "denominador" ###
50 ### denominador a ser integrado em theta no calculo de EAP ###
51 #####
52 denominador <- function(par) {
53   veross(par)*dnorm(par,0,1)
54 }
55
56 #####
57 ### Geracao dos parametros dos itens e do valor real de theta ###
58 #####
59
60 set.seed(23456)
61 n.itens <- 400 # numero de itens no banco de quest es
62 b.par <- rnorm(n.itens,0,1) # valores dos parametros de dificuldade
63 a.par <- runif(n.itens,0.3,2.5) # valores dos parametros de discrimina o
64 c.par <- rep(0.2,n.itens)# valores dos parametros de acerto ao acaso
65
66 n.resp <- 3000 # numero de respondentes
67
68 r.max <- 0.2 # taxa de exposi o mxima aceit vel
69
70 amp <- 0.001 # amplitude dos intervalos na integracao via retangulos
71 param <- seq(-3,3,amp) # valores de theta utilizados na integracao numerica
72
73 n.max <- 60 # numero maximo de itens respondidos
74
75 #####
76 ## Inicio das repeticoes
77 #####
78
79 n.rep <- 20 # numero de repeticoes na simulacao
80
81 repet.overlap <- numeric(0) # vetor com taxas de sobreposicao de cada
   simulacao
82 repet.it.exp <- numeric(0) # vetor com proporcao de itens expostos de cada
   simulacao
83 repet.it.sup <- numeric(0) # vetor com proporcao de itens superexpostos de
   cada simulacao

```

```
84 repet.it.subut <- numeric(0) # vetor com proporcao de itens subutilizados
    de cada simulacao
85 repet.fid <- numeric(0) # vetor com a fidelidade (correlacao entre theta e
    sua estimativa) de cada simulacao
86 repet.vicio <- numeric(0) # vetor com vicio medio das estimativas de theta
    de cada simulacao
87 repet.vicio.A <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ theta < -1.5 de cada simulacao
88 repet.vicio.B <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ -1.5 < theta < 1.5 de cada simulacao
89 repet.vicio.C <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ theta > 1.5 de cada simulacao
90 repet.eqm <- numeric(0) # vetor com erro quadratico medio de cada simulacao
91 repet.eqm.A <- numeric(0) # vetor com erro quadratico medio p/ theta < -1.5
    de cada simulacao
92 repet.eqm.B <- numeric(0) # vetor com erro quadratico medio p/ -1.5 < theta
    < 1.5 de cada simulacao
93 repet.eqm.C <- numeric(0) # vetor com erro quadratico medio p/ theta > 1.5
    de cada simulacao
94 repet.it.resp.lim <- numeric(0) # vetor com proporcao de respondentes no
    limite de itens de cada simulacao
95 repet.no.it.resp <- numeric(0) # vetor com numero medio de itens
    respondidos de cada simulacao
96 repet.dp.no.it.resp <- numeric(0) # vetor com desvio-padrao do numero de
    itens respondidos de cada simulacao
97
98
99 ## Inicio do loop das repeticoes das simulacoes
100 for (m in 1:n.rep) {
101
102   print(m)
103
104   vec.theta <- rnorm(n.resp,0,1) # proficiencia real dos respondentes
105   vec.theta.est <- numeric(0) # vetor com estimativas de theta
106
107   vec.n.it.resp <- numeric(0) # vetor com numero de itens respondidos
108
109   prop <- rep(0,n.itens) # proporcao de respondentes que responderam cada
    item
110
111   # matriz indicadora de aplicacao dos itens aos respondentes (linha=
    respondente, coluna=item)
112   mat.overlap <- matrix(0,n.resp,n.itens)
113
114   for (j in 1:n.resp) {
115
116     # print(j)
117
118     theta.est <- 0 # estimativa de theta
119
```

```

120 psd.acum <- numeric(0) # vetor com sequencia de estimativas do desvio-
    padrao a posteriori de theta.est
121
122 tol <- 0.1 # limite de tolerancia para convergencia da estimativa de
    theta.est[j] para psd < tol
123 psd <- 1 # valor inicial do desvio padrao a posteriori de theta.est
124
125 #####
126 # banco de itens: colunas da matriz "banco"
127 # 1-indice, 2-a.par, 3-b.par, 4-c.par,
128 # 5-resposta (0=erro, 1=acerto, 9=ainda nao respondido),
129 # 6-ordem de selecao.
130 #####
131
132 banco <- cbind(c(1:n.itens), a.par, b.par, c.par, rep(9, n.itens), rep(0, n.
    itens))
133
134 n.it.resp <- 0 # numero cumulativo de itens respondidos
135
136 ## Escolha de 6 itens iniciais com "a<1", sendo 3 itens com "b" em
    [-1,0], e 3 itens com "b" em [0,1]
137
138 itens.neg <- banco[banco[,2]<1,]
139 itens.neg <- itens.neg[itens.neg[,3]<0,]
140 itens.neg <- itens.neg[itens.neg[,3]>-1,]
141
142 itens.pos <- banco[banco[,2]<1,]
143 itens.pos <- itens.pos[itens.pos[,3]>=0,]
144 itens.pos <- itens.pos[itens.pos[,3]<1,]
145
146 n.neg <- 3
147 n.pos <- 3
148
149 ind.neg <- sample(itens.neg[,1], n.neg)
150 ind.pos <- sample(itens.pos[,1], n.pos)
151
152 ind.aux <- sort(c(ind.neg, ind.pos))
153
154 aa <- bb <- cc <- numeric(0)
155
156 for (i in 1:length(ind.aux)) {
157   aa <- c(aa, banco[banco[,1]==ind.aux[i], 2])
158   bb <- c(bb, banco[banco[,1]==ind.aux[i], 3])
159   cc <- c(cc, banco[banco[,1]==ind.aux[i], 4])
160 }
161
162 ## geracao das respostas para os 6 itens iniciais
163
164 p.acerto <- cc + (1-cc)/(1+exp(-aa*(vec.theta[j]-bb)))
165 u <- sign(sign(p.acerto-runif(length(ind.aux), 0, 1))+1) # resposta
166

```

```

167 n.it.resp <- n.neg + n.pos
168
169 ## Estimacao inicial de "theta"
170 ## calculo de theta.est via EAP com priori N(0,1)
171
172 banco[ind.aux,5] <- u
173 banco[ind.aux,6] <- c(1:6)
174
175 # os vetores a.aux, b.aux, c.aux e resp.aux contem os valores dos
176 # dos itens e as respostas para os itens ja' respondidos
177
178 a.aux <- aa
179 b.aux <- bb
180 c.aux <- cc
181
182 resp.aux <- u
183
184 theta.est <- integrate( Numerador, -3, 3) $value / integrate( Denominador, -3, 3) $
185 value
186 theta.est2 <- integrate( Numerador2, -3, 3) $value / integrate( Denominador
187 , -3, 3) $value
188
189 psd <- theta.est2 - theta.est^2
190 psd.acum <- c(psd.acum, psd)
191
192 ## Escolha a partir do setimo item segundo a maxima informacao
193
194 while ((n.it.resp <= n.max) && (psd > tol)) {
195
196   banco.aux <- banco[banco[,5]==9,]
197
198   ind.n.esc <- banco.aux[,1]
199   a.n.esc <- banco.aux[,2]
200   b.n.esc <- banco.aux[,3]
201   c.n.esc <- banco.aux[,4]
202
203   vec.prob <- c.n.esc + (1-c.n.esc) / (1+exp(-a.n.esc*(theta.est-b.n.esc)))
204
205   vec.info <- (a.n.esc^2) * ((1-vec.prob) / vec.prob) * (((vec.prob-c.n.esc) /
206   (1-c.n.esc))^2)
207
208   ind <- ind.n.esc[vec.info==max(vec.info)]
209
210   prop[ind] <- prop[ind] + 1
211
212   n.it.resp <- n.it.resp + 1 # numero de itens respondidos
213
214   a.esc <- banco[ind,2]
215   b.esc <- banco[ind,3]
216   c.esc <- banco[ind,4]

```

```

214
215 p.acerto <- c.esc+(1-c.esc)/(1+exp(-a.esc*(vec.theta[j]-b.esc)))
216 u <- sign(sign(p.acerto-runif(1,0,1))+1) # resposta
217
218 banco[ind,5] <- u
219 banco[ind,6] <- n.it.resp
220
221 #####
222 ## calculo de theta.est via EAP com priori N(0,1)
223 #####
224
225 # os vetores a.aux, b.aux, c.aux e resp.aux contem os valores dos
parametros
226 # dos itens e as respostas para os itens ja' respondidos
227
228 a.aux <- banco[banco[,5]!=9,2]
229 b.aux <- banco[banco[,5]!=9,3]
230 c.aux <- banco[banco[,5]!=9,4]
231
232 resp.aux <- banco[banco[,5]!=9,5]
233
234 theta.est <- integrate(numerador,-3,3)$value/integrate(denominador
,-3,3)$value
235 theta.est2 <- integrate(numerador2,-3,3)$value/integrate(denominador
,-3,3)$value
236
237 psd <- theta.est2-theta.est^2
238 psd.acum <- c(psd.acum,psd)
239
240 }
241
242 vec.theta.est <- c(vec.theta.est,theta.est)
243 vec.n.it.resp <- c(vec.n.it.resp,n.it.resp)
244
245 vec.aux <- banco[,5]
246 vec.aux[vec.aux>1] <- 0
247 mat.overlap[j,] <- vec.aux
248
249 }
250
251 overlap.theta.dif <- overlap.vec <- numeric(0)
252
253 overlap.soma <- 0
254 tam.am <- 10000 # tamanho da amostra de pares de respondentes para
estimacao da taxa de overlap
255
256 for (k in 1:tam.am) {
257   am.resp <- sample(c(1:n.resp),2,replace=FALSE)
258   am.overlap <- sum(mat.overlap[am.resp[1],]*mat.overlap[am.resp[2],])
259   overlap.soma <- overlap.soma + am.overlap
260   overlap.vec <- c(overlap.vec,am.overlap)

```

```

261   overlap.theta.dif <- c(overlap.theta.dif, abs(vec.theta[am.resp[1]]-vec.
262     theta[am.resp[2]]))
263 }
264 ## proporcao media de coincidencia de itens entre respondentes
265 overlap.rate <- 100*overlap.soma/(n.itens*tam.am)
266 #overlap.rate
267
268 ## Taxa de exposi o dos itens
269
270 tx.expos <- prop/n.resp
271 tx.expos
272 prop.it.exp <- length(tx.expos[tx.expos>0])/n.itens
273 prop.it.sup.exp <- length(tx.expos[tx.expos>r.max])/n.itens
274 prop.it.subut <- length(tx.expos[tx.expos<0.02])/n.itens
275
276 #mean(tx.expos) # taxa media de exposicao por item
277 #sd(tx.expos) # desvio-padrao da taxa de exposicao dos itens
278
279 #plot(table(prop))
280
281 ## Fidelidade
282 fid <-cor(vec.theta,vec.theta.est)
283 #fid
284
285 ## V cio
286 vicio <- sum(vec.theta.est-vec.theta)/n.resp
287 #vicio
288
289 ## V cio separado por tr s grupos
290 vicio.A <- sum(vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
291   /length(vec.theta[vec.theta < -1.5])
292 vicio.B <- sum(vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta
293   ) <= 1.5])/
294   length(vec.theta[abs(vec.theta) <= 1.5])
295 vicio.C <- sum(vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
296   /length(vec.theta[vec.theta > -1.5])
297
298 #vicio.A
299 #vicio.B
300 #vicio.C
301
302 ## Erro quadr tico m dio (MSE)
303 eqm <-sum((vec.theta.est-vec.theta)^2)/n.resp
304 #eqm
305
306 ## Erro quadr tico m dio (MSE) separado por tr s grupos
307 eqm.A <- sum((vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
308   ^2)/length(vec.theta[vec.theta < -1.5])
309 eqm.B <- sum((vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta)
310   <= 1.5])^2)/

```

