

CURVA DE CRESCIMENTO EM MATEMÁTICA VIA TEORIA DA RESPOSTA AO ITEM

HELITON RIBEIRO TAVARES
ADRIANA MORAES DE CARVALHO
WALTER LANA LEITE

RESUMO

Neste trabalho são propostos modelos para acompanhar a evolução do desempenho educacional médio em Matemática de um grupo de indivíduos avaliados ao longo do tempo no contexto da Teoria da Resposta ao Item, possibilitando a estimação de habilidades médias em períodos não avaliados. Foram avaliadas as curvas de crescimento linear, quadrática, logística e Gompertz. A característica longitudinal induz uma dependência entre as habilidades nos vários tempos de avaliação, sendo propostas algumas estruturas de covariância para modelar essa dependência. Os parâmetros dos itens foram considerados conhecidos de uma calibração prévia. Realizou-se uma aplicação em dados provenientes do Plano de Desenvolvimento da Escola (PDE-Escola/MEC). Verificou-se que a curva de crescimento logística, associada à estrutura de correlação AR(1), ajustou-se melhor aos dados, com uma altíssima correlação entre as habilidades estimadas, e que o maior ganho em habilidade ocorre até o final do 6º ano escolar.

PALAVRAS-CHAVE TEORIA DA RESPOSTA AO ITEM • CURVAS DE CRESCIMENTO • MATEMÁTICA • AVALIAÇÃO DA EDUCAÇÃO.

CURVA DE CRECIMIENTO PARA MATEMÁTICAS VÍA TEORÍA DE RESPUESTA AL ÍTEM

RESUMEN

En este trabajo se proponen modelos para acompañar la evolución del desempeño educacional medio en Matemáticas de un grupo de individuos evaluados a lo largo del tiempo usando la Teoría de la Respuesta al Ítem, a fin de posibilitar la estimación de habilidades medias en periodos no evaluados. Se evaluaron las curvas de crecimiento lineal, cuadrática, logística y Gompertz. La característica longitudinal induce una dependencia entre las habilidades en los varios tiempos de evaluación, proponiéndose algunas estructuras de covariancia para modelar dicha dependencia. Los parámetros de los ítems fueron considerados conocidos de una calibración previa. Se realizó una aplicación en datos provenientes del Plano de Desenvolvimento da Escola (PDE-Escola/MEC) [Plan de Desarrollo de la Escuela]. Se verificó que la curva de crecimiento logística, asociado a la estructura de correlación AR(1), se ajustó mejor a los datos, con una altísima correlación entre las habilidades estimadas, y que la mayor ganancia en habilidad ocurre hasta fines del 6º año escolar.

PALABRAS CLAVE TEORÍA DE LA RESPUESTA AL ÍTEM • CURVAS DE CRECIMIENTO • MATEMÁTICAS • EVALUACIÓN DE LA EDUCACIÓN.

GROWTH CURVE FOR MATHEMATICS VIA ITEM RESPONSE THEORY

ABSTRACT

This study proposes models to follow the evolution of average educational performance of a group of individuals in Mathematics. They were evaluated over time using Item Response Theory, enabling estimation of average skills in periods not evaluated and improving the interpretation of scales. The linear, quadratic, logistic and Gompertz growth curves were evaluated. The longitudinal feature leads to interdependence among skills at the various evaluation times, with the proposal of some covariance structures to model this dependence. The parameters of the items were known from a prior calibration. Real data from the Plano de Desenvolvimento da Escola (PDE-Escola/MEC) [School Development Plan] of the Ministry of Education were used. It was found that the logistic growth curve, associated with the correlation structure AR(1), fit better to data, with a very high correlation between the estimated skills and that the greatest gain in skill occurs by the end of the 6th school year.

KEYWORDS ITEM RESPONSE THEORY • GROWTH CURVES • MATHEMATICS • EVALUATION OF EDUCATION.

INTRODUÇÃO

A avaliação educacional é um ponto importante no ensino como política de prestação de contas ou como planejamento (RAUDENBUSH; FOTIU; CHEONG, 1998), sendo cada vez mais discutida por profissionais que buscam receber mais informações dos alunos do que simplesmente o número de itens (questões) respondidos corretamente em um teste. A Teoria da Resposta ao Item (TRI) entrou nesse contexto com uma abordagem inovadora, propondo modelos que relacionam os traços latentes (habilidades, proficiências ou desempenho) do aluno com os itens apresentados a ele em um teste, modelando assim a relação entre a probabilidade de um aluno responder corretamente a determinado item e sua habilidade na área de conhecimento avaliada (ANDRADE; TAVARES; VALLE, 2000). Sobretudo, a principal contribuição da TRI foi garantir que os resultados de avaliações realizadas em momentos diferentes possam ser diretamente comparados, possibilitando o avanço em outras áreas, tais como a mensuração mais precisa do crescimento na habilidade ao longo do tempo.

