

# Redução da Quantidade de Itens para Estimação de Proficiência em Teste Adaptativo Computadorizado com Filtro de Kalman

Rafael M. Albuquerque<sup>1</sup>, Charles C. Cavalcante<sup>1</sup>, Jorge H. S. de Lira<sup>1</sup>

<sup>1</sup>Departamento de Engenharia de Teleinformática – Universidade Federal do Ceará (UFC) – Fortaleza – CE – Brasil

rafael.mor.alb@alu.ufc.br, charles@gtel.ufc.br, jorge.lira@mat.ufc.br

**Abstract.** *Large-scale assessments are important instruments not only in measuring the quality of education, but also in implementing educational policies. However, despite the latest advances, traditional large-scale assessments still face problems, such as the time required for institutions to receive their feedback and the number of items in the test, which is sometimes exhaustive. To cope with this issue, we propose the implementation of a computerized adaptive test based on the item response theory with the application of a Kalman filter to reduce the number of test items. This makes it possible to reduce about 70% of the required number of items in a test, when compared to a traditional test.*

**Resumo.** *Avaliações em larga escala são importantes instrumentos não só na aferição da qualidade da educação, mas também na implementação de políticas educacionais. Contudo, apesar dos últimos avanços, avaliações em larga escala tradicionais ainda enfrentam problemáticas, como o tempo demandado para as instituições receberem suas devolutivas e a quantidade de itens no teste que por vezes é exaustiva. Diante desta limitação, propõe-se a implementação de um teste adaptativo computadorizado baseado na teoria de resposta ao item com a aplicação de um filtro de Kalman para a redução de itens do teste. Com isso, pode-se reduzir cerca de 70% da quantidade necessária de itens de um teste, quando comparado a um teste tradicional.*

## 1. Introdução

Nas últimas décadas, notavelmente, as avaliações educacionais em larga escala tornaram-se ferramentas cada vez mais relevantes no processo de aferição da qualidade do ensino e essenciais na definição de políticas educacionais em todo o mundo. No Brasil, pode-se citar o Sistema de Avaliação da Educação Básica (SAEB) como exemplo de avaliação em escala nacional e o Sistema Permanente de Avaliação da Educação Básica do Ceará (SPAECE) em escala estadual, de acordo com Esteves (2021). Conforme Min e Aryadoust (2021), a aplicação da Teoria de Resposta ao Item (TRI) tem sido predominante como método de estimação dos níveis de conhecimento dos indivíduos que participam dessas avaliações.

Porém, nesse tipo de avaliação, os respondentes podem enfrentar extensos e exaustivos testes por causa da grande quantidade de itens. O Exame Nacional do Ensino Médio (ENEM), por exemplo, possui testes com 45 itens para cada habilidade (traço

latente) a ser mensurada, sendo distribuídos em dois dias de aplicação. Outra fragilidade das avaliações em larga escala tradicionais, de acordo com Spenassato *et al.* (2016), é a existência de itens no teste que não avaliam de forma precisa o nível de conhecimento do indivíduo por não estarem de acordo com suas habilidades e competências.

Nesse sentido, uma solução apropriada, segundo Araújo *et al.* (2020), é a utilização de Teste Adaptativo Computadorizado (TAC), um recurso computacional que permite automatizar o processo de aplicação e correção de avaliações, além de selecionar os itens a serem respondidos conforme o desempenho dos examinandos durante o teste. Dessa forma, a utilização do teste adaptativo é também justificada pela apresentação apenas de itens em consonância com o real nível de conhecimento do indivíduo avaliado.

Com base nisso, este trabalho propõe o desenvolvimento de um TAC baseado na TRI com a otimização do processo de estimação de proficiências por meio do filtro de Kalman (1960). A abordagem permitiu redução da quantidade de itens necessários para uma estimação adequada da proficiência dos indivíduos, tornando os testes menos exaustivos.

O restante desse trabalho está organizado como segue. A Seção 2 descreve o referencial teórico e a metodologia utilizada é explicada na Seção 3. Na Seção 4, são detalhadas as fases de um teste adaptativo computadorizado, e na Seção 5, a filtragem de Kalman é detalhada. Finalmente, a Seção 5 mostra os resultados obtidos e, na Seção 6, são listadas as conclusões do trabalho.

## 2. Referencial Teórico

O trabalho desenvolvido por Spenassato (2016) apresenta as vantagens do uso de Testes Adaptativos Computadorizados em relação ao teste da área de Matemática e suas Tecnologias do Enem 2012, composto por 45 itens, realizando uma simulação com dados reais. Como resultados, definiu-se um comprimento de 33 itens para o TAC, que gerou redução de 26,67% no comprimento do teste e, conseqüentemente, redução no tempo do teste e fadiga dos respondentes.