```

306   length(vec.theta[abs(vec.theta) <= 1.5])
307 eqm.C <- sum((vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
308           ^2)/length(vec.theta[vec.theta > -1.5])
309 #eqm.A
310 #eqm.B
311 #eqm.C
312
313 ## Proporcao de respondentes que atingiu o numero maximo de itens permitido
314 prop.it.resp.lim <- length(vec.n.it.resp[vec.n.it.resp==(n.max+1)])/n.resp
315 #prop.it.resp.lim
316
317 ## Numero medio de itens respondidos por respondente
318 no.med.it.resp <- mean(vec.n.it.resp)
319 #no.med.it.resp
320
321 ## Desvio-padrao do numero de itens respondidos
322 dp.no.it.resp <- sd(vec.n.it.resp)
323 #dp.no.it.resp
324
325 repet.overlap <- c(repet.overlap, overlap.rate)
326 repet.it.exp <- c(repet.it.exp, prop.it.exp)
327 repet.it.sup <- c(repet.it.sup, prop.it.sup.exp)
328 repet.it.subut <- c(repet.it.subut, prop.it.subut)
329 repet.fid <- c(repet.fid, fid)
330 repet.vicio <- c(repet.vicio, vicio)
331 repet.vicio.A <- c(repet.vicio.A, vicio.A)
332 repet.vicio.B <- c(repet.vicio.B, vicio.B)
333 repet.vicio.C <- c(repet.vicio.C, vicio.C)
334 repet.eqm <- c(repet.eqm, eqm)
335 repet.eqm.A <- c(repet.eqm.A, eqm.A)
336 repet.eqm.B <- c(repet.eqm.B, eqm.B)
337 repet.eqm.C <- c(repet.eqm.C, eqm.C)
338 repet.it.resp.lim <- c(repet.it.resp.lim, prop.it.resp.lim)
339 repet.no.it.resp <- c(repet.no.it.resp, no.med.it.resp)
340 repet.dp.no.it.resp <- c(repet.dp.no.it.resp, dp.no.it.resp)
341
342 } # fim do loop das repeticoes das simulacoes
343
344 resultados <- rbind(repet.overlap, repet.it.exp, repet.it.sup, repet.it.
345   subut, repet.fid,
346   repet.vicio, repet.vicio.A, repet.vicio.B, repet.vicio.C, repet.eqm,
347   repet.eqm.A,
348   repet.eqm.B, repet.eqm.C, repet.it.resp.lim, repet.no.it.resp, repet.dp.
349   no.it.resp)
350
351 vec.media <- round(apply(resultados, 1, mean), digits=5)
352 vec.erro.padrao <- round(apply(resultados, 1, sd), digits=5)
353 sumario.media.ep <- cbind(vec.media, vec.erro.padrao)
354 sumario.media.ep <- matrix(sumario.media.ep, ncol=2)

```

## Código em R para implementação do método Randomesque

```

1 #####
2 ### METODO RANDOMESQUE (REPETICOES)
3 ### ESCOLHA DE ITENS POR MAXIMA INFORMACAO
4 ### NUMERO ALEATORIO DE ITENS POR TESTE
5 ### NUMERO MAXIMO DE ITENS RESPONDIDOS = 60 (n.max)
6 ### NUMERO DE ITENS NO BANCO DE ITENS = 400 (n.itens)
7 ### N MERO DE RESPONDENTES = 3000 (n.resp)
8 #####
9
10 #####
11 ### CAT
12 ### - Estimacao da proficiencia via EAP.
13 ### - Escolha do item com maior valor para a informacao ponderada em theta
    estimado.
14 ### - Criterio de parada: PSD.
15 ### - Integracao via retangulos
16 #####
17 #####
18 ### Funcao "veross" ###
19 ### calcula o valor da verossimilhanca ###
20 #####
21 veross <- function(par) {
22   prod <- 1
23   for (k in 1:length(resp.aux)) {
24     p.aux <- c.aux[k] + (1-c.aux[k]) / (1+exp(-a.aux[k]*(par-b.aux[k])))
25     prod <- prod*(p.aux^resp.aux[k]) * ((1-p.aux)^(1-resp.aux[k]))
26   }
27   return(prod)
28 }
29
30 #####
31 ### Funcao "numerador" ###
32 ### numerador a ser integrado em theta no calculo de EAP ###
33 #####
34 numerador <- function(par) {
35   par*veross(par)*dnorm(par,0,1)
36 }
37
38 #####
39 ### Funcao "numerador2" ###
40 ### numerador a ser integrado em theta no calculo de EAP ###
41 #####
42 numerador2 <- function(par) {
43   (par^2)*veross(par)*dnorm(par,0,1)
44 }
45
46 #####
47 ### Funcao "denominador" ###
48 ### denominador a ser integrado em theta no calculo de EAP ###
49 #####

```

```

50 denominador <- function(par) {
51   veross(par)*dnorm(par,0,1)
52 }
53
54 #####
55 ### Geracao dos parametros dos itens e do valor real de theta ###
56 #####
57
58 set.seed(23456)
59 n.itens <- 400 # numero de itens no banco de quest es
60 b.par <- rnorm(n.itens,0,1) # valores dos parametros de dificuldade
61 a.par <- runif(n.itens,0.3,2.5) # valores dos parametros de discrimina o
62 c.par <- rep(0.2,n.itens)# valores dos parametros de acerto ao acaso
63
64 n.resp <- 3000 # numero de respondentes
65
66 r.max <- 0.2 # taxa de exposi o mxima aceit vel
67
68 amp <- 0.001 # amplitude dos intervalos na integracao via retangulos
69 param <- seq(-3,3,amp) # valores de theta utilizados na integracao numerica
70
71 n.max <- 60 # numero maximo de itens respondidos
72
73 #####
74 ## Inicio das repeticoes
75 #####
76
77 n.rep <- 20 # numero de repeticoes na simulacao
78
79 repet.overlap <- numeric(0) # vetor com taxas de sobreposicao de cada
   simulacao
80 repet.it.exp <- numeric(0) # vetor com proporcao de itens expostos de cada
   simulacao
81 repet.it.sup <- numeric(0) # vetor com proporcao de itens superexpostos de
   cada simulacao
82 repet.it.subut <- numeric(0) # vetor com proporcao de itens subutilizados
   de cada simulacao
83 repet.fid <- numeric(0) # vetor com a fidelidade (correlacao entre theta e
   sua estimativa) de cada simulacao
84 repet.vicio <- numeric(0) # vetor com vicio medio das estimativas de theta
   de cada simulacao
85 repet.vicio.A <- numeric(0) # vetor com vicio medio das estimativas de
   theta p/ theta < -1.5 de cada simulacao
86 repet.vicio.B <- numeric(0) # vetor com vicio medio das estimativas de
   theta p/ -1.5 < theta < 1.5 de cada simulacao
87 repet.vicio.C <- numeric(0) # vetor com vicio medio das estimativas de
   theta p/ theta > 1.5 de cada simulacao
88 repet.eqm <- numeric(0) # vetor com erro quadratico medio de cada simulacao
89 repet.eqm.A <- numeric(0) # vetor com erro quadratico medio p/ theta < -1.5
   de cada simulacao
90 repet.eqm.B <- numeric(0) # vetor com erro quadratico medio p/ -1.5 < theta

```

```

    < 1.5 de cada simulacao
91 repet.eqm.C <- numeric(0) # vetor com erro quadratico medio p/ theta > 1.5
    de cada simulacao
92 repet.it.resp.lim <- numeric(0) # vetor com proporcao de respondentes no
    limite de itens de cada simulacao
93 repet.no.it.resp <- numeric(0) # vetor com numero medio de itens
    respondidos de cada simulacao
94 repet.dp.no.it.resp <- numeric(0) # vetor com desvio-padrao do numero de
    itens respondidos de cada simulacao
95
96
97 ## Inicio do loop das repeticoes das simulacoes
98 for (m in 1:n.rep) {
99
100 print(m)
101
102 vec.theta <- rnorm(n.resp,0,1) # proficiencia real dos respondentes
103 vec.theta.est <- numeric(0) # vetor com estimativas de theta
104
105 vec.n.it.resp <- numeric(0) # vetor com numero de itens respondidos
106
107 prop <- rep(0,n.itens) # proporcao de respondentes que responderam cada
    item
108
109
110 # matriz indicadora de aplicacao dos itens aos respondentes (linha=
    respondente, coluna=item)
111 mat.overlap <- matrix(0,n.resp,n.itens)
112
113 for (j in 1:n.resp) {
114
115 # print(j)
116
117 theta.est <- 0 # estimativa de theta
118
119 psd.acum <- numeric(0) # vetor com sequencia de estimativas do desvio-
    padrao a posteriori de theta.est
120
121
122 tol <- 0.1
123 psd <- 1
124
125 #####
126 # banco de itens: colunas da matriz "banco"
127 # 1-indice, 2-a.par, 3-b.par, 4-c.par,
128 # 5-resposta (0=erro, 1=acerto, 9=ainda nao respondido),
129 # 6-ordem de selecao.
130 #####
131
132 banco <- cbind(c(1:n.itens),a.par,b.par,c.par,rep(9,n.itens),rep(0,n.
    itens))

```

```

133
134 n.it.resp <- 0 # numero cumulativo de itens respondidos
135
136 while ((n.it.resp <= n.max)&&(psd>tol)) {
137
138     banco.aux <- banco[banco[,5]==9,]
139
140     ind.n.esc <- banco.aux[,1]
141     a.n.esc <- banco.aux[,2]
142     b.n.esc <- banco.aux[,3]
143     c.n.esc <- banco.aux[,4]
144
145     vec.prob <- c.n.esc + (1-c.n.esc)/(1+exp(-a.n.esc*(theta.est-b.n.esc)))
146
147     vec.info <- (a.n.esc^2)*((1-vec.prob)/vec.prob)*((vec.prob-c.n.esc)/
148     (1-c.n.esc))^2)
149
150     banco.aux[banco.aux[,1]==ind.n.esc,6] <- vec.info
151
152     top5.info <- sort(vec.info,decreasing=TRUE)[1:5]
153     esc.info.top5 <- sample(top5.info, size=1, replace=FALSE)
154     ind <- banco.aux[banco.aux[,6]==esc.info.top5,1]
155
156     prop[ind] <- prop[ind] + 1
157
158     n.it.resp <- n.it.resp + 1 # numero de itens respondidos
159
160     a.esc <- banco[ind,2]
161     b.esc <- banco[ind,3]
162     c.esc <- banco[ind,4]
163
164     p.acerto <- c.esc+(1-c.esc)/(1+exp(-a.esc*(vec.theta[j]-b.esc)))
165     u <- sign(sign(p.acerto-runif(1,0,1))+1) # resposta
166
167     banco[ind,5] <- u
168     banco[ind,6] <- n.it.resp
169
170     #####
171     ## calculo de theta.est via EAP com priori N(0,1)
172     #####
173
174     # os vetores a.aux, b.aux, c.aux e resp.aux contem os valores dos
175     # dos itens e as respostas para os itens ja' respondidos
176
177     a.aux <- banco[banco[,5]!=9,2]
178     b.aux <- banco[banco[,5]!=9,3]
179     c.aux <- banco[banco[,5]!=9,4]
180
181     resp.aux <- banco[banco[,5]!=9,5]

```