A TRI vem tornando-se uma componente essencial para o pesquisador em ciências do comportamento. Ela fornece um poderoso meio para estudar respostas individuais a uma variedade de estímulos e tem sido estendida e desenvolvida para cobrir campos essenciais do conhecimento científico (VAN DER LINDEN; HAMBLETON, 2013). Tornou-se a técnica predominante no contexto de avaliações educacionais no Brasil e no mundo, com algumas iniciativas na construção de indicadores gerais em outras áreas (MOREIRA JR., 2010). A TRI foi usada pela primeira vez em 1995, na análise de dados do Sistema Nacional de Ensino Básico (Saeb). A partir dos resultados obtidos no Saeb, outras avaliações em larga escala também foram planejadas e implementadas de modo a serem analisadas por meio da TRI, das quais destacam-se: Avaliação Nacional do Rendimento Escolar (Prova Brasil); Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (Saresp); e, com maior ênfase atualmente, Exame Nacional do Ensino Médio (Enem).

Como antecipado, uma das principais vantagens da TRI é possibilitar a comparabilidade de seus resultados em distintos testes, edições ou níveis escolares, desde que devidamente planejados para essa finalidade. Após estimados os parâmetros dos itens (etapa conhecida como calibração), as habilidades dos respondentes também são estimadas na mesma escala, ou seja, são todas diretamente comparáveis, mesmo que os indivíduos respondam a conjuntos de itens diferentes (BAKER; KIM, 2004).

No Saresp, por exemplo, que é aplicado anualmente, são selecionados 13 blocos de oito itens cada, totalizando 104 itens cobrindo todo ou quase todo o conteúdo de determinada disciplina para cada nível escolar. A partir desses 13 blocos, são elaborados 26 cadernos distintos de prova em esquema de Blocos Incompletos Balanceados (BIB), cada um com três blocos, ou seja, 24 itens. Dessa forma, cada aluno só responde a 24 itens, mas é avaliado um total de 104, permitindo a interpretação mais precisa dos resultados do sistema escolar. Parte dos itens de um ano (chamados itens de ligação) é colocada nos blocos dos anos ou níveis seguintes, possibilitando que parâmetros de provas aplicadas em anos

e/ou níveis distintos estejam todos na mesma escala, de forma que todos os itens (isto é, seus parâmetros) estejam em uma escala única, totalmente comparável, constituindo, assim, as habilidades médias de cada ano, em cada nível e disciplina (ANDRADE; TAVARES; VALLE, 2000).

Estudos longitudinais têm se apresentado como a fronteira da pesquisa científica e aplicada em muitas áreas, incluindo o campo educacional em que o interesse reside em acompanhar a eficácia e equidade escolar. No entanto, alguns autores ressaltam a impossibilidade de fazer inferências confiáveis sobre a eficácia escolar com base em uma única medida do desempenho (GOLDSTEIN et al., 2000), apontando para a necessidade de estudos longitudinais. Nesse sentido, já há iniciativas, como o Plano de Desenvolvimento da Escola (PDE-Escola), do Ministério da Educação, conduzido pelo Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (Inep) e realizado de 1999 a 2003, com um *baseline* no início de 1999 e mais cinco testes ao final de cada ano, totalizando seis testes (CARNOY et al., 2008). Soma-se a esse o Geres (Estudo Longitudinal da Geração Escolar) (FRANCO; BROOKE; ALVES, 2008; BROOKE et al., 2014), com duas aplicações em 2005 (março e novembro) e continuado anualmente até 2008, no mês de novembro. Tanto o PDE como o Geres avaliaram Leitura e Matemática, além da aplicação de questionários socioeconômicos. Também há iniciativas na educação superior, particularmente na área de medicina, com o Teste do Progresso (SAKAI; FERREIRA FILHO; MATSUO, 2011).