Jatobá (2019) cria em seu trabalho uma abordagem personalizada no processo de seleção de itens em TAC, denominada ALICAT, que considera o uso de mais de uma Regra de Seleção de Item (ISR) e objetiva reduzir o comprimento de provas dicotômicas (verdadeiro ou falso). Como resultados, obteve-se que, com a utilização da abordagem ALICAT, conseguiu-se reduzir 53,3% da extensão de uma prova de 45 itens.

Em seu trabalho, Souza (2019) utiliza diferentes métodos de estimação, como *Bayes Modal* (BM) e *Expected a Posteriori* (EAP), em diferentes cenários, afim de verificar os melhores resultados para aplicação de TAC em microdados do ENEM. Os resultados indicaram que os testes podem ser encerrados com menos que 45 itens sem grandes prejuízos nos resultados. De modo que, por meio do método de estimação BM, por exemplo, com 20 itens aplicados obtém-se resultados significantes, o que representa redução de até 55,55% no comprimento do teste.

## 3. Teoria de Resposta ao Item: Definição e Modelagem

A TRI é uma metodologia amplamente utilizada em avaliações de larga escala pois permite estimar habilidades e níveis de conhecimento dos indivíduos em uma mesma escala de proficiência. Essa abordagem permite mensurar a habilidade de um indivíduo

de tal forma que o seu valor de proficiência tenha um significado pedagógico quando comparado com a escala de proficiência, conforme Andrade *et al.* (2000).

Além disso, a teoria permite analisar a qualidade dos itens de um teste utilizando suas parametrizações. A parametrização de itens é a fase inicial da criação de uma escala de proficiência sendo executada por meio da realização de pré-testes com uma amostra da população. Com a parametrização dos itens, a TRI permite a criação de um banco de itens que pode ser utilizado para medir a proficiência de novos indivíduos. Neste trabalho, o banco de itens foi simulado seguindo critérios de forma que os itens criados apresentassem parâmetros realistas como proposto por Barrada *et al.* (2010).

Uma outra característica importante que a TRI possui é a comparabilidade entre populações diferentes ou entre indivíduos da mesma população mesmo que tenham sido submetidos a testes diferentes, desde que seus testes possuam itens em comum. Com isso, conforme Couto e Primi (2011), indivíduos que realizam testes contendo itens de um mesmo banco de itens tem suas proficiências geradas na mesma escala de proficiência.

A TRI é um conjunto de modelos matemáticos que mensuram a probabilidade de um indivíduo responder corretamente a um item como função dos parâmetros daquele item e da habilidade do respondente. Essa relação, para Andrade *et al.* (2000), é sempre expressa de maneira que quanto maior a habilidade, maior a probabilidade de acerto no item. Dentre o conjunto de modelos presentes na TRI, o modelo logístico de 3 parâmetros para itens dicotômicos é amplamente difundido, caracterizando cada item de um teste por 3 parâmetros, sendo eles: discriminação, dificuldade e probabilidade de acerto casual.

Nesse modelo, a probabilidade de um indivíduo  $j$  com habilidade  $\theta_j$  obter sucesso no  $i$ -ésimo item é dada por Baker (1992):

$$P(U_{ij} = 1|\theta_j) = c_i + \frac{1 - c_i}{1 + \exp\{-a_i(\theta_j - b_i)\}} \quad (1)$$

Com  $i = 1, 2, \dots, I$ , e  $j = 1, 2, \dots, n$ , em que:

- $U_{ij}$  é uma variável dicotômica que assume os valores 1, quando o indivíduo  $j$  responde corretamente o item  $i$ , ou 0 quando o indivíduo  $j$  não responde corretamente ao item  $i$ ;
- $\theta_j$  representa a habilidade (traço latente) do  $j$ -ésimo indivíduo;
- $P(U_{ij} = 1|\theta_j)$  é a probabilidade de um indivíduo  $j$  com habilidade  $\theta_j$  responder corretamente o item  $i$  e é chamada de Função de Resposta do Item;
- $b_i$  é o parâmetro de dificuldade (ou de posição) do item  $i$ , medido na mesma escala da habilidade;
- $a_i$  é o parâmetro de discriminação (ou de inclinação) do item  $i$ , com valor proporcional à inclinação da Curva Característica do Item (CCI) no ponto  $b_i$ ;
- $c_i$  é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item  $i$  (muitas vezes referido como a probabilidade de acerto casual).