```

182   theta.est <- integrate(numerador,-3,3)$value/integrate(denominador
183   ,-3,3)$value
184   theta.est2 <- integrate(numerador2,-3,3)$value/integrate(denominador
185   ,-3,3)$value
186   psd <- theta.est2-theta.est^2
187   psd.acum <- c(psd.acum,psd)
188 }
189
190 vec.theta.est <- c(vec.theta.est,theta.est)
191 vec.n.it.resp <- c(vec.n.it.resp,n.it.resp)
192
193 vec.aux <- banco[,5]
194 vec.aux[vec.aux>1] <- 0
195 mat.overlap[j,] <- vec.aux
196
197 }
198
199 overlap.theta.dif <- overlap.vec <- numeric(0)
200
201 overlap.soma <- 0
202 tam.am <- 10000 # tamanho da amostra de pares de respondentes para
203   estimacao da taxa de overlap
204
205 for (k in 1:tam.am) {
206   am.resp <- sample(c(1:n.resp),2,replace=FALSE)
207   am.overlap <- sum(mat.overlap[am.resp[1],]*mat.overlap[am.resp[2],])
208   overlap.soma <- overlap.soma + am.overlap
209   overlap.vec <- c(overlap.vec,am.overlap)
210   overlap.theta.dif <- c(overlap.theta.dif,abs(vec.theta[am.resp[1]]-vec.
211     theta[am.resp[2]]))
212 }
213
214 ## proporcao media de coincidencia de itens entre respondentes
215 overlap.rate <- 100*overlap.soma/(n.itens*tam.am)
216 #overlap.rate
217
218 ## Taxa de exposi o dos itens
219
220 tx.expos <- prop/n.resp
221 tx.expos
222 prop.it.exp <- length(tx.expos[tx.expos>0])/n.itens
223 prop.it.sup.exp <- length(tx.expos[tx.expos>r.max])/n.itens
224 prop.it.subut <- length(tx.expos[tx.expos<0.02])/n.itens
225
226 #mean(tx.expos) # taxa media de exposicao por item
227 #sd(tx.expos) # desvio-padrao da taxa de exposicao dos itens
228 #plot(table(prop))

```

```

229 fid <-cor(vec.theta,vec.theta.est)
230 #fid
231
232 ## V c i o
233 vicio <- sum(vec.theta.est-vec.theta)/n.resp
234 #vicio
235
236 ## V c i o separado por tr s grupos
237 vicio.A <- sum(vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
      /length(vec.theta[vec.theta < -1.5])
238 vicio.B <- sum(vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta
      ) <= 1.5])/
239     length(vec.theta[abs(vec.theta) <= 1.5])
240 vicio.C <- sum(vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
      /length(vec.theta[vec.theta > -1.5])
241
242 #vicio.A
243 #vicio.B
244 #vicio.C
245
246 ## Erro quadr tico m dio (MSE)
247 eqm <-sum((vec.theta.est-vec.theta)^2)/n.resp
248 #eqm
249
250 ## Erro quadr tico m dio (MSE) separado por tr s grupos
251 eqm.A <- sum((vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
      ^2)/length(vec.theta[vec.theta < -1.5])
252 eqm.B <- sum((vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta)
      <= 1.5])^2)/
253     length(vec.theta[abs(vec.theta) <= 1.5])
254 eqm.C <- sum((vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
      ^2)/length(vec.theta[vec.theta > -1.5])
255
256 #eqm.A
257 #eqm.B
258 #eqm.C
259
260 ## Proporcão de respondentes que atingiu o número máximo de itens permitido
261 prop.it.resp.lim <- length(vec.n.it.resp[vec.n.it.resp==(n.max+1)])/n.resp
262 #prop.it.resp.lim
263
264 ## Número médio de itens respondidos por respondente
265 no.med.it.resp <- mean(vec.n.it.resp)
266 #no.med.it.resp
267
268 ## Desvio-padrão do número de itens respondidos
269 dp.no.it.resp <- sd(vec.n.it.resp)
270 #dp.no.it.resp
271
272 repet.overlap <- c(repet.overlap,overlap.rate)
273 repet.it.exp <- c(repet.it.exp,prop.it.exp)

```

```

274 repet.it.sup <- c(repet.it.sup,prop.it.sup.exp)
275 repet.it.subut <- c(repet.it.subut,prop.it.subut)
276 repet.fid <- c(repet.fid,fid)
277 repet.vicio <- c(repet.vicio,vicio)
278 repet.vicio.A <- c(repet.vicio.A,vicio.A)
279 repet.vicio.B <- c(repet.vicio.B,vicio.B)
280 repet.vicio.C <- c(repet.vicio.C,vicio.C)
281 repet.eqm <- c(repet.eqm,eqm)
282 repet.eqm.A <- c(repet.eqm.A,eqm.A)
283 repet.eqm.B <- c(repet.eqm.B,eqm.B)
284 repet.eqm.C <- c(repet.eqm.C,eqm.C)
285 repet.it.resp.lim <- c(repet.it.resp.lim,prop.it.resp.lim)
286 repet.no.it.resp <- c(repet.no.it.resp,no.med.it.resp)
287 repet.dp.no.it.resp <- c(repet.dp.no.it.resp,dp.no.it.resp)
288
289 } # fim do loop das repeticoes das simulacoes
290
291 resultados <- rbind(repet.overlap, repet.it.exp, repet.it.sup, repet.it.
  subut, repet.fid,
292   repet.vicio, repet.vicio.A, repet.vicio.B, repet.vicio.C, repet.eqm,
  repet.eqm.A,
293   repet.eqm.B, repet.eqm.C, repet.it.resp.lim, repet.no.it.resp, repet.dp.
  no.it.resp)
294
295 vec.media <- round(apply(resultados,1,mean),digits=5)
296 vec.erro.padrao <- round(apply(resultados,1,sd),digits=5)
297 sumario.media.ep <- cbind(vec.media,vec.erro.padrao)
298 sumario.media.ep <- matrix(sumario.media.ep,ncol=2)

```

### Código em R para implementação do método Simpson e Hetter

```

1 #####
2 ### METODO DE SYMPSON E HETTER (REPETICOES)
3 ### ESCOLHA DE ITENS POR MAXIMA INFORMACAO
4 ### NUMERO ALEATORIO DE ITENS POR TESTE
5 ### NUMERO MAXIMO DE ITENS RESPONDIDOS = 60 (n.max)
6 ### NUMERO DE ITENS NO BANCO DE ITENS = 400 (n.itens)
7 ### N MERO DE RESPONDENTES = 3000 (n.resp)
8 #####
9
10 #####
11 ### CAT
12 ### - Estimacao da proficiencia via EAP.
13 ### - Escolha do item com maior valor para a informacao ponderada em theta
  estimado.
14 ### - Criterio de parada: PSD.
15 ### - Integracao via retangulos
16 #####
17
18 #####
19 ### Funcao "veross"      ###
20 ### calcula o valor da verossimilhanca ###

```

```

21 #####
22 veross <- function(par) {
23   prod <- 1
24   for (k in 1:length(resp.aux)) {
25     p.aux <- c.aux[k] + (1-c.aux[k]) / (1+exp(-a.aux[k]*(par-b.aux[k])))
26     prod <- prod*(p.aux^resp.aux[k])*(1-p.aux)^(1-resp.aux[k])
27   }
28   return(prod)
29 }
30
31 #####
32 ### Funcao "numerador" ###
33 ### numerador a ser integrado em theta no calculo de EAP ###
34 #####
35 numerador <- function(par) {
36   par*veross(par)*dnorm(par,0,1)
37 }
38
39 #####
40 ### Funcao "numerador2" ###
41 ### numerador a ser integrado em theta no calculo de EAP ###
42 #####
43 numerador2 <- function(par) {
44   (par^2)*veross(par)*dnorm(par,0,1)
45 }
46
47 #####
48 ### Funcao "denominador" ###
49 ### denominador a ser integrado em theta no calculo de EAP ###
50 #####
51 denominador <- function(par) {
52   veross(par)*dnorm(par,0,1)
53 }
54
55 #####
56 ### Geracao dos parametros dos itens e do valor real de theta ###
57 #####
58
59 set.seed(23456)
60 n.itens <- 400 # numero de itens no banco de quest es
61 b.par <- rnorm(n.itens,0,1) # valores dos parametros de dificuldade
62 a.par <- runif(n.itens,0.3,2.5) # valores dos parametros de discrimina o
63 c.par <- rep(0.2,n.itens)# valores dos parametros de acerto ao acaso
64
65 #####
66 ### Aplicacao do CAT apos determinacao dos parametros de exposicao dos
   itens
67 #####
68 n.resp <- 3000 # numero de respondentes
69
70 amp <- 0.001 # amplitude dos intervalos na integracao via retangulos

```

```
71 param <- seq(-3,3,amp) # valores de theta utilizados na integracao numerica
72
73 n.max <- 60 # numero maximo de itens respondidos
74
75 #####
76 ### Metodo de Sympson e Hetter para determinacao do valor do
77 ### parametro de exposicao de cada item
78 #####
79
80 r.max <- 0.2 # taxa de exposicao maxima
81
82 n.resp.SH <- 3000 # numero de respondentes ficticios na etapa de ajuste
83
84 #####
85 ## Inicio das repeticoes
86 #####
87
88 n.rep <- 20 # numero de repeticoes na simulacao
89
90 repet.overlap <- numeric(0) # vetor com taxas de sobreposicao de cada
    simulacao
91 repet.it.exp <- numeric(0) # vetor com proporcao de itens expostos de cada
    simulacao
92 repet.it.sup <- numeric(0) # vetor com proporcao de itens superexpostos de
    cada simulacao
93 repet.it.subut <- numeric(0) # vetor com proporcao de itens subutilizados
    de cada simulacao
94 repet.fid <- numeric(0) # vetor com a fidelidade (correlacao entre theta e
    sua estimativa) de cada simulacao
95 repet.vicio <- numeric(0) # vetor com vicio medio das estimativas de theta
    de cada simulacao
96 repet.vicio.A <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ theta < -1.5 de cada simulacao
97 repet.vicio.B <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ -1.5 < theta < 1.5 de cada simulacao
98 repet.vicio.C <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ theta > 1.5 de cada simulacao
99 repet.eqm <- numeric(0) # vetor com erro quadratico medio de cada simulacao
100 repet.eqm.A <- numeric(0) # vetor com erro quadratico medio p/ theta < -1.5
    de cada simulacao
101 repet.eqm.B <- numeric(0) # vetor com erro quadratico medio p/ -1.5 < theta
    < 1.5 de cada simulacao
102 repet.eqm.C <- numeric(0) # vetor com erro quadratico medio p/ theta > 1.5
    de cada simulacao
103 repet.it.resp.lim <- numeric(0) # vetor com proporcao de respondentes no
    limite de itens de cada simulacao
104 repet.no.it.resp <- numeric(0) # vetor com numero medio de itens
    respondidos de cada simulacao
105 repet.dp.no.it.resp <- numeric(0) # vetor com desvio-padrao do numero de
    itens respondidos de cada simulacao
106
```