Em termos de planejamento, é fundamental para os gestores avaliar como as habilidades médias de suas unidades estão evoluindo ao longo do tempo ou dos níveis educacionais, possibilitando produzir previsões a médio e longo prazos, bem como melhorando a análise dos conteúdos absorvidos nos vários níveis escolares (SMITH; YEN, 2006; SCHAFER et al., 2012). O presente artigo trata exatamente deste problema, ainda não considerado na literatura no contexto da Teoria da Resposta ao Item, de forma que sua contribuição será de grande relevância não só do ponto de vista teórico, mas sobretudo na correção de rumos adotados pelos gestores, visando ao alcance de possíveis metas previamente definidas.

Este trabalho está organizado em três seções, além dessa introdução. A seguir, é feita a formalização do modelo estatístico, finalizando com breves comentários sobre a estimação dos parâmetros do modelo. Posteriormente, apresenta-se uma aplicação nos dados do projeto PDE e, por fim, são tecidas algumas considerações sobre o trabalho.

METODOLOGIA

O método consiste, prioritariamente, em modelar as médias das distribuições latentes sem precisar obtê-las por meio das estimativas das habilidades, mas sim a partir de funções que servirão para estimar as médias em um período contínuo do tempo. De forma geral, a seguir apresentam-se os requisitos teóricos necessários à modelagem, a saber: a distribuição multivariada da habilidade; o modelo de resposta ao item; as funções que modelarão as médias; e as estruturas de covariância entre as habilidades dos respondentes nos diferentes instantes de avaliação. Será precisa alguma notação mínima para uma maior compreensão da modelagem, mas sem os desenvolvimentos necessários para chegarmos ao processo de estimação dos parâmetros, o que não é o escopo deste artigo.

A DISTRIBUIÇÃO LATENTE DO VETOR DE HABILIDADES

Considera-se a situação em que um grupo de N indivíduos é avaliado ao longo do tempo, em T testes, e que as habilidades individuais no tempo $t = 1, \dots, T$ provêm de uma variável θ_t com média μ_t e variância σ_t^2 . Dada a característica longitudinal, espera-se uma estrutura de dependência entre as habilidades nos T testes, uma vez que os mesmos indivíduos são submetidos a tais testes. Supõe-se que o vetor de habilidades dos respondentes nos T testes, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_T)'$, tenha distribuição contínua multivariada com vetor de parâmetros $\boldsymbol{\eta}$ de componentes finitas e função densidade de probabilidade $g(\boldsymbol{\theta}|\boldsymbol{\eta})$. No caso Normal T -variado tem-se que $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, onde $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_T)$ é o vetor de médias e $\boldsymbol{\Sigma}$ é a matriz $T \times T$ de covariâncias com elementos σ_{ts} .

A FUNÇÃO DE RESPOSTA DO ITEM

Na maioria dos casos, particularmente na área educacional, o modelo adotado para a função de resposta do item deve satisfazer uma condição básica: quanto maior a habilidade, maior será a probabilidade de acerto ao item considerado. Cada item tem características próprias que devem ser representadas por seus parâmetros. Frequentemente adotam-se modelos unidimensionais, ou seja, o item mede o desempenho específico em uma única habilidade (Matemática, por exemplo), sendo a mesma habilidade em todos os itens que compõem o teste. Isso significa que a habilidade de um indivíduo não se altera durante um teste t , mas certamente poderá ser diferente nos T testes.

Sejam I_t o número de itens no teste t e I o número total de itens nos T testes, vamos considerar que todos os itens sejam de múltipla escolha, dicotômicos ou dicotomizados do tipo certo/errado. A resposta do indivíduo j ao item i do teste t será representada pela variável aleatória Bernoulli:

$$U_{jit} = \begin{cases} 1, & \text{se o item } i \text{ é respondido corretamente pelo indivíduo } j \text{ no teste } t; \\ 0, & \text{se o item } i \text{ não é respondido corretamente pelo indivíduo } j \text{ no teste } t; \end{cases}$$

com $i = 1, \dots, I_t$, $j = 1, \dots, N$ e $t = 1, \dots, T$

A Função de Resposta do Item ou *Curva Característica do Item (CCI)* mais apropriada para a maioria dos casos é o modelo logístico de três parâmetros (ML3), dado por:

$$P(U_{jit} = 1 | \theta_{jt}, \zeta_i) = c_i + (1 - c_i) \frac{1}{1 + e^{-D a_i (\theta_{jt} - b_i)}} \quad (1),$$

onde: θ_{jt} é a habilidade do j -ésimo respondente durante o teste t ; $\zeta_i = (a_i, b_i, c_i)'$ é o conjunto de características associadas ao item i , no qual b_i corresponde ao parâmetro de dificuldade (ou de posição), a_i refere-se ao parâmetro de discriminação e c_i é o parâmetro de acerto casual do item, que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente ao item; e D é um fator de escala. Utiliza-se $D = 1,7$ quando deseja-se que a função logística forneça

resultados semelhantes ao da função ogiva normal, ou $D = 1$ em caso contrário.