A representação gráfica do modelo logístico é chamada de Curva Característica do Item (CCI). Na Figura 1, é exibida uma CCI do modelo logístico de 3 parâmetros.

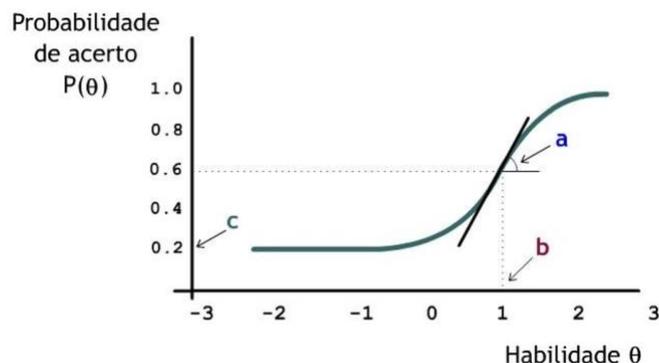


Figura 1. Exemplo de uma curva característica de um item.

#### 4. Teste Adaptativo Computadorizado

Com os avanços da tecnologia das últimas décadas, principalmente na área de computação, novas formas de avaliação têm ganhado destaque, como os testes adaptativos computadorizados. Esse tipo de abordagem visa individualizar cada teste de forma que sua acurácia seja maximizada para o nível de conhecimento do indivíduo.

Para tornar isso possível, um TAC é constituído pelas seguintes componentes:

- Um banco de itens calibrado pela TRI;
- Método de inicialização do teste;
- Critério de seleção de itens;
- Método de estimação de proficiência;
- Critério de parada do teste.

O fluxograma da Figura 2 ilustra como funciona um TAC.

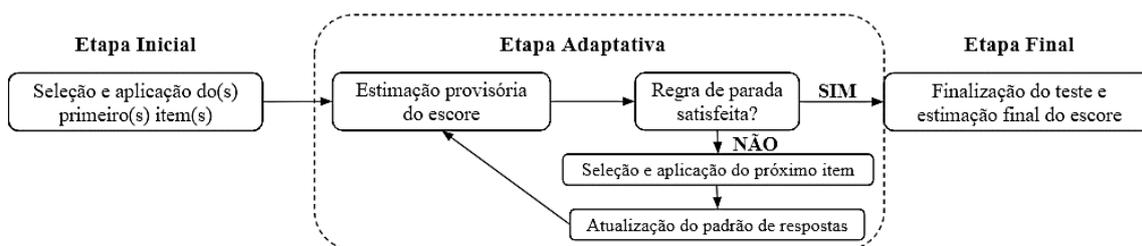


Figura 2. Fluxograma de um TAC.

As fases desse fluxograma são mais detalhadas a seguir.

##### 4.1. Início do Teste

O início de todo teste adaptativo consiste na atribuição de um valor inicial de proficiência para o indivíduo a ser avaliado. Essa definição depende do público a ser avaliado, que pode apresentar proficiências semelhantes ou díspares e, portanto, a escolha fica a critério do aplicador, que interpretará as necessidades avaliativas. Nesse caso, por tratar-se de uma aplicação geral para testes de proficiência, onde não há, *a priori*, estipulação do grau avaliativo, adota-se o método de inicialização do tipo fixo, que consiste na atribuição do mesmo valor inicial de proficiência para todos os indivíduos a serem avaliados.

Sendo assim, parte-se do princípio que todos os indivíduos iniciam o teste no mesmo ponto da escala de proficiência adotada, o qual é geralmente escolhido como o valor médio da escala de proficiência. Em outras palavras, em uma escala de 0 a 500, por exemplo, o algoritmo considera que, no ponto de partida, todos os indivíduos possuem o mesmo valor inicial 250 na escala de proficiência.

#### 4.2. Critério de Seleção

A cada passo do processo iterativo de estimação da proficiência do indivíduo, um novo item do banco de itens deve ser exibido para ele. Esse novo item deve ser selecionado de acordo com a sua atual proficiência estimada, ou seja, se ele tiver acertado o item anterior, é esperado que o algoritmo selecione um item com dificuldade superior para ser exibido na próxima iteração. De forma análoga, quando o indivíduo responde incorretamente um item, o algoritmo deve selecionar um item de dificuldade inferior à dificuldade do item anterior para a próxima iteração.