```

107
108 ## Inicio do loop das repeticoes das simulacoes
109 for (m in 1:n.rep) {
110
111 print(m)
112
113 vec.theta.est <- numeric(0) # vetor com estimativas de theta
114
115 theta.SH <- rnorm(n.resp.SH,0,1) # proficiencias dos respondentes ficticios
    na etapa de ajuste
116
117 PAS <- rep(1,n.itens) # probabilidade de aplicacao de cada item dado que
    foi escolhido
118
119 PA <- rep(1,n.itens) # probabilidade de cada item ser aplicado
120
121 while (max(PA) > (r.max+0.01)) {
122
123   nA <- rep(0,n.itens) # numero de respondentes que responderam cada item
124   nS <- rep(0,n.itens) # numero de respondentes para os quais cada item foi
    selecionado
125
126   for (j in 1:n.resp.SH) {
127
128     # Aplicando o CAT para o j-esimo respondente
129 #print(j)
130
131     theta <- theta.SH[j] # valor real da proficiencia do respondente
132     theta.est <- 0 # estimativa de theta
133
134     psd.acum <- numeric(0) # vetor com sequencia de estimativas do desvio-
    padrao a posteriori de theta.est
135
136     tol <- 0.1
137     psd <- 1
138
139     #####
140     # banco de itens: colunas da matriz "banco"
141     # 1-indice, 2-a.par, 3-b.par, 4-c.par,
142     # 5-resposta (0=erro, 1=acerto, 8=selecionado mas nao respondido, 9=
    ainda nao selecionado),
143     # 6-ordem de selecao.
144     #####
145
146     banco <- cbind(c(1:n.itens),a.par,b.par,c.par,rep(9,n.itens),rep(0,n.
    itens))
147
148     n.it.resp <- 0 # numero cumulativo de itens respondidos
149
150     while ((n.it.resp <= n.max)&&(psd>tol)) {
151

```

```

152 #   banco.aux <- banco[banco[,5]==9,]
153   banco.aux <- matrix(banco[banco[,5]==9,], ncol=6)
154
155   ind.n.esc <- banco.aux[,1]
156   a.n.esc <- banco.aux[,2]
157   b.n.esc <- banco.aux[,3]
158   c.n.esc <- banco.aux[,4]
159
160   vec.prob <- c.n.esc + (1-c.n.esc)/(1+exp(-a.n.esc*(theta.est-b.n.esc)
161   ))
162
163   vec.info <- (a.n.esc^2)*((1-vec.prob)/vec.prob)*(((vec.prob-c.n.esc)/
164   (1-c.n.esc))^2)
165
166   ind <- ind.n.esc[vec.info==max(vec.info)]
167
168   nS[ind] <- nS[ind]+1
169
170   x <- runif(1,0,1)
171
172   if (x < PAS[ind]) {
173
174     nA[ind] <- nA[ind] + 1
175
176     n.it.resp <- n.it.resp + 1 # numero de itens respondidos
177
178     a.esc <- banco[ind,2]
179     b.esc <- banco[ind,3]
180     c.esc <- banco[ind,4]
181
182     p.acerto <- c.esc+(1-c.esc)/(1+exp(-a.esc*(theta-b.esc)))
183     u <- sign(sign(p.acerto-runif(1,0,1))+1) # resposta
184
185     banco[ind,5] <- u
186     banco[ind,6] <- n.it.resp
187
188     #####
189     ## calculo de theta.est via EAP com priori N(0,1)
190     #####
191
192     # os vetores a.aux, b.aux, c.aux e resp.aux contem os valores dos
193     # parametros
194     # dos itens e as respostas para os itens ja' respondidos
195
196     a.aux <- banco[banco[,5]<2,2]
197     b.aux <- banco[banco[,5]<2,3]
198     c.aux <- banco[banco[,5]<2,4]
199
200     resp.aux <- banco[banco[,5]<2,5]
201
202     theta.est <- integrate(numerador,-3,3)$value/integrate(denominador

```

```

    ,-3,3)$value
200     theta.est2 <- integrate(numerador2,-3,3)$value/integrate(
denominador,-3,3)$value
201
202     psd <- theta.est2-theta.est^2
203     psd.acum <- c(psd.acum,psd)
204     }
205
206     if (x>=PAS[ind]) banco[ind,5] <- 8
207
208     }
209
210     vec.theta.est <- c(vec.theta.est,theta.est)
211
212     }
213
214     PS <- nS/n.resp.SH
215 # PS <- length(banco[banco[,5]<=8,5])
216     PA <- nA/n.resp.SH
217
218     PAS[PS>0] <- nA[nS>0]/nS[nS>0]
219
220     PAS[PS>r.max] <- r.max/PS[PS>r.max]*.95
221     PAS[PS<=r.max] <- 1
222
223 #print(c("PAS<1"))
224 #print(PAS[PAS<1])
225
226 #print(c("PA>r.max"))
227 #print(PA[PA>r.max])
228
229 }
230
231
232 #####
233 ### Aplicacao do CAT com maxima informacao e restricao de SH
234 #####
235
236 prop <- rep(0,n.itens)
237 vec.theta.est <- numeric(0)
238 vec.n.it.resp <- numeric(0)
239
240 vec.theta <- rnorm(n.resp,0,1) # proficiencia real do respondente
241
242 n.max <- 60 # numero maximo de itens respondidos
243
244 # matriz indicadora de aplicacao dos itens aos respondentes (linha=
respondente, coluna=item)
245 mat.overlap <- matrix(0,n.resp,n.itens)
246
247 for (j in 1:n.resp) {

```

```

248
249 # print(j)
250
251 theta.est <- 0 # estimativa de theta
252
253 psd.acum <- numeric(0) # vetor com sequencia de estimativas do desvio-
    padrao a posteriori de theta.est
254
255 tol <- 0.1 # limite de tolerancia para convergencia da estimativa de
    theta.est[j] para psd < tol
256 psd <- 1 # valor inicial do desvio padrao a posteriori de theta.est
257
258 #####
259 # banco de itens: colunas da matriz "banco"
260 # 1-indice, 2-a.par, 3-b.par, 4-c.par,
261 # 5-resposta (0=erro, 1=acerto, 8=escolhido mas nao respondido, 9=nao
    escolhido),
262 # 6-ordem de selecao.
263 #####
264
265 banco <- cbind(c(1:n.itens), a.par, b.par, c.par, rep(9, n.itens), rep(0, n.
    itens), PAS)
266
267 n.it.resp <- 0 # numero cumulativo de itens respondidos
268
269 while ((n.it.resp <= n.max) && (psd > tol)) {
270
271     banco.aux <- banco[banco[, 5] == 9, ]
272
273     ind.n.esc <- banco.aux[, 1]
274     a.n.esc <- banco.aux[, 2]
275     b.n.esc <- banco.aux[, 3]
276     c.n.esc <- banco.aux[, 4]
277     PAS.n.esc <- banco.aux[, 7]
278
279     vec.prob <- c.n.esc + (1 - c.n.esc) / (1 + exp(-a.n.esc * (theta.est - b.n.esc)))
280
281     vec.info <- (a.n.esc^2) * ((1 - vec.prob) / vec.prob) * (((vec.prob - c.n.esc) /
    (1 - c.n.esc))^2)
282
283     ind <- ind.n.esc[vec.info == max(vec.info)]
284
285     x <- runif(1, 0, 1)
286
287     if (x < PAS[ind]) {
288
289         prop[ind] <- prop[ind] + 1
290
291         n.it.resp <- n.it.resp + 1 # numero de itens respondidos
292
293         a.esc <- banco[ind, 2]

```

```

294     b.esc <- banco[ind,3]
295     c.esc <- banco[ind,4]
296
297     p.acerto <- c.esc+(1-c.esc)/(1+exp(-a.esc*(vec.theta[j]-b.esc)))
298     u <- sign(sign(p.acerto-runif(1,0,1))+1) # resposta
299
300     banco[ind,5] <- u
301     banco[ind,6] <- n.it.resp
302
303     #####
304     ## calculo de theta.est via EAP com priori N(0,1)
305     #####
306
307     # os vetores a.aux, b.aux, c.aux e resp.aux contem os valores dos
parametros
308     # dos itens e as respostas para os itens ja' respondidos
309
310     a.aux <- banco[banco[,5]<2,2]
311     b.aux <- banco[banco[,5]<2,3]
312     c.aux <- banco[banco[,5]<2,4]
313
314     resp.aux <- banco[banco[,5]<2,5]
315
316     theta.est <- integrate(numerador,-3,3)$value/integrate(denominador
,-3,3)$value
317     theta.est2 <- integrate(numerador2,-3,3)$value/integrate(denominador
,-3,3)$value
318
319     psd <- theta.est2-theta.est^2
320     psd.acum <- c(psd.acum,psd)
321   }
322
323   if (x >= PAS[ind]) banco[ind,5] <- 8
324
325 }
326
327 vec.theta.est <- c(vec.theta.est,theta.est)
328 vec.n.it.resp <- c(vec.n.it.resp,n.it.resp)
329
330 vec.aux <- banco[,5]
331 vec.aux[vec.aux>1] <- 0
332 mat.overlap[j,] <- vec.aux
333
334 }
335
336
337 overlap.theta.dif <- overlap.vec <- numeric(0)
338
339 overlap.soma <- 0
340 tam.am <- 10000 # tamanho da amostra de pares de respondentes para
estimacao da taxa de overlap

```

```

341
342 for (k in 1:tam.am) {
343   am.resp <- sample(c(1:n.resp),2,replace=FALSE)
344   am.overlap <- sum(mat.overlap[am.resp[1],]*mat.overlap[am.resp[2],])
345   overlap.soma <- overlap.soma + am.overlap
346   overlap.vec <- c(overlap.vec,am.overlap)
347   overlap.theta.dif <- c(overlap.theta.dif,abs(vec.theta[am.resp[1]]-vec.
      theta[am.resp[2]]))
348 }
349
350 ## proporcao media de coincidencia de itens entre respondentes
351 overlap.rate <- 100*overlap.soma/(n.itens*tam.am)
352 #overlap.rate
353
354 ## Taxa de exposi o dos itens
355
356 tx.expos <- prop/n.resp
357 tx.expos
358 prop.it.exp <- length(tx.expos[tx.expos>0])/n.itens
359 prop.it.sup.exp <- length(tx.expos[tx.expos>r.max])/n.itens
360 prop.it.subut <- length(tx.expos[tx.expos<0.02])/n.itens
361
362 #mean(tx.expos) # taxa media de exposicao por item
363 #sd(tx.expos) # desvio-padrao da taxa de exposicao dos itens
364
365 ## Fidelidade
366 fid <-cor(vec.theta,vec.theta.est)
367 #fid
368
369 ## V cio
370 vicio <- sum(vec.theta.est-vec.theta)/n.resp
371 #vicio
372
373 ## V cio separado por tr s grupos
374 vicio.A <- sum(vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
      /length(vec.theta[vec.theta < -1.5])
375 vicio.B <- sum(vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta
      ) <= 1.5])/
376   length(vec.theta[abs(vec.theta) <= 1.5])
377 vicio.C <- sum(vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
      /length(vec.theta[vec.theta > -1.5])
378
379 #vicio.A
380 #vicio.B
381 #vicio.C
382
383 ## Erro quadr tico m dio (MSE)
384 eqm <-sum((vec.theta.est-vec.theta)^2)/n.resp
385 #eqm
386
387 ## Erro quadr tico m dio (MSE) separado por tr s grupos

```