Embora estejam sendo considerados T testes tecnicamente distintos, alguns itens poderão estar em dois ou mais testes, de forma que $I \leq \sum_{t=1}^T I_t$. Neste trabalho será considerada uma ordenação nos I itens que compõem o conjunto dos T testes, representando-os por $\zeta = (\zeta_1, \dots, \zeta_I)$.

ESTRUTURAS DE COVARIÂNCIA

Quando o interesse consiste em acompanhar o desempenho de um mesmo grupo de avaliados ao longo do tempo, é importante incorporar possíveis estruturas de covariância entre as habilidades desses respondentes. A seguir são apresentadas algumas das principais estruturas para modelar essa dependência. Mais detalhes podem ser obtidos em Tavares e Andrade (2006).

MATRIZ DE COVARIÂNCIA DIAGONAL

Se considerarmos que as habilidades dos avaliados em um tempo são não correlacionadas com as habilidades nos outros $T - 1$ tempos, é possível afirmar que a matriz de covariâncias das habilidades nas T condições de avaliação é dada por:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_T^2 \end{bmatrix} \quad (2)$$

MATRIZ DE COVARIÂNCIA UNIFORME

Com essa estrutura, as variâncias são sempre as mesmas, iguais a σ^2 , ao longo do tempo, e a correlação entre as habilidades dos alunos é igual a $\rho \in (-1,1)$. É dada por:

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix} \quad (3)$$

MATRIZ DE COVARIÂNCIA BANDAS

Essa estrutura é um pouco semelhante à anterior, sendo que a diferença está no fato de ter uma dependência imediata, no sentido de que apenas as habilidades nos tempos imediatamente anterior e posterior a determinado tempo são correlacionadas entre si. Essa correlação também será denotada por ρ , cuja variação depende do valor de T .

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & 0 & \dots & 0 \\ \rho & 1 & \rho & \dots & 0 \\ 0 & \rho & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \quad (4)$$

MATRIZ DE COVARIÂNCIA AR(1)

Esta estrutura, conhecida como *autorregressiva de ordem 1*, assume que as correlações entre as habilidades diminuem à medida que aumenta a distância entre os instantes das observações. As variâncias são sempre as mesmas em todos os momentos. A forma dessa matriz é mostrada a seguir, com $\rho \in (-1,1)$:

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix} \quad (5)$$

Apresentadas as principais estruturas de covariância constantes da literatura, passa-se, a seguir, a abordar os modelos para curvas de crescimento utilizados nesse estudo.

CURVAS DE CRESCIMENTO

Nesta seção o interesse concentra-se no vetor de médias, $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_T)'$, que representa as habilidades médias em

cada tempo ou teste, sendo exploradas algumas formas de crescimento para as médias μ_t , do tipo:

$$\mu_t = f(t|\mathbf{a}) \quad (6),$$

onde f é uma função duplamente diferenciável com p parâmetros e $\mathbf{a} = (\alpha_1, \alpha_2, \dots, \alpha_p)'$ refere-se ao vetor de parâmetros da função. Assumindo-se que os testes foram realizados nos tempos t_1, t_2, \dots, t_T , o vetor de médias nos T testes pode ser escrito como:

$$\boldsymbol{\mu} = (f(t_1|\mathbf{a}), f(t_2|\mathbf{a}), \dots, f(t_T|\mathbf{a}))' \quad (7)$$

A seguir apresentam-se as principais funções propostas na teoria de curvas de crescimento para modelar o crescimento das médias populacionais μ_t para várias condições de avaliação (JONES, 1993). A Figura 1 traz a representação gráfica dessas curvas.