Dentre os métodos de seleção de itens existentes, optou-se pela utilização do método da máxima informação, em conformidade com Ayala (2009). Nesse método, o item é selecionado de acordo com o seu valor de informação dada a proficiência atual do indivíduo na iteração. Assim, o item a ser selecionado será aquele que possuir o maior valor de informação para a proficiência atual estimada do indivíduo dentre todos os itens do banco que ainda não foram exibidos para ele. O cálculo da informação de um item é dado, seguindo Ayala (2009), por:

$$I(\theta) = \frac{a^2 [P(\theta) - c]^2 [1 - P(\theta)]^2}{(1 - c)^2 [1 - P(\theta)] P(\theta)}, \quad (2)$$

em que  $I(\theta)$  é a informação do item dada a proficiência  $\theta$  do indivíduo,  $a$  é o parâmetro de discriminação do item,  $c$  é o parâmetro de chute do item e  $P(\theta)$  é a probabilidade de um indivíduo de proficiência  $\theta$  acertar o item e é dada pela expressão em (1).

A escolha deste método foi motivada pelo fato que ele seleciona o item que mais se adequa ao nível de proficiência atual do indivíduo e, por consequência, aumenta a precisão da estimação da sua proficiência.

#### 4.3. Método de Estimação da Proficiência

Nas condições da teoria de resposta ao item, há alguma probabilidade de que um indivíduo tenha sucesso ou falhe em qualquer item, não importa o quão fácil ou difícil seja o item. Mesmo o indivíduo mais capaz pode falhar em todos os itens.

A proficiência estimada para um indivíduo é aquela para a qual o padrão de respostas observado é mais provável ou aquela para a qual o padrão de respostas se ajusta melhor ao modelo. Pode-se, entretanto, ter uma ideia aproximada sobre a medida de habilidade de um indivíduo antes da coleta de dados atual e ser desejado incorporar essa ideia na medida recém-estimada. Em seguida, combina-se as calibrações dos itens, a ideia aproximada anterior e as respostas observadas para obter-se uma medida de proficiência aprimorada, dita *a posteriori*, conforme Mislevy e Stocking (1989).

O método descrito é conhecido como *Expected a Posteriori* (EAP), e consiste em um método bayesiano para a estimação da habilidade de um indivíduo baseada em um conhecimento prévio de sua habilidade, dado pela seguinte expressão:

$$EAP(\theta|\mathbf{u}) = \frac{\int_{-\infty}^{\infty} \theta L(\mathbf{u}|\theta) f(\theta) d\theta}{\int_{-\infty}^{\infty} L(\mathbf{u}|\theta) f(\theta) d\theta}, \quad (3)$$

em que  $EAP(\theta|\mathbf{u})$  é a proficiência estimada,  $L(\mathbf{u}|\theta)$  é a função de máxima verossimilhança e  $f(\theta)$  é a função de probabilidade *a priori*.

A proficiência do indivíduo deve ser atualizada a cada nova iteração. Dessa forma, para cada nova resposta, recalcula-se a proficiência com base na nova resposta do indivíduo concatenada ao histórico de respostas anteriores e nos coeficientes dos itens respondidos.

#### 4.4. Critério de Parada do Teste

Sem a definição de um critério de parada preestabelecido, o teste adaptativo repete-se de forma indefinida. Sendo assim, é necessária a definição de um limite para que o teste possa ser encerrado. Os principais critérios utilizados são a quantidade máxima de itens a serem exibidos ou um valor mínimo para o erro padrão de estimação.

No critério da quantidade máxima de itens, a cada iteração (resposta do indivíduo), o algoritmo verifica se a quantidade máxima de itens a serem exibidos foi atingida. Caso tenha sido atingida, o algoritmo encerra o teste e a proficiência do indivíduo é a sua última proficiência estimada. Em caso contrário, o algoritmo continua o seu fluxo adaptativo selecionando outro item para o indivíduo.

E como o próprio nome sugere, no critério do erro padrão de estimação mínimo, a cada iteração é analisado se o erro mínimo configurado foi atingido. Se o erro mínimo for atingido, o algoritmo encerra o teste e a proficiência do indivíduo é a sua última proficiência estimada. Se o erro não for atingido, o algoritmo continua o fluxo adaptativo da próxima iteração.

Neste trabalho, foi selecionado o critério de erro padrão de estimação mínimo igual a 0,3 definido segundo Stafford *et al.* (2018), pois tem-se como objetivo reduzir ao máximo a quantidade de itens apresentados ao indivíduo mantendo-se o erro padrão de estimação dentro do limite aceitável.