```

388 eqm.A <- sum((vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
  ^2)/length(vec.theta[vec.theta < -1.5])
389 eqm.B <- sum((vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta)
  <= 1.5])^2)/
390   length(vec.theta[abs(vec.theta) <= 1.5])
391 eqm.C <- sum((vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
  ^2)/length(vec.theta[vec.theta > -1.5])
392
393 #eqm.A
394 #eqm.B
395 #eqm.C
396
397 ## Proporcao de respondentes que atingiu o numero maximo de itens permitido
398 prop.it.resp.lim <- length(vec.n.it.resp[vec.n.it.resp==(n.max+1)])/n.resp
399 #prop.it.resp.lim
400
401 ## Numero medio de itens respondidos por respondente
402 no.med.it.resp <- mean(vec.n.it.resp)
403 #no.med.it.resp
404
405 ## Desvio-padrao do numero de itens respondidos
406 dp.no.it.resp <- sd(vec.n.it.resp)
407 #dp.no.it.resp
408
409 repet.overlap <- c(repet.overlap, overlap.rate)
410 repet.it.exp <- c(repet.it.exp, prop.it.exp)
411 repet.it.sup <- c(repet.it.sup, prop.it.sup.exp)
412 repet.it.subut <- c(repet.it.subut, prop.it.subut)
413 repet.fid <- c(repet.fid, fid)
414 repet.vicio <- c(repet.vicio, vicio)
415 repet.vicio.A <- c(repet.vicio.A, vicio.A)
416 repet.vicio.B <- c(repet.vicio.B, vicio.B)
417 repet.vicio.C <- c(repet.vicio.C, vicio.C)
418 repet.eqm <- c(repet.eqm, eqm)
419 repet.eqm.A <- c(repet.eqm.A, eqm.A)
420 repet.eqm.B <- c(repet.eqm.B, eqm.B)
421 repet.eqm.C <- c(repet.eqm.C, eqm.C)
422 repet.it.resp.lim <- c(repet.it.resp.lim, prop.it.resp.lim)
423 repet.no.it.resp <- c(repet.no.it.resp, no.med.it.resp)
424 repet.dp.no.it.resp <- c(repet.dp.no.it.resp, dp.no.it.resp)
425
426 } # fim do loop das repeticoes das simulacoes
427
428 resultados <- rbind(repet.overlap, repet.it.exp, repet.it.sup, repet.it.
  subut, repet.fid,
429   repet.vicio, repet.vicio.A, repet.vicio.B, repet.vicio.C, repet.eqm,
  repet.eqm.A,
430   repet.eqm.B, repet.eqm.C, repet.it.resp.lim, repet.no.it.resp, repet.dp.
  no.it.resp)
431
432 vec.media <- round(apply(resultados, 1, mean), digits=5)

```

```

433 vec.erro.padrao <- round(apply(resultados,1,sd),digits=5)
434 sumario.media.ep <- cbind(vec.media,vec.erro.padrao)
435 sumario.media.ep <- matrix(sumario.media.ep,ncol=2)

```

### Código em R para implementação do método de Máxima Informação Ponderada

```

1 #####
2 ### METODO DE ESCOLHA POR MAXIMA INFORMACAO PONDERADA POR ((1-p)^alfa)*I(
   theta^)(REPETI ES)
3 ### NUMERO ALEATORIO DE ITENS POR TESTE
4 ### NUMERO MAXIMO DE ITENS RESPONDIDOS = 60 (n.max)
5 ### NUMERO DE ITENS NO BANCO DE ITENS = 400 (n.itens)
6 ### NUMERO DE RESPONDENTES = 3000 (n.resp)
7 #####
8
9 #####
10 ### CAT
11 ### - Estimacao da proficiencia via EAP.
12 ### - Escolha do item com maior valor para a informacao ponderada em theta
   estimado.
13 ### - Criterio de parada: PSD.
14 ### - Integracao via retangulos
15 #####
16
17 #####
18 ### Funcao "veross" #####
19 ### calcula o valor da verossimilhanca ###
20 #####
21 veross <- function(par) {
22   prod <- 1
23   for (k in 1:length(resp.aux)) {
24     p.aux <- c.aux[k] + (1-c.aux[k])/(1+exp(-a.aux[k]*(par-b.aux[k])))
25     prod <- prod*(p.aux^resp.aux[k])*(1-p.aux)^(1-resp.aux[k])
26   }
27   return(prod)
28 }
29
30 #####
31 ### Funcao "numerador" #####
32 ### numerador a ser integrado em theta no calculo de EAP ###
33 #####
34 numerador <- function(par) {
35   par*veross(par)*dnorm(par,0,1)
36 }
37
38 #####
39 ### Funcao "numerador2" #####
40 ### numerador a ser integrado em theta no calculo de EAP ###
41 #####
42 numerador2 <- function(par) {
43   (par^2)*veross(par)*dnorm(par,0,1)
44 }

```

```

45
46 #####
47 ### Funcao "denominador" ###
48 ### denominador a ser integrado em theta no calculo de EAP ###
49 #####
50 denominador <- function(par) {
51   veross(par)*dnorm(par,0,1)
52 }
53
54 #####
55 ### Geracao dos parametros dos itens e do valor real de theta ###
56 #####
57
58 set.seed(23456)
59 n.itens <- 400 # numero de itens no banco de quest es
60 b.par <- rnorm(n.itens,0,1) # valores dos parametros de dificuldade
61 a.par <- runif(n.itens,0.3,2.5) # valores dos parametros de discrimina o
62 c.par <- rep(0.2,n.itens)# valores dos parametros de acerto ao acaso
63
64 n.resp <- 3000 # numero de respondentes
65
66 r.max <- 0.2 # taxa maxima de exposicao desejada
67
68 alfa <- 8 # parametro que altera os pesos da informacao de cada item nao
   escolhido
69
70 amp <- 0.001 # amplitude dos intervalos na integracao via retangulos
71 param <- seq(-3,3,amp) # valores de theta utilizados na integracao numerica
72
73 n.max <- 60 # numero maximo de itens respondidos
74
75 #####
76 ## Inicio das repeticoes
77 #####
78
79 n.rep <- 5 # numero de repeticoes na simulacao
80
81 repet.overlap <- numeric(0) # vetor com taxas de sobreposicao de cada
   simulacao
82 repet.it.exp <- numeric(0) # vetor com proporcao de itens expostos de cada
   simulacao
83 repet.it.sup <- numeric(0) # vetor com proporcao de itens superexpostos de
   cada simulacao
84 repet.it.subut <- numeric(0) # vetor com proporcao de itens subutilizados
   de cada simulacao
85 repet.fid <- numeric(0) # vetor com a fidelidade (correlacao entre theta e
   sua estimativa) de cada simulacao
86 repet.vicio <- numeric(0) # vetor com vicio medio das estimativas de theta
   de cada simulacao
87 repet.vicio.A <- numeric(0) # vetor com vicio medio das estimativas de
   theta p/ theta < -1.5 de cada simulacao

```

```

88 repet.vicio.B <- numeric(0) # vetor com vicio medio das estimativas de
   theta p/ -1.5 < theta < 1.5 de cada simulacao
89 repet.vicio.C <- numeric(0) # vetor com vicio medio das estimativas de
   theta p/ theta > 1.5 de cada simulacao
90 repet.eqm <- numeric(0) # vetor com erro quadratico medio de cada simulacao
91 repet.eqm.A <- numeric(0) # vetor com erro quadratico medio p/ theta < -1.5
   de cada simulacao
92 repet.eqm.B <- numeric(0) # vetor com erro quadratico medio p/ -1.5 < theta
   < 1.5 de cada simulacao
93 repet.eqm.C <- numeric(0) # vetor com erro quadratico medio p/ theta > 1.5
   de cada simulacao
94 repet.it.resp.lim <- numeric(0) # vetor com proporcao de respondentes no
   limite de itens de cada simulacao
95 repet.no.it.resp <- numeric(0) # vetor com numero medio de itens
   respondidos de cada simulacao
96 repet.dp.no.it.resp <- numeric(0) # vetor com desvio-padrao do numero de
   itens respondidos de cada simulacao
97
98 ## Inicio do loop das repeticoes das simulacoes
99 for (m in 1:n.rep) {
100
101 print(m)
102
103 vec.theta <- rnorm(n.resp,0,1) # proficiencia real do respondente
104 vec.theta.est <- numeric(0) # vetor com estimativas de theta
105
106 vec.n.it.resp <- numeric(0) # vetor com numero de itens respondidos
107
108 prop <- rep(0,n.itens) # proporcao de respondentes que responderam cada
   item
109
110 # matriz indicadora de aplicacao dos itens aos respondentes (linha=
   respondente, coluna=item)
111 mat.overlap <- matrix(0,n.resp,n.itens)
112
113 for (j in 1:n.resp) {
114
115 # print(j)
116
117 theta.est <- 0 # estimativa de theta
118
119 psd.acum <- numeric(0) # vetor com sequencia de estimativas do desvio-
   padrao a posteriori de theta.est
120
121 tol <- 0.1
122 psd <- 1
123
124 #####
125 # banco de itens: colunas da matriz "banco"
126 # 1-indice, 2-a.par, 3-b.par, 4-c.par,
127 # 5-resposta (0=erro, 1=acerto, 9=ainda nao respondido),

```

```

128 # 6-ordem de selecao.
129 #####
130
131 banco <- cbind(c(1:n.itens),a.par,b.par,c.par,rep(9,n.itens),rep(0,n.
  itens))
132
133 n.it.resp <- 0 # numero cumulativo de itens respondidos
134
135 while ((n.it.resp <= n.max)&&(psd>tol)) {
136
137   banco.aux <- banco[banco[,5]==9,]
138
139   ind.n.esc <- banco.aux[,1]
140   a.n.esc <- banco.aux[,2]
141   b.n.esc <- banco.aux[,3]
142   c.n.esc <- banco.aux[,4]
143
144   vec.prob <- c.n.esc + (1-c.n.esc)/(1+exp(-a.n.esc*(theta.est-b.n.esc)))
145
146   vec.info <- (a.n.esc^2)*((1-vec.prob)/vec.prob)*((vec.prob-c.n.esc)/
  (1-c.n.esc))^2)
147 #   vec.info <- vec.info/mean(vec.info)
148   vec.info <- vec.info/sum(vec.info)
149
150 #   const <- 1
151
152   pesos <- (((1-prop[ind.n.esc]/j)^alfa)*vec.info)/sum(((1-prop[ind.n.esc
  ]/j)^alfa)*vec.info)
153
154   ind <- ind.n.esc[pesos==max(pesos)]
155
156   prop[ind] <- prop[ind] + 1
157
158   n.it.resp <- n.it.resp + 1 # numero de itens respondidos
159
160   a.esc <- banco[ind,2]
161   b.esc <- banco[ind,3]
162   c.esc <- banco[ind,4]
163
164   p.acerto <- c.esc+(1-c.esc)/(1+exp(-a.esc*(vec.theta[j]-b.esc)))
165   u <- sign(sign(p.acerto-runif(1,0,1))+1) # resposta
166
167   banco[ind,5] <- u
168   banco[ind,6] <- n.it.resp
169
170   #####
171   ## calculo de theta.est via EAP com priori N(0,1)
172   #####
173
174   # os vetores a.aux, b.aux, c.aux e resp.aux contem os valores dos
  parametros

```

```

175 # dos itens e as respostas para os itens ja' respondidos
176
177 a.aux <- banco[banco[,5] !=9,2]
178 b.aux <- banco[banco[,5] !=9,3]
179 c.aux <- banco[banco[,5] !=9,4]
180
181 resp.aux <- banco[banco[,5] !=9,5]
182
183 theta.est <- integrate(numerador,-3,3)$value/integrate(denominador
184 ,-3,3)$value
185 theta.est2 <- integrate(numerador2,-3,3)$value/integrate(denominador
186 ,-3,3)$value
187
188 psd <- theta.est2-theta.est^2
189 psd.acum <- c(psd.acum,psd)
190
191 }
192
193 vec.theta.est <- c(vec.theta.est,theta.est)
194 vec.n.it.resp <- c(vec.n.it.resp,n.it.resp)
195
196 vec.aux <- banco[,5]
197 vec.aux[vec.aux>1] <- 0
198 mat.overlap[j,] <- vec.aux
199
200 }
201 overlap.theta.dif <- overlap.vec <- numeric(0)
202
203 overlap.soma <- 0
204 tam.am <- 10000 # tamanho da amostra de pares de respondentes para
205 estimacao da taxa de overlap
206
207 for (k in 1:tam.am) {
208   am.resp <- sample(c(1:n.resp),2,replace=FALSE)
209   am.overlap <- sum(mat.overlap[am.resp[1],]*mat.overlap[am.resp[2],])
210   overlap.soma <- overlap.soma + am.overlap
211   overlap.vec <- c(overlap.vec,am.overlap)
212   overlap.theta.dif <- c(overlap.theta.dif,abs(vec.theta[am.resp[1]]-vec.
213     theta[am.resp[2]]))
214 }
215
216 ## proporcao media de coincidencia de itens entre respondentes
217 overlap.rate <- 100*overlap.soma/(n.itens*tam.am)
218 #overlap.rate
219
220 ## Taxa de exposi o dos itens
221
222 tx.expos <- prop/n.resp
223 tx.expos
224 prop.it.exp <- length(tx.expos[tx.expos>0])/n.itens
225 prop.it.sup.exp <- length(tx.expos[tx.expos>r.max])/n.itens

```