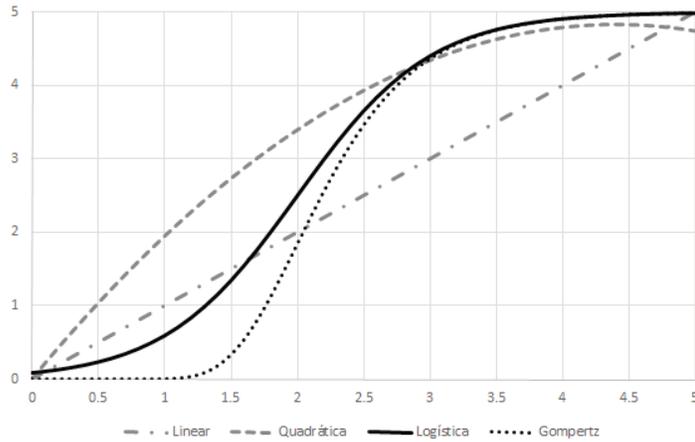
CURVA DE CRESCIMENTO LINEAR

O modelo para curva de crescimento linear é dado por:

$$\mu_t = \alpha_0 + \alpha_1 t \quad (8)$$

Embora simples, essa curva não é realista em muitos casos. No entanto, para determinado intervalo central, ela normalmente aproxima-se muito bem de funções não lineares. A maior restrição em seu uso reside em projeções para o início ou final da escala, e naturalmente falha em maior grau para tempos fora daqueles incluídos na pesquisa. Para garantir o crescimento dessa função, devemos ter $\alpha_1 > 0$.

FIGURA 1 - Modelos para curvas de crescimento



Fonte: Elaboração dos autores.

CURVA DE CRESCIMENTO QUADRÁTICA

O modelo para curva de crescimento quadrática é dado por:

$$\mu_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 \quad (9)$$

A função quadrática é mais flexível do que a linear, mas trata-se de uma parábola, de forma que, mesmo com uma escolha realista dos parâmetros, ela mantém o crescimento até certo ponto e depois começa a decrescer, apresentando um comportamento não desejado para uma curva de crescimento. No entanto, ela também pode ajustar-se muito bem aos dados em determinado intervalo, falhando com maior intensidade nas extremidades da escala e, principalmente, em projeções fora do intervalo coberto pela pesquisa. Para garantir o crescimento dessa função em um caso típico, devemos ter $\alpha_2 < 0$ e $\alpha_1 > 0$.

CURVA DE CRESCIMENTO LOGÍSTICA

O modelo para curva de crescimento logística é dado por:

$$\mu_t = \frac{\alpha_0}{1 + e^{-\alpha_1(t - \alpha_2)}} \quad (10)$$

Esse modelo é bem mais realista do que os anteriores, apresentando um comportamento não linear e sempre crescente quando $\alpha_0 > 0$ e $\alpha_1 > 0$. Nesse modelo a assíntota inferior é zero e denota aproximadamente a habilidade média para o menor nível do sistema educacional incluído na análise. De forma similar, o parâmetro α_0 representa a habilidade média do maior nível do sistema educacional inserido na análise. Essa curva assume um crescimento bem modesto da habilidade média para os instantes iniciais, depois aumenta gradualmente a velocidade de crescimento e, após certo ponto (denominado *ponto de inflexão*), volta a diminuir a velocidade. Seu crescimento é considerado simétrico, de forma que a velocidade observada antes do ponto de inflexão é igual àquela verificada após o referido ponto. Esse ponto de inflexão em que muda a velocidade de crescimento também é muito importante nesta abordagem, pois mostra o ponto limite em que há ganho acelerado de habilidade, e ele é α_2 para a função logística.

CURVA DE CRESCIMENTO GOMPERTZ

O modelo para curva de crescimento Gompertz é dado por:

$$\mu_t = \alpha_0 + \alpha_1 e^{-e^{-\alpha_2(t-\alpha_3)}} \quad (11)$$

Esse modelo é ainda mais realista do que os anteriores, também apresentando um comportamento não linear e sempre crescente quando $\alpha_1 > 0$ e $\alpha_2 > 0$. Nesse modelo a assíntota inferior é α_0 , e denota aproximadamente a habilidade média para o menor nível do sistema educacional. De forma similar, o parâmetro α_1 representa a habilidade média do maior nível do sistema educacional. Essa curva começa com um crescimento bem modesto da habilidade média, depois aumenta bastante a velocidade de crescimento e, após o ponto de inflexão, diminui a velocidade. No entanto, aqui ocorre uma diferença considerável com relação à curva logística, pois não há similaridade no comportamento antes e depois do ponto de inflexão. Ela demora um pouco mais para começar a crescer, mas quando iniciado adota um crescimento

mais acelerado do que a logística; porém, após algum tempo, elas praticamente se igualam. O ponto limite em que há ganho acelerado de habilidade é α_3 para a Gompertz.