### 5. Filtro de Kalman

O filtro de Kalman, idealizado por Kalman (1960), propõe uma solução recursiva para o problema de filtragem linear de dados discretos e é um dos algoritmos de estimativa mais importantes e comuns. Este método produz estimativas de variáveis ocultas com base em medições imprecisas e incertas. Além disso, o algoritmo de estimação fornece uma previsão do estado futuro do sistema, com base nas estimativas anteriores.

Atualmente, o filtro de Kalman é usado em alvos de rastreamento (radar), sistemas de localização e navegação, sistemas de controle, computação gráfica e muito mais. Neste trabalho, propõe-se a sua utilização para otimizar a estimação de proficiências, visto que, em um teste adaptativo computadorizado, o processo de estimação tem altos valores de incerteza, principalmente nas iterações iniciais da estimação. Com isso, espera-se que seja possível convergir para o valor real de proficiência do indivíduo com quantidade menor de iterações, ou seja, com a menor quantidade de itens possível.

### 5.1. O Algoritmo

O filtro de Kalman opera em um processo adaptativo, no qual cada iteração contém uma fase de predição e uma de correção. A Figura 3 ilustra esse processo adaptativo.

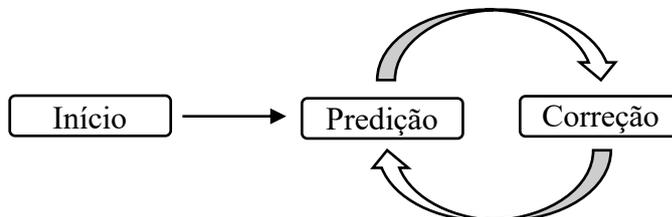


Figura 3. Processo adaptativo do Filtro de Kalman.

Depois de iniciado, o filtro deve prever o estado do sistema na próxima iteração e fornecer a incerteza dessa previsão. Assim que uma medição é recebida, o algoritmo de estimação deve corrigir (ou atualizar) a previsão e a incerteza do estado atual. Dessa forma, o filtro prevê e corrige os próximos estados, conforme Kalman (1960).

A fase de predição consiste nos passos de previsão do novo estado e de cálculo da incerteza dessa previsão. Cada passo é representado, respectivamente, pelas Equações (4) e (5), dadas a seguir:

$$\hat{x}_{n+1,n} = F\hat{x}_{n,n} + Gu_n \quad (4)$$

$$P_{n+1,n} = FP_{n,n}F^T + Q \quad (5)$$

em que  $x$  é o vetor de estados,  $F$  é a matriz de transição de estados,  $G$  é a matriz de controle,  $u$  é o vetor de variáveis de entrada,  $P$  é a matriz de estimativa de incerteza e  $Q$  é a matriz de ruído do processo.

Na fase de correção acontece a correção dos valores do estado e da incerteza das medidas anteriores baseado na nova medida recebida. Para isso, primeiramente é calculado o ganho de Kalman da iteração e a partir dele o estado e a incerteza são corrigidos. As equações do ganho de Kalman, correção do estado e correção da incerteza são, respectivamente, dadas por (6), (7) e (8), representadas a seguir:

$$K_n = P_{n,n-1}H^T(HP_{n,n-1}H^T + R_n)^{-1}, \quad (6)$$

$$\hat{x}_{n,n} = \hat{x}_{n,n-1} + K_n(z_n - H\hat{x}_{n,n-1}), \quad (7)$$

$$P_{n,n} = (I - K_nH)P_{n,n-1}(I - K_nH)^T + K_nR_nK_n^T, \quad (8)$$

em que  $K$  é o ganho de Kalman,  $H$  é a matriz de observação,  $R$  é a medida de incerteza,  $z$  é o vetor de saída e  $I$  é a matriz identidade.

### 6. Análise dos Resultados

Posteriormente à implementação das metodologias acima especificadas, foram realizadas simulações computacionais para avaliar a robustez do algoritmo. O teste adaptativo foi criado seguindo os seguintes requisitos:

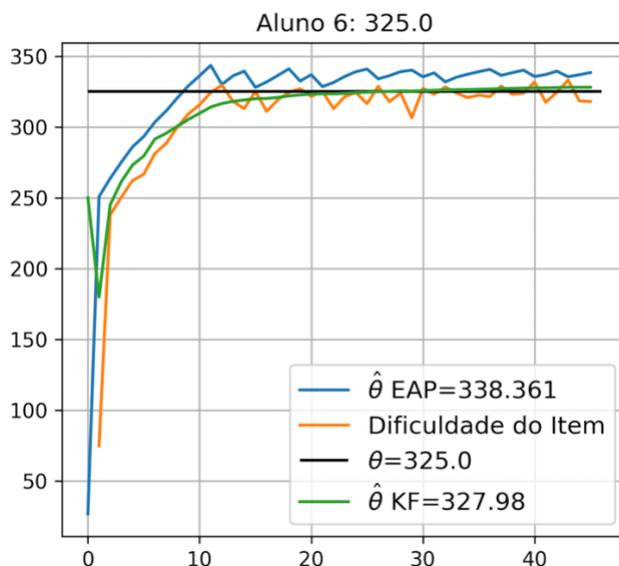
- Banco de itens - 1000 itens simulados usando o modelo de 3 parâmetros;
- Inicialização do teste - inicializador fixo em 250 (escala de 0 a 500);
- Método de seleção de itens - máxima informação;

- Método de estimação de proficiência - EAP com filtro de Kalman;
- Critério de parada - erro padrão de estimação mínimo de 0.3.

Por causa da ausência de um banco de itens reais contendo uma quantidade significativa de itens que tornasse possível a realização do TAC, foi necessário realizar a simulação de um banco contendo 1000 itens. Os parâmetros de discriminação, dificuldade e chute dos itens são gerados seguindo, respectivamente, as distribuições de probabilidade:  $N(1.2, 0.25)$ ,  $N(0, 1)$  e  $N(0.25, 0.02)$ , conforme Barrada *et.al* (2010). O filtro de Kalman é utilizado para otimizar a estimação da proficiência.

### 6.1. Resultados da Simulação

Com o banco de itens obtido, o passo seguinte foi a realização da simulação do teste adaptativo de um indivíduo escolhido ao acaso (utiliza-se aqui o Aluno 5). Inicialmente, o algoritmo foi configurado para encerrar o teste pelo critério da quantidade máxima de itens, sendo atribuído o valor máximo igual a 45 (propositalmente a mesma quantidade de itens por teste do ENEM). Essa modificação inicial teve como objetivo a análise da convergência da proficiência comparando-se o método EAP e o método utilizando o filtro de Kalman. O resultado desse experimento é mostrado na Figura 4.



**Figura 4. Processo adaptativo do Filtro de Kalman.**

No gráfico da Figura 4, tem-se na cor preta a proficiência do indivíduo simulado, em azul é a proficiência estimada pelo método EAP, em laranja tem-se a dificuldade dos itens que foram exibidos ao indivíduo e, por fim, em verde tem-se a proficiência do indivíduo estimada utilizando o filtro de Kalman. No eixo das abscissas tem-se a quantidade de itens exibida ao indivíduo e no eixo das ordenadas tem-se os valores de proficiência e dificuldade dos itens (ambos na mesma escala). Vale ressaltar que, para facilitar a interpretação, os valores de proficiência e dificuldade dos itens foram transformados da escala de média 0 e desvio padrão 1 para a escala de média 250 e desvio padrão 50, obtendo-se assim um valor entre 0 e 500.

Pela Figura 4, pode-se observar que com cerca de 45 itens, ambos os métodos se aproximam do valor da proficiência do indivíduo simulado, que é 325. No entanto, o método otimizado pelo Filtro de Kalman (em verde), visivelmente, converge para o valor

da proficiência com menos iterações (cerca de 13), atingindo o valor de 327,98, do que o método EAP (em azul), que atinge o valor de 338,36. Além disso, o Filtro de Kalman possui uma curva consideravelmente mais suave, implicando em uma variância menor.