```

222 prop.it.subut <- length(tx.expos[tx.expos<0.02])/n.itens
223
224 #mean(tx.expos) # taxa media de exposicao por item
225 #sd(tx.expos) # desvio-padrao da taxa de exposicao dos itens
226
227 ## Fidelidade
228 fid <-cor(vec.theta,vec.theta.est)
229 #fid
230
231 ## Vicio
232 vicio <- sum(vec.theta.est-vec.theta)/n.resp
233 #vicio
234
235 ## Vicio separado por tr s grupos
236 vicio.A <- sum(vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
237           /length(vec.theta[vec.theta < -1.5])
238 vicio.B <- sum(vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta
239           ) <= 1.5])/
240           length(vec.theta[abs(vec.theta) <= 1.5])
241 vicio.C <- sum(vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
242           /length(vec.theta[vec.theta > -1.5])
243
244 #vicio.A
245 #vicio.B
246 #vicio.C
247
248 ## Erro quadr tico m dio (MSE)
249 eqm <-sum((vec.theta.est-vec.theta)^2)/n.resp
250 #eqm
251
252 ## Erro quadr tico m dio (MSE) separado por tr s grupos
253 eqm.A <- sum((vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
254           ^2)/length(vec.theta[vec.theta < -1.5])
255 eqm.B <- sum((vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta)
256           <= 1.5])^2)/
257           length(vec.theta[abs(vec.theta) <= 1.5])
258 eqm.C <- sum((vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
259           ^2)/length(vec.theta[vec.theta > -1.5])
260
261 #eqm.A
262 #eqm.B
263 #eqm.C
264
265 ## Proporciao de respondentes que atingiu o numero maximo de itens permitido
266 prop.it.resp.lim <- length(vec.n.it.resp[vec.n.it.resp==(n.max+1)])/n.resp
267 #prop.it.resp.lim
268
269 ## Numero medio de itens respondidos por respondente
270 no.med.it.resp <- mean(vec.n.it.resp)
271 #no.med.it.resp

```

```

267 ## Desvio-padroao do numero de itens respondidos
268 dp.no.it.resp <- sd(vec.n.it.resp)
269 #dp.no.it.resp
270
271 repet.overlap <- c(repet.overlap, overlap.rate)
272 repet.it.exp <- c(repet.it.exp, prop.it.exp)
273 repet.it.sup <- c(repet.it.sup, prop.it.sup.exp)
274 repet.it.subut <- c(repet.it.subut, prop.it.subut)
275 repet.fid <- c(repet.fid, fid)
276 repet.vicio <- c(repet.vicio, vicio)
277 repet.vicio.A <- c(repet.vicio.A, vicio.A)
278 repet.vicio.B <- c(repet.vicio.B, vicio.B)
279 repet.vicio.C <- c(repet.vicio.C, vicio.C)
280 repet.eqm <- c(repet.eqm, eqm)
281 repet.eqm.A <- c(repet.eqm.A, eqm.A)
282 repet.eqm.B <- c(repet.eqm.B, eqm.B)
283 repet.eqm.C <- c(repet.eqm.C, eqm.C)
284 repet.it.resp.lim <- c(repet.it.resp.lim, prop.it.resp.lim)
285 repet.no.it.resp <- c(repet.no.it.resp, no.med.it.resp)
286 repet.dp.no.it.resp <- c(repet.dp.no.it.resp, dp.no.it.resp)
287
288 } # fim do loop das repeticoes das simulacoes
289
290 resultados <- rbind(repet.overlap, repet.it.exp, repet.it.sup, repet.it.
  subut, repet.fid,
291   repet.vicio, repet.vicio.A, repet.vicio.B, repet.vicio.C, repet.eqm,
  repet.eqm.A,
292   repet.eqm.B, repet.eqm.C, repet.it.resp.lim, repet.no.it.resp, repet.dp.
  no.it.resp)
293
294 vec.media <- round(apply(resultados, 1, mean), digits=5)
295 vec.erro.padroao <- round(apply(resultados, 1, sd), digits=5)
296 sumario.media.ep <- cbind(vec.media, vec.erro.padroao)
297 sumario.media.ep <- matrix(sumario.media.ep, ncol=2)
298 sumario.media.ep

```

### Código em R para implementação do método de Informação Aleatória Ponderada

```

1 #####
2 ### METODO DE ESCOLHA INFORMACAO ALEATORIA PONDERADA POR ((1-p)^alfa)*I(
  theta^) (REPETI ES)
3 ### NUMERO ALEATORIO DE ITENS POR TESTE
4 ### NUMERO MAXIMO DE ITENS RESPONDIDOS = 60 (n.max)
5 ### NUMERO DE ITENS NO BANCO DE ITENS = 400 (n.itens)
6 ### NUMERO DE RESPONDENTES = 3000 (n.resp)
7 #####
8
9 #####
10 ### CAT
11 ### - Estimacao da proficiencia via EAP.
12 ### - Escolha do item com maior valor para a informacao ponderada em theta
  estimado.

```

```

13 ### - Criterio de parada: PSD.
14 ### - Integracao via retangulos
15 #####
16
17 #####
18 ### Funcao "veross" ###
19 ### calcula o valor da verossimilhanca ###
20 #####
21 veross <- function(par) {
22   prod <- 1
23   for (k in 1:length(resp.aux)) {
24     p.aux <- c.aux[k] + (1-c.aux[k]) / (1+exp(-a.aux[k]*(par-b.aux[k])))
25     prod <- prod*(p.aux^resp.aux[k])*(1-p.aux)^(1-resp.aux[k])
26   }
27   return(prod)
28 }
29
30 #####
31 ### Funcao "numerador" ###
32 ### numerador a ser integrado em theta no calculo de EAP ###
33 #####
34 numerador <- function(par) {
35   par*veross(par)*dnorm(par,0,1)
36 }
37
38 #####
39 ### Funcao "numerador2" ###
40 ### numerador a ser integrado em theta no calculo de EAP ###
41 #####
42 numerador2 <- function(par) {
43   (par^2)*veross(par)*dnorm(par,0,1)
44 }
45
46 #####
47 ### Funcao "denominador" ###
48 ### denominador a ser integrado em theta no calculo de EAP ###
49 #####
50 denominador <- function(par) {
51   veross(par)*dnorm(par,0,1)
52 }
53
54 #####
55 ### Geracao dos parametros dos itens e do valor real de theta ###
56 #####
57
58 set.seed(23456)
59 n.itens <- 400 # numero de itens no banco de quest es
60 b.par <- rnorm(n.itens,0,1) # valores dos parametros de dificuldade
61 a.par <- runif(n.itens,0.3,2.5) # valores dos parametros de discrimina o
62 c.par <- rep(0.2,n.itens)# valores dos parametros de acerto ao acaso
63

```

```
64 n.resp <- 3000 # numero de respondentes
65
66 r.max <- 0.2 # taxa maxima de exposicao desejada
67
68 alfa <- 3 # parametro que altera os pesos da informacao de cada item nao
    escolhido
69
70 amp <- 0.001 # amplitude dos intervalos na integracao via retangulos
71 param <- seq(-3,3,amp) # valores de theta utilizados na integracao numerica
72
73 n.max <- 60 # numero maximo de itens respondidos
74
75 #####
76 ## Inicio das repeticoes
77 #####
78
79 n.rep <- 20 # numero de repeticoes na simulacao
80
81 repet.overlap <- numeric(0) # vetor com taxas de sobreposicao de cada
    simulacao
82 repet.it.exp <- numeric(0) # vetor com proporcao de itens expostos de cada
    simulacao
83 repet.it.sup <- numeric(0) # vetor com proporcao de itens superexpostos de
    cada simulacao
84 repet.it.subut <- numeric(0) # vetor com proporcao de itens subutilizados
    de cada simulacao
85 repet.fid <- numeric(0) # vetor com a fidelidade (correlacao entre theta e
    sua estimativa) de cada simulacao
86 repet.vicio <- numeric(0) # vetor com vicio medio das estimativas de theta
    de cada simulacao
87 repet.vicio.A <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ theta < -1.5 de cada simulacao
88 repet.vicio.B <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ -1.5 < theta < 1.5 de cada simulacao
89 repet.vicio.C <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ theta > 1.5 de cada simulacao
90 repet.eqm <- numeric(0) # vetor com erro quadratico medio de cada simulacao
91 repet.eqm.A <- numeric(0) # vetor com erro quadratico medio p/ theta < -1.5
    de cada simulacao
92 repet.eqm.B <- numeric(0) # vetor com erro quadratico medio p/ -1.5 < theta
    < 1.5 de cada simulacao
93 repet.eqm.C <- numeric(0) # vetor com erro quadratico medio p/ theta > 1.5
    de cada simulacao
94 repet.it.resp.lim <- numeric(0) # vetor com proporcao de respondentes no
    limite de itens de cada simulacao
95 repet.no.it.resp <- numeric(0) # vetor com numero medio de itens
    respondidos de cada simulacao
96 repet.dp.no.it.resp <- numeric(0) # vetor com desvio-padrao do numero de
    itens respondidos de cada simulacao
97
98
```

```

99 ## Inicio do loop das repeticoes das simulacoes
100 for (m in 1:n.rep) {
101
102 print(m)
103
104 prop <- rep(0,n.itens) # proporcao de respondentes que responderam cada
    item
105
106 vec.theta <- rnorm(n.resp,0,1) # proficiencia real do respondente
107 vec.theta.est <- numeric(0) # vetor com estimativas de theta
108
109 vec.n.it.resp <- numeric(0) # vetor com numero de itens respondidos
110
111 # matriz indicadora de aplicacao dos itens aos respondentes (linha=
    respondente, coluna=item)
112 mat.overlap <- matrix(0,n.resp,n.itens)
113
114 for (j in 1:n.resp) {
115
116 # print(j)
117
118 theta.est <- 0 # estimativa de theta
119
120 psd.acum <- numeric(0) # vetor com sequencia de estimativas do desvio-
    padrao a posteriori de theta.est
121
122 tol <- 0.1
123 psd <- 1
124
125 #####
126 # banco de itens: colunas da matriz "banco"
127 # 1-indice, 2-a.par, 3-b.par, 4-c.par,
128 # 5-resposta (0=erro, 1=acerto, 9=ainda nao respondido),
129 # 6-ordem de selecao.
130 #####
131
132 banco <- cbind(c(1:n.itens),a.par,b.par,c.par,rep(9,n.itens),rep(0,n.
    itens))
133
134 n.it.resp <- 0 # numero cumulativo de itens respondidos
135
136 while ((n.it.resp <= n.max)&&(psd>tol)) {
137
138     banco.aux <- banco[banco[,5]==9,]
139
140     ind.n.esc <- banco.aux[,1]
141     a.n.esc <- banco.aux[,2]
142     b.n.esc <- banco.aux[,3]
143     c.n.esc <- banco.aux[,4]
144
145     vec.prob <- c.n.esc + (1-c.n.esc)/(1+exp(-a.n.esc*(theta.est-b.n.esc)))

```