ESTIMAÇÃO DOS PARÂMETROS DO MODELO

Embora não seja o escopo deste artigo, é importante traçar em poucas palavras o caminho adotado para estimação dos parâmetros das curvas de crescimento e da estrutura de covariância. Assim, I_t é o conjunto dos índices dos itens presentes no teste t , $\mathbf{U}_{j,t} = (U_{j1t}, U_{j2t}, \dots, U_{jI_t t})'$ corresponde ao vetor aleatório de respostas do indivíduo j no teste t , $\mathbf{U}_j = (\mathbf{U}'_{j,1}, \mathbf{U}'_{j,2}, \dots, \mathbf{U}'_{j,T})'$ refere-se ao vetor aleatório de respostas do indivíduo j em todos os testes e $\mathbf{U}_{\dots} = (\mathbf{U}'_{\dots,1}, \mathbf{U}'_{\dots,2}, \dots, \mathbf{U}'_{\dots,N})'$ é o vetor total de respostas. De forma similar, definimos os valores efetivamente observados por u_{jit} , $\mathbf{u}_{j,t}$, \mathbf{u}_j , e \mathbf{u}_{\dots} , respectivamente.

Os parâmetros da distribuição latente de $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_T)'$ são $\boldsymbol{\eta} = (\boldsymbol{\alpha}, \boldsymbol{\Sigma})$, e são referidos como *parâmetros populacionais*. Como os parâmetros dos itens são conhecidos, o interesse reside em estimar apenas os parâmetros populacionais de $g(\boldsymbol{\theta}|\boldsymbol{\eta})$.

O processo de estimação dos parâmetros adotado neste trabalho foi o *máxima verossimilhança marginal* (TAVARES, 2001), que consiste em maximizar a verossimilhança $L(\boldsymbol{\eta})$ dos dados. Também assume-se que a distribuição de $\boldsymbol{\theta}$ é normal multivariada $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Nessas condições, tem-se que a equação de estimação para $\boldsymbol{\eta}$ é:

$$\frac{\partial \log L(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \int_R \left[\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right]^{-*} g^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0 \quad (12)$$

com

$$g^*(\boldsymbol{\theta}) = \frac{1}{N} \sum_{j=1}^N g_j^*(\boldsymbol{\theta}) \quad \text{e} \quad g_j^*(\boldsymbol{\theta}) = \frac{P(\mathbf{u}_j | \boldsymbol{\theta}, \zeta) g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\mathbf{u}_j | \zeta, \boldsymbol{\eta})} \quad (13),$$

onde em (12) a equação foi dividida por N , sem alteração da solução, que pode ser obtida por meio de algum método numérico fornecendo as estimativas de máxima verossimilhança marginal dos parâmetros do modelo.

APLICAÇÃO AO PROJETO PDE

Nesta seção são discutidos os resultados da metodologia proposta a dados reais, considerando cada curva de crescimento apresentada e cada uma das estruturas de covariância adotadas. O foco principal é a modelagem da curva de crescimento, visto que a estrutura de dependência já tem sido tratada na literatura de TRI (TAVARES; ANDRADE, 2006). Em uma etapa prévia à aplicação a dados reais, foi elaborado um código denominado *LongIRT* na linguagem Ox (DOORNIK; OOMS, 2007), visto que ainda não há programa pronto para o caso deste artigo. De forma a testar se as estimativas obtidas eram precisas, houve uma etapa intensa de estudos com dados simulados, concluindo-se que as estimativas obtidas pelo programa reproduziam com grande precisão os parâmetros verdadeiros do modelo.

A aplicação refere-se ao Plano de Desenvolvimento da Escola (PDE-Escola), conduzido pelo Ministério da Educação (MEC) e coordenado pelo Fundescola (Fundo de Fortalecimento da Escola), em parceria com governos estaduais e municipais, com o objetivo de melhorar a qualidade do ensino. O estudo envolveu alunos da 4^a à 8^a série do ensino fundamental (5^o ao 8^o ano), que foram submetidos a provas em Língua Portuguesa e Matemática. Os momentos de avaliação foram abril de 1999 (*baseline*) e novembro de 1999, 2000, 2001, 2002 e 2003.