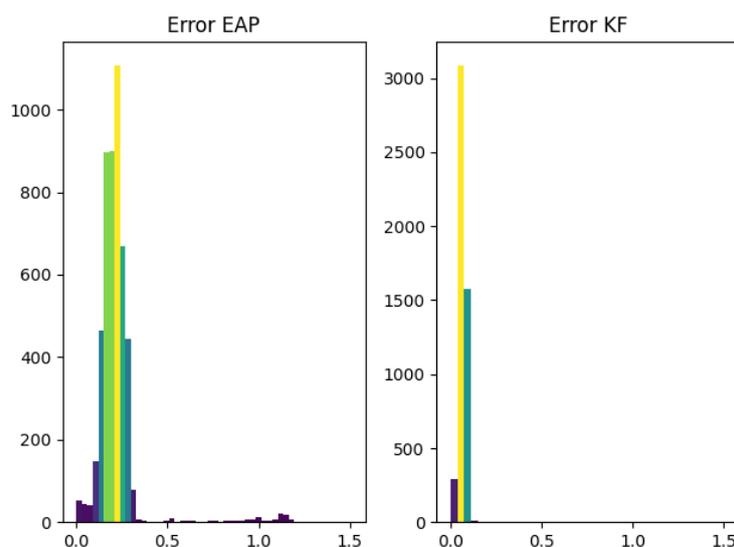
### 6.2. Simulação de Monte Carlo

Partindo dessas observações iniciais, como critério de robustez para os resultados do algoritmo, foi realizada uma simulação de Monte Carlo, na qual foram simulados 5000 indivíduos respondendo testes. Os resultados são exibidos na Tabela 1 e Figura 5.

**Tabela 1. Comparativo entre média e desvio padrão do erro de estimação entre os métodos EAP e Filtro de Kalman.**

Método	Média	Desvio Padrão
<i>Expected a Posteriori</i>	0,222606	0,138206
Filtro de Kalman	0,070162	0,049416

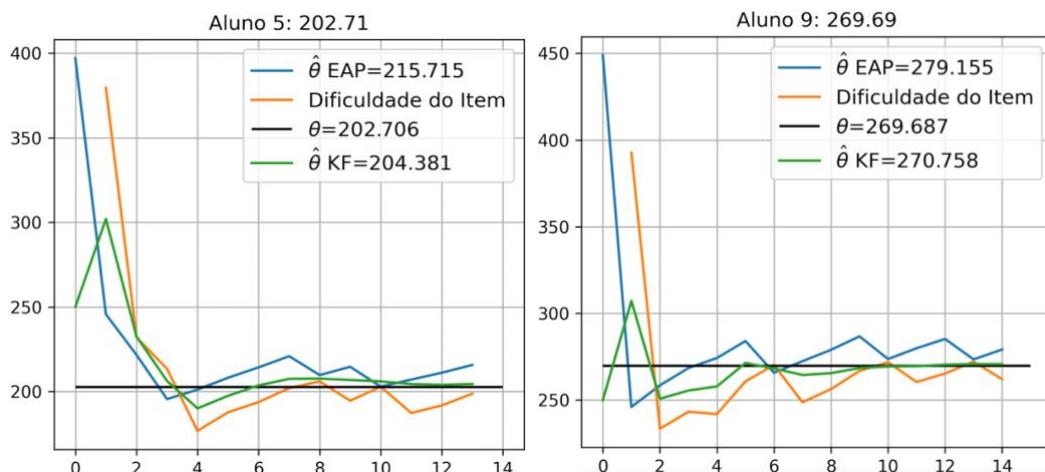
Observando a Tabela 1 e a Figura 5, nota-se que, com o filtro, as estimações possuem menores valores de erro médio e de desvio padrão do que o método *Expected a Posteriori*. Logo, a utilização do filtro de Kalman produz uma estimação da proficiência do indivíduo que converge com menos iterações e mais precisamente para o valor real do que utilizando o método EAP.



**Figura 5. Histogramas de erro de estimação do método *Expected a Posteriori* (à esquerda) e do método do Filtro de Kalman (à direita).**

### 6.3. Redução da Quantidade de Itens

A partir dos resultados da simulação de Monte Carlo, foi aplicado o critério de parada do erro padrão de estimação mínimo igual a 0.3, por se tratar de uma margem de erro aceitável segundo Stafford *et al.* (2018). Dessa forma, novas simulações foram geradas e observou-se que o algoritmo atinge o erro mínimo de parada com cerca de 13 itens apresentados. A critério de ilustração, na Figura 6, são mostrados os resultados de duas amostras de simulação escolhidos ao acaso, para ilustrar situações onde a proficiência inicial (meio da escala) está acima e abaixo das proficiências reais (Alunos 5 e 9).



**Figura 6. Resultados da simulação de dois testes adaptativos.**

Pode-se também observar pela Figura 6, que o método de estimação EAP sem o filtro de Kalman, resulta na proficiência do indivíduo 5 como 215,71, enquanto com a utilização do filtro de Kalman, a proficiência estimada do mesmo indivíduo é de 204,38. Logo, obteve-se uma estimação mais precisa com a utilização do filtro de Kalman, visto que a proficiência do indivíduo simulado é 202,71.

Um resultado similar foi obtido com o indivíduo 9, que tem proficiência igual a 269,69. Pelo método EAP, sem filtro de Kalman, obteve-se o valor 279,15. Com o método que utiliza filtro de Kalman, obteve-se 270,76.