```

146
147   vec.info <- (a.n.esc^2)*((1-vec.prob)/vec.prob)*(((vec.prob-c.n.esc)/
148   #   vec.info <- vec.info/mean(vec.info)
149   vec.info <- vec.info/sum(vec.info)
150
151 #   const <- 1
152
153   pesos <- (((1-prop[ind.n.esc]/j)^alfa)*vec.info)/sum(((1-prop[ind.n.esc
154 ]/j)^alfa)*vec.info)
155
156   ind <- sample(ind.n.esc, size=1, replace=FALSE, prob=pesos)
157
158   prop[ind] <- prop[ind] + 1
159
160   n.it.resp <- n.it.resp + 1 # numero de itens respondidos
161
162   a.esc <- banco[ind,2]
163   b.esc <- banco[ind,3]
164   c.esc <- banco[ind,4]
165
166   p.acerto <- c.esc+(1-c.esc)/(1+exp(-a.esc*(vec.theta[j]-b.esc)))
167   u <- sign(sign(p.acerto-runif(1,0,1))+1) # resposta
168
169   banco[ind,5] <- u
170   banco[ind,6] <- n.it.resp
171
172   #####
173   ## calculo de theta.est via EAP com priori N(0,1)
174   #####
175
176   # os vetores a.aux, b.aux, c.aux e resp.aux contem os valores dos
177   # dos itens e as respostas para os itens ja' respondidos
178
179   a.aux <- banco[banco[,5]!=9,2]
180   b.aux <- banco[banco[,5]!=9,3]
181   c.aux <- banco[banco[,5]!=9,4]
182
183   resp.aux <- banco[banco[,5]!=9,5]
184
185   theta.est <- integrate(numerador,-3,3)$value/integrate(denominador
186   ,-3,3)$value
187   theta.est2 <- integrate(numerador2,-3,3)$value/integrate(denominador
188   ,-3,3)$value
189
190   psd <- theta.est2-theta.est^2
191   psd.acum <- c(psd.acum,psd)
192
193 }

```

```

192 vec.theta.est <- c(vec.theta.est,theta.est)
193 vec.n.it.resp <- c(vec.n.it.resp,n.it.resp)
194
195 vec.aux <- banco[,5]
196 vec.aux[vec.aux>1] <- 0
197 mat.overlap[j,] <- vec.aux
198
199 }
200 overlap.theta.dif <- overlap.vec <- numeric(0)
201
202 overlap.soma <- 0
203 tam.am <- 10000 # tamanho da amostra de pares de respondentes para
                # estimacao da taxa de overlap
204
205 for (k in 1:tam.am) {
206   am.resp <- sample(c(1:n.resp),2,replace=FALSE)
207   am.overlap <- sum(mat.overlap[am.resp[1],]*mat.overlap[am.resp[2],])
208   overlap.soma <- overlap.soma + am.overlap
209   overlap.vec <- c(overlap.vec,am.overlap)
210   overlap.theta.dif <- c(overlap.theta.dif,abs(vec.theta[am.resp[1]]-vec.
                theta[am.resp[2]]))
211 }
212
213 ## proporcao media de coincidencia de itens entre respondentes
214 overlap.rate <- 100*overlap.soma/(n.itens*tam.am)
215 #overlap.rate
216
217
218
219 ## Taxa de exposi o dos itens
220
221 tx.expos <- prop/n.resp
222 tx.expos
223 prop.it.exp <- length(tx.expos[tx.expos>0])/n.itens
224 prop.it.sup.exp <- length(tx.expos[tx.expos>r.max])/n.itens
225 prop.it.subut <- length(tx.expos[tx.expos<0.02])/n.itens
226
227 #mean(tx.expos) # taxa media de exposicao por item
228 #sd(tx.expos) # desvio-padrao da taxa de exposicao dos itens
229
230 ## Fidelidade
231 fid <-cor(vec.theta,vec.theta.est)
232 #fid
233
234 ## V cio
235 vicio <- sum(vec.theta.est-vec.theta)/n.resp
236 #vicio
237
238 ## V cio separado por tr s grupos
239 vicio.A <- sum(vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
                /length(vec.theta[vec.theta < -1.5])

```

```

240 vicio.B <- sum(vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta)
    ) <= 1.5])/
241     length(vec.theta[abs(vec.theta) <= 1.5])
242 vicio.C <- sum(vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
    /length(vec.theta[vec.theta > -1.5])
243
244 #vicio.A
245 #vicio.B
246 #vicio.C
247
248 ## Erro quadrático médio (MSE)
249 eqm <- sum((vec.theta.est-vec.theta)^2)/n.resp
250 #eqm
251
252 ## Erro quadrático médio (MSE) separado por três grupos
253 eqm.A <- sum((vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
    ^2)/length(vec.theta[vec.theta < -1.5])
254 eqm.B <- sum((vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta)
    <= 1.5])^2)/
255     length(vec.theta[abs(vec.theta) <= 1.5])
256 eqm.C <- sum((vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
    ^2)/length(vec.theta[vec.theta > -1.5])
257
258 #eqm.A
259 #eqm.B
260 #eqm.C
261
262 ## Proporção de respondentes que atingiu o número máximo de itens permitido
263 prop.it.resp.lim <- length(vec.n.it.resp[vec.n.it.resp==(n.max+1)])/n.resp
264 #prop.it.resp.lim
265
266 ## Número médio de itens respondidos por respondente
267 no.med.it.resp <- mean(vec.n.it.resp)
268 #no.med.it.resp
269
270 ## Desvio-padrão do número de itens respondidos
271 dp.no.it.resp <- sd(vec.n.it.resp)
272 #dp.no.it.resp
273
274 repet.overlap <- c(repet.overlap, overlap.rate)
275 repet.it.exp <- c(repet.it.exp, prop.it.exp)
276 repet.it.sup <- c(repet.it.sup, prop.it.sup.exp)
277 repet.it.subut <- c(repet.it.subut, prop.it.subut)
278 repet.fid <- c(repet.fid, fid)
279 repet.vicio <- c(repet.vicio, vicio)
280 repet.vicio.A <- c(repet.vicio.A, vicio.A)
281 repet.vicio.B <- c(repet.vicio.B, vicio.B)
282 repet.vicio.C <- c(repet.vicio.C, vicio.C)
283 repet.eqm <- c(repet.eqm, eqm)
284 repet.eqm.A <- c(repet.eqm.A, eqm.A)
285 repet.eqm.B <- c(repet.eqm.B, eqm.B)

```

```

286 repet.eqm.C <- c(repet.eqm.C, eqm.C)
287 repet.it.resp.lim <- c(repet.it.resp.lim, prop.it.resp.lim)
288 repet.no.it.resp <- c(repet.no.it.resp, no.med.it.resp)
289 repet.dp.no.it.resp <- c(repet.dp.no.it.resp, dp.no.it.resp)
290
291 } # fim do loop das repeticoes das simulacoes
292
293 resultados <- rbind(repet.overlap, repet.it.exp, repet.it.sup, repet.it.
  subut, repet.fid,
294   repet.vicio, repet.vicio.A, repet.vicio.B, repet.vicio.C, repet.eqm,
  repet.eqm.A,
295   repet.eqm.B, repet.eqm.C, repet.it.resp.lim, repet.no.it.resp, repet.dp.
  no.it.resp)
296
297 vec.media <- round(apply(resultados, 1, mean), digits=5)
298 vec.erro.padrao <- round(apply(resultados, 1, sd), digits=5)
299 sumario.media.ep <- cbind(vec.media, vec.erro.padrao)
300 sumario.media.ep <- matrix(sumario.media.ep, ncol=2)
301 sumario.media.ep

```

Código em R para implementação do método da Máxima Informação Ponderada por  $[(1-p)^\alpha]I(\hat{\theta})$

```

1 #####
2 ### M TODO DE ESCOLHA POR M XIMA INFORMA O PONDERADA POR ((1-p)^alfa)*
  I(theta^ (REPETI ES)
3 ### N MERO ALEAT RIO DE ITENS POR TESTE
4 ### N MERO M XIMO DE ITENS RESPONDIDOS = 60 (n.max)
5 ### N MERO DE ITENS NO BANCO DE ITENS = 400 (n.itens)
6 ### N MERO DE RESPONDENTES = 3000 (n.resp)
7 #####
8
9 #####
10 ### CAT
11 ### - Estimacao da proficiencia via EAP.
12 ### - Escolha do item com maior valor para a informacao ponderada em theta
  estimado.
13 ### - Criterio de parada: PSD.
14 ### - Integracao via retangulos
15 #####
16
17 #####
18 ### Funcao "veross" ###
19 ### calcula o valor da verossimilhanca ###
20 #####
21 veross <- function(par) {
22   prod <- 1
23   for (k in 1:length(resp.aux)) {
24     p.aux <- c.aux[k] + (1-c.aux[k]) / (1+exp(-a.aux[k]*(par-b.aux[k])))
25     prod <- prod*(p.aux^resp.aux[k])*(1-p.aux)^(1- resp.aux[k])
26   }
27   return(prod)

```

```

28 }
29
30 #####
31 ### Funcao "numerador" ###
32 ### numerador a ser integrado em theta no calculo de EAP ###
33 #####
34 numerador <- function(par) {
35   par*veross(par)*dnorm(par,0,1)
36 }
37
38 #####
39 ### Funcao "numerador2" ###
40 ### numerador a ser integrado em theta no calculo de EAP ###
41 #####
42 numerador2 <- function(par) {
43   (par^2)*veross(par)*dnorm(par,0,1)
44 }
45
46 #####
47 ### Funcao "denominador" ###
48 ### denominador a ser integrado em theta no calculo de EAP ###
49 #####
50 denominador <- function(par) {
51   veross(par)*dnorm(par,0,1)
52 }
53
54 #####
55 ### Geracao dos parametros dos itens e do valor real de theta ###
56 #####
57
58 set.seed(23456)
59 n.itens <- 400 # numero de itens no banco de quest es
60 b.par <- rnorm(n.itens,0,1) # valores dos parametros de dificuldade
61 a.par <- runif(n.itens,0.3,2.5) # valores dos parametros de discrimina o
62 c.par <- rep(0.2,n.itens)# valores dos parametros de acerto ao acaso
63
64 n.resp <- 3000 # numero de respondentes
65
66 r.max <- 0.2 # taxa maxima de exposicao desejada
67
68 amp <- 0.001 # amplitude dos intervalos na integracao via retangulos
69 param <- seq(-3,3,amp) # valores de theta utilizados na integracao numerica
70
71 n.max <- 60 # numero maximo de itens respondidos
72
73 #####
74 ## Inicio das repeticoes
75 #####
76
77 n.rep <- 5 # numero de repeticoes na simulacao
78

```