A modelagem da curva de crescimento se restringirá à disciplina Matemática. Para a obtenção dos resultados, foram selecionados apenas os alunos que compareceram a todos os testes, totalizando 1.987. Além disso, como os tempos entre as avaliações não foram regulares, resolveu-se adotar como instantes de avaliação os tempos 4, 11, 23, 35, 47 e 59, equivalentes aos meses de realização da avaliação, em que $t = 1$ equivale a janeiro de 1999. Dos 40 itens aplicados em cada teste, exceto o *baseline* em que foram 36 itens, utilizaram-se, na análise, apenas 34 itens do teste 1, 38 do teste 2, 36 do teste 3, 34 do teste 4, 40 do teste 5 e 34 do teste 6, o que totalizou 167 itens. Os demais foram excluídos por problemas na calibração e sugestão de especialistas da área pedagógica.

Na Tabela 5 são apresentados os valores adotados para os parâmetros dos itens, obtidos em etapa prévia. Empregou-se a métrica normal ($D = 1$) para o fator de escada, pois os parâmetros dos itens estavam na escala logística (ANDRADE; TAVARES; VALLE, 2000).

O estudo foi planejado com vistas à equalização dos resultados ao longo do tempo, garantindo a comparabilidade por meio de itens comuns, como mostra a Tabela 1, em que se pode observar que o teste 2 utilizou 10 itens do teste 1, o teste 3 usou 5 itens comuns aos testes 1 e 2, e também 5 exclusivamente do teste 2, totalizando 10 itens comuns com o teste 2. No geral, o teste 4 utilizou 1, 3 e 5 itens dos testes 1, 2 e 3, respectivamente, o teste 5 usou 3 e 7 itens dos testes 3 e 4 e o teste 6 utilizou 1, 1 e 8 itens dos testes 3, 4 e 5.

TABELA 1 – Itens comuns entre os testes

	TESTE 1	TESTE 2	TESTE 3	TESTE 4	TESTE 5	TESTE 6
Teste 1	34	10	5	1	0	0
Teste 2	10	38	10	4	0	0
Teste 3	5	10	36	7	3	1
Teste 4	1	4	7	34	10	2
Teste 5	0	0	3	10	40	10
Teste 6	0	0	1	2	10	34

Fonte: Elaboração dos autores.

Na Tabela 2 são apresentadas as estimativas dos parâmetros das curvas de crescimento e suas respectivas projeções (médias), de acordo com a estrutura de covariância adotada. Cabe informar que as estimativas das proficiências médias, quando realizadas as análises com a base de dados completa, ou seja, com todos os alunos envolvidos no estudo, utilizando-se o *software* BILOG-MG (ZIMOWSKI et al., 2003), foram, respectivamente para os seis testes, 0, 0.25, 0.74, 1.31, 1.54 e 1.66, em que o *baseline* foi a população referência. Ao se considerar apenas o conjunto de 1.987 presentes em todos os testes, as proficiências médias foram 0.37, 0.53, 0.95, 1.37, 1.46 e 1.65. Esses últimos resultados foram obtidos no próprio LongIRT, levando em conta a estrutura diagonal, sem curvas de crescimento, fixados os parâmetros dos itens.

TABELA 2 – Estimativas dos parâmetros das curvas de crescimento para cada estrutura de covariância