## 7. Conclusões

Apesar de consolidados avanços na área de educometria, mais especificamente no que se refere à medição e comparabilidade em avaliações de larga escala com a teoria de resposta ao item, os métodos de avaliação tradicionais ainda apresentam alguns desafios com relação ao tempo de apropriação dos resultados e a quantidade exaustiva de itens dos testes.

Testes adaptativos computadorizados se mostram como uma boa alternativa para contornar esses problemas, visto que, por sua natureza tecnológica, é possível entregar os resultados de forma imediata à finalização de cada teste. Além disso, pode-se reduzir custos com correção, impressão dos testes, distribuição, etc.

Como proposta para a redução da quantidade de itens em um teste adaptativo, neste trabalho foi aplicado um filtro de Kalman sobre as estimativas de proficiência do método EAP. Essa abordagem permitiu estimar a proficiência de indivíduos com uma redução de cerca de 70% da quantidade de itens no teste, se comparado a um teste tradicional.

Portanto, diante da expressiva redução da quantidade de itens para estimação de proficiência, a utilização do filtro de Kalman mostra-se uma promissora alternativa diante dos tradicionais testes longos e exaustivos, confirmando Araújo *et al.* (2020). O algoritmo, inclusive, sobressai-se quando comparado a outros trabalhos similares que utilizaram outras metodologias em busca do mesmo objetivo, como o de Spenassato *et al.* (2016), que obteve redução de 26,6% dos itens, Jatobá (2019) que conseguiu a redução de 53,3% e Souza (2019), que reduziu até 55,55% dos itens.

## Referências

- Andrade, D. F., Tavares, H. R., & Valle, R. C. (2000). *Teoria da Resposta ao Item. Conceitos e Aplicações. Associação Brasileira de Estatística*. São Paulo.
- Araújo, A. L. S. O., J. S. Santos, M. R. A. Melo, W. L. Andrade, D. D. S. Guerreiro, & J. C. A. de Figueiredo (2020). *Teoria de Resposta ao Item (2 ed.)*, Capítulo 8. Porto Alegre: SBC.
- Ayala, R. J. (2009). *The Theory & Practice of Item Response Theory. Guilford Press*. New York.
- Baker, F. B. (1992). *Item Response Theory: Parameter Estimation Techniques. Marcel Dekker, Inc.* New York.
- Barrada, J. R., Olea, J., Ponsoda, V., & Abad, F. J. (2010). A Method for the Comparison of Item Selection Rules in Computerized Adaptive Testing. *Applied Psychological Measurement*, 34(6), 438–452.
- Couto, G.; Primi, R. (2011). Teoria de resposta ao item (TRI): Conceitos elementares dos modelos para itens dicotômicos. *Boletim de Psicologia*, São Paulo, v. 61, n. 134, p. 1 – 15.
- Esteves, T (2021). Avaliação em larga escala. In. Rothen, J. C.; Esteves, T.; Oliveira, I. S. *Glossário Brasileiro de avaliação educacional*.
- Fonseca, J. C. G. (2019). *Uma Aplicação do Teste Adaptativo Computadorizado via Filtro de Kalman Não-Linear*. Mestrado em estatística, Universidade de Brasília.
- Jatobá, V. M. G. (2019). *Uma abordagem personalizada no processo de seleção de itens em Testes Adaptativos Computadorizados*. 60 p. Dissertação (Mestrado em Sistemas de Informação) — Universidade de São Paulo.
- Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME - Journal of Basic Engineering*, 82, 35-45.
- Min, S. & V. Aryadoust (2021, March). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Studies in Educational Evaluation* 68.
- Mislevy, R. J. & Stocking, M. L. (1989). A consumer's guide to LOGIST & BILOG. *Applied Psychological Measurement*, 13, 57-75.
- Souza, W. F. de. (2019). *Testes Adaptativos Computadorizados Aplicados às Provas do ENEM*. 65 p. Monografia (Estatística) — Universidade Federal de Ouro Preto.
- Spenassato, D., Trierweiler, A. C., De Andrade, D. F., & Bornia, A. C. (2016). Testes Adaptativos Computadorizados Aplicados em Avaliações Educacionais. *Revista Brasileira de Informática na Educação*, 24 (2).
- Stafford, R. E., Runyon, C. R., Casabianca J. M., & Dodd, B. G. (2018). Comparing Computer Adaptive Testing Stopping Rules Under the Generalized Partial-Credit Model. *Psychonomic Society. Behavior Research Methods*, 51, 1305–1320.