```

79 repet.overlap <- numeric(0) # vetor com taxas de sobreposicao de cada
    simulacao
80 repet.it.exp <- numeric(0) # vetor com proporcao de itens expostos de cada
    simulacao
81 repet.it.sup <- numeric(0) # vetor com proporcao de itens superexpostos de
    cada simulacao
82 repet.it.subut <- numeric(0) # vetor com proporcao de itens subutilizados
    de cada simulacao
83 repet.fid <- numeric(0) # vetor com a fidelidade (correlacao entre theta e
    sua estimativa) de cada simulacao
84 repet.vicio <- numeric(0) # vetor com vicio medio das estimativas de theta
    de cada simulacao
85 repet.vicio.A <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ theta < -1.5 de cada simulacao
86 repet.vicio.B <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ -1.5 < theta < 1.5 de cada simulacao
87 repet.vicio.C <- numeric(0) # vetor com vicio medio das estimativas de
    theta p/ theta > 1.5 de cada simulacao
88 repet.eqm <- numeric(0) # vetor com erro quadratico medio de cada simulacao
89 repet.eqm.A <- numeric(0) # vetor com erro quadratico medio p/ theta < -1.5
    de cada simulacao
90 repet.eqm.B <- numeric(0) # vetor com erro quadratico medio p/ -1.5 < theta
    < 1.5 de cada simulacao
91 repet.eqm.C <- numeric(0) # vetor com erro quadratico medio p/ theta > 1.5
    de cada simulacao
92 repet.it.resp.lim <- numeric(0) # vetor com proporcao de respondentes no
    limite de itens de cada simulacao
93 repet.no.it.resp <- numeric(0) # vetor com numero medio de itens
    respondidos de cada simulacao
94 repet.dp.no.it.resp <- numeric(0) # vetor com desvio-padrao do numero de
    itens respondidos de cada simulacao
95
96 ## Inicio do loop das repeticoes das simulacoes
97 for (m in 1:n.rep) {
98
99   print(m)
100
101   vec.theta <- rnorm(n.resp,0,1) # proficiencia real do respondente
102   vec.theta.est <- numeric(0) # vetor com estimativas de theta
103
104   vec.n.it.resp <- numeric(0) # vetor com numero de itens respondidos
105
106   prop <- rep(0,n.itens) # proporcao de respondentes que responderam cada
    item
107
108   # matriz indicadora de aplicacao dos itens aos respondentes (linha=
    respondente, coluna=item)
109   mat.overlap <- matrix(0,n.resp,n.itens)
110
111   for (j in 1:n.resp) {
112

```

```

113 # print(j)
114
115 theta.est <- 0 # estimativa de theta
116
117 psd.acum <- numeric(0) # vetor com sequencia de estimativas do desvio-
    padrao a posteriori de theta.est
118
119 tol <- 0.1
120 psd <- 1
121
122 #####
123 # banco de itens: colunas da matriz "banco"
124 # 1-indice, 2-a.par, 3-b.par, 4-c.par,
125 # 5-resposta (0=erro, 1=acerto, 9=ainda nao respondido),
126 # 6-ordem de selecao.
127 #####
128
129 banco <- cbind(c(1:n.itens), a.par, b.par, c.par, rep(9, n.itens), rep(0, n.
    itens))
130
131 n.it.resp <- 0 # numero cumulativo de itens respondidos
132
133 while ((n.it.resp <= n.max)&&(psd>tol)) {
134
135     alfa <- runif(1,0.5,8.5) # parametro que altera os pesos da informacao
    de cada item nao escolhido
136
137     banco.aux <- banco[banco[,5]==9,]
138
139     ind.n.esc <- banco.aux[,1]
140     a.n.esc <- banco.aux[,2]
141     b.n.esc <- banco.aux[,3]
142     c.n.esc <- banco.aux[,4]
143
144     vec.prob <- c.n.esc + (1-c.n.esc)/(1+exp(-a.n.esc*(theta.est-b.n.esc)))
145
146     vec.info <- (a.n.esc^2)*((1-vec.prob)/vec.prob)*((vec.prob-c.n.esc)/
    (1-c.n.esc))^2
147 #     vec.info <- vec.info/mean(vec.info)
148     vec.info <- vec.info/sum(vec.info)
149
150 #     const <- 1
151
152     pesos <- (((1-prop[ind.n.esc]/j)^alfa)*vec.info)/sum(((1-prop[ind.n.esc]
    ]/j)^alfa)*vec.info)
153
154     ind <- ind.n.esc[pesos==max(pesos)]
155
156     prop[ind] <- prop[ind] + 1
157
158     n.it.resp <- n.it.resp + 1 # numero de itens respondidos

```

```

159
160 a.esc <- banco[ind,2]
161 b.esc <- banco[ind,3]
162 c.esc <- banco[ind,4]
163
164 p.acerto <- c.esc+(1-c.esc)/(1+exp(-a.esc*(vec.theta[j]-b.esc)))
165 u <- sign(sign(p.acerto-runif(1,0,1))+1) # resposta
166
167 banco[ind,5] <- u
168 banco[ind,6] <- n.it.resp
169
170 #####
171 ## calculo de theta.est via EAP com priori N(0,1)
172 #####
173
174 # os vetores a.aux, b.aux, c.aux e resp.aux contem os valores dos
parametros
175 # dos itens e as respostas para os itens ja' respondidos
176
177 a.aux <- banco[banco[,5]!=9,2]
178 b.aux <- banco[banco[,5]!=9,3]
179 c.aux <- banco[banco[,5]!=9,4]
180
181 resp.aux <- banco[banco[,5]!=9,5]
182
183 theta.est <- integrate(numerador,-3,3)$value/integrate(denominador
,-3,3)$value
184 theta.est2 <- integrate(numerador2,-3,3)$value/integrate(denominador
,-3,3)$value
185
186 psd <- theta.est2-theta.est^2
187 psd.acum <- c(psd.acum,psd)
188
189 }
190
191 vec.theta.est <- c(vec.theta.est,theta.est)
192 vec.n.it.resp <- c(vec.n.it.resp,n.it.resp)
193
194 vec.aux <- banco[,5]
195 vec.aux[vec.aux>1] <- 0
196 mat.overlap[j,] <- vec.aux
197
198 }
199 overlap.theta.dif <- overlap.vec <- numeric(0)
200
201 overlap.soma <- 0
202 tam.am <- 10000 # tamanho da amostra de pares de respondentes para
estimacao da taxa de overlap
203
204 for (k in 1:tam.am) {
205 am.resp <- sample(c(1:n.resp),2,replace=FALSE)

```

```

206 am.overlap <- sum(mat.overlap[am.resp[1],]*mat.overlap[am.resp[2],])
207 overlap.soma <- overlap.soma + am.overlap
208 overlap.vec <- c(overlap.vec, am.overlap)
209 overlap.theta.dif <- c(overlap.theta.dif, abs(vec.theta[am.resp[1]]-vec.
  theta[am.resp[2]]))
210 }
211
212 ## proporcao media de coincidencia de itens entre respondentes
213 overlap.rate <- 100*overlap.soma/(n.itens*tam.am)
214 overlap.rate
215
216 plot(overlap.theta.dif, overlap.vec, xlab=c("Diferen a entre as
  proficiencias"),
217 ylab=c("N mero de itens coincidentes"))
218
219 boxplot(overlap.theta.dif~overlap.vec, ylim=c(0,6), ylab=c("Diferen a entre
  as proficiencias"),
220 xlab=c("N mero de itens coincidentes"), main=c("M todo de M xima
  Informa o Ponderada", "com alfa aleatorio"))
221
222 ## Taxa de exposi o dos itens
223
224 tx.expos <- prop/n.resp
225 tx.expos
226 prop.it.exp <- length(tx.expos[tx.expos>0])/n.itens
227 prop.it.sup.exp <- length(tx.expos[tx.expos>r.max])/n.itens
228 prop.it.subut <- length(tx.expos[tx.expos<0.02])/n.itens
229
230 #mean(tx.expos) # taxa media de exposicao por item
231 #sd(tx.expos) # desvio-padrao da taxa de exposicao dos itens
232
233 ## Fidelidade
234 fid <-cor(vec.theta, vec.theta.est)
235 #fid
236
237 ## V cio
238 vicio <- sum(vec.theta.est-vec.theta)/n.resp
239 #vicio
240
241 ## V cio separado por tr s grupos
242 vicio.A <- sum(vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
  /length(vec.theta[vec.theta < -1.5])
243 vicio.B <- sum(vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta
  ) <= 1.5])/
244 length(vec.theta[abs(vec.theta) <= 1.5])
245 vicio.C <- sum(vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
  /length(vec.theta[vec.theta > -1.5])
246
247 #vicio.A
248 #vicio.B
249 #vicio.C

```

```

250
251 ## Erro quadrático médio (MSE)
252 eqm <- sum((vec.theta.est-vec.theta)^2)/n.resp
253 #eqm
254
255 ## Erro quadrático médio (MSE) separado por três grupos
256 eqm.A <- sum((vec.theta.est[vec.theta < -1.5]-vec.theta[vec.theta < -1.5])
257             ^2)/length(vec.theta[vec.theta < -1.5])
258 eqm.B <- sum((vec.theta.est[abs(vec.theta) <= 1.5]-vec.theta[abs(vec.theta)
259             <= 1.5])^2)/
260             length(vec.theta[abs(vec.theta) <= 1.5])
261 eqm.C <- sum((vec.theta.est[vec.theta > -1.5]-vec.theta[vec.theta > -1.5])
262             ^2)/length(vec.theta[vec.theta > -1.5])
263
264 #eqm.A
265 #eqm.B
266 #eqm.C
267
268 ## Proporção de respondentes que atingiu o número máximo de itens permitido
269 prop.it.resp.lim <- length(vec.n.it.resp[vec.n.it.resp==(n.max+1)])/n.resp
270 #prop.it.resp.lim
271
272 ## Número médio de itens respondidos por respondente
273 no.med.it.resp <- mean(vec.n.it.resp)
274 #no.med.it.resp
275
276 ## Desvio-padrão do número de itens respondidos
277 dp.no.it.resp <- sd(vec.n.it.resp)
278 #dp.no.it.resp
279
280 repet.overlap <- c(repet.overlap, overlap.rate)
281 repet.it.exp <- c(repet.it.exp, prop.it.exp)
282 repet.it.sup <- c(repet.it.sup, prop.it.sup.exp)
283 repet.it.subut <- c(repet.it.subut, prop.it.subut)
284 repet.fid <- c(repet.fid, fid)
285 repet.vicio <- c(repet.vicio, vicio)
286 repet.vicio.A <- c(repet.vicio.A, vicio.A)
287 repet.vicio.B <- c(repet.vicio.B, vicio.B)
288 repet.vicio.C <- c(repet.vicio.C, vicio.C)
289 repet.eqm <- c(repet.eqm, eqm)
290 repet.eqm.A <- c(repet.eqm.A, eqm.A)
291 repet.eqm.B <- c(repet.eqm.B, eqm.B)
292 repet.eqm.C <- c(repet.eqm.C, eqm.C)
293 repet.it.resp.lim <- c(repet.it.resp.lim, prop.it.resp.lim)
294 repet.no.it.resp <- c(repet.no.it.resp, no.med.it.resp)
295 repet.dp.no.it.resp <- c(repet.dp.no.it.resp, dp.no.it.resp)
296
297 } # fim do loop das repetições das simulações
298
299 resultados <- rbind(repet.overlap, repet.it.exp, repet.it.sup, repet.it.
300                    subut, repet.fid,

```

```
297 repet.vicio, repet.vicio.A, repet.vicio.B, repet.vicio.C, repet.eqm,  
    repet.eqm.A,  
298 repet.eqm.B, repet.eqm.C, repet.it.resp.lim, repet.no.it.resp, repet.dp.  
    no.it.resp)  
299  
300 vec.media <- round(apply(resultados,1,mean),digits=5)  
301 vec.erro.padrao <- round(apply(resultados,1,sd),digits=5)  
302 sumario.media.ep <- cbind(vec.media,vec.erro.padrao)  
303 sumario.media.ep <- matrix(sumario.media.ep,ncol=2)  
304 sumario.media.ep
```