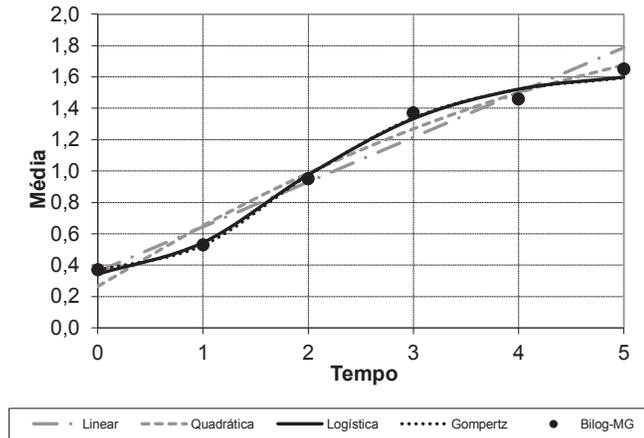
ESTRUTURA	CURVA	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	α_0	α_1	α_2	α_3
Diagonal	Linear	0.3598	0.645	0.931	1.217	1.502	1.788	0.0742	0.286		
	Quadrática	0.2651	0.653	0.988	1.270	1.500	1.678	-0.1750	0.466	-0.0262	
	Logística	0.3420	0.544	0.976	1.334	1.522	1.599	1.6411	0.090	18.767	
	Gompertz	0.3680	0.528	0.973	1.340	1.521	1.594	0.3637	1.273	1.0218	2.702
Uniforme	Linear	0.3549	0.639	0.922	1.206	1.490	1.774	0.0711	0.284		
	Quadrática	0.3284	0.644	0.943	1.226	1.492	1.742	-0.004	0.340	-0.0082	
	Logística	0.3948	0.570	0.940	1.293	1.538	1.672	1.7851	0.072	21.534	
	Gompertz	0.3752	0.566	0.932	1.287	1.528	1.666	0.3462	1.463	0.7278	2.878
Bandas	Linear	-0.0120	0.308	0.627	0.947	1.267	1.587	-0.332	0.320		
	Quadrática	-0.0770	0.321	0.679	0.999	1.280	1.522	-0.514	0.456	-0.0195	
	Logística	0.0417	0.100	0.381	0.942	1.363	1.504	1.5466	0.130	31.593	
	Gompertz	0.0184	0.218	0.659	1.080	1.347	1.486	0.000	1.609	0.8061	2.859
AR(1)	Linear	0.3726	0.652	0.932	1.211	1.491	1.770	0.093	0.280		
	Quadrática	0.3599	0.661	0.950	1.229	1.497	1.753	0.0481	0.317	-0.0055	
	Logística	0.4180	0.595	0.961	1.309	1.553	1.690	1.8115	0.070	21.265	
	Gompertz	0.3966	0.587	0.952	1.306	1.547	1.685	0.3678	1.460	0.7278	2.878

Fonte: Elaboração dos autores.

Nota-se que, de forma geral, as estimativas dos parâmetros das curvas de crescimento na Tabela 2 estão próximas, mas com alguma diferença identificável, principalmente para a estrutura de bandas, com estimativas dos parâmetros das curvas de crescimento claramente distante das demais. Isso pode ter ocorrido por instabilidade do processo de estimação, visto que para valores altos de correlação essa matriz não é positiva definida (TAVARES, 2001). As estruturas uniforme e autorregressiva de ordem 1 foram as que ficaram mais próximas quanto às estimativas dos parâmetros das curvas de crescimento, como seria de se esperar com base na Tabela 3.

Na Figura 2 observam-se as projeções das médias para cada curva de crescimento e os resultados pontuais obtidos no BILOG-MG. As curvas logística e Gompertz são as que mais se aproximam dos resultados daquele aplicativo, sem fácil identificação de qual das duas realmente sobressai no ajuste dos dados.

FIGURA 2 – Evolução das habilidades médias



Fonte: Elaboração dos autores.

Relativamente à estrutura de dependência observada, apresentam-se na Tabela 3 as correlações entre os escores (número de acertos) nos seis instantes de avaliação. Nota-se claramente uma alta correlação entre os escores, certamente presente também nas proficiências individuais. Percebe-se um suave decaimento das correlações ao longo do tempo, menos acentuado do que a estrutura AR(1), mas suficientemente claro a ponto de descartar as estruturas diagonal e bandas. No entanto, é possível que as diferentes estruturas de covariância tenham um impacto modesto nas estimativas dos parâmetros das curvas de crescimento e das proficiências médias, mas trazendo bastante informação para a diminuição dos erros-padrão de estimativa (TAVARES, 2001).

TABELA 3 - Matriz de correlações observadas

	Teste ₁	Teste ₂	Teste ₃	Teste ₄	Teste ₅	Teste ₆
Teste ₁	1.000	0.717	0.628	0.593	0.580	0.485
Teste ₂	0.717	1.000	0.686	0.650	0.641	0.544
Teste ₃	0.628	0.686	1.000	0.654	0.647	0.581
Teste ₄	0.593	0.650	0.654	1.000	0.674	0.601
Teste ₅	0.580	0.641	0.647	0.674	1.000	0.640
Teste ₆	0.485	0.544	0.581	0.601	0.640	1.000

Fonte: Elaboração dos autores.

Para identificar, entre estas duas estruturas de covariância, qual curva de crescimento melhor se ajusta aos dados, utilizamos os critérios AIC (Akaike's Information Criterion) e BIC (Bayesian Information Criterion), obtidos por:

$$AIC = -2\log L(\hat{\eta}) + 2n_p \quad (14)$$

$$BIC = -2\log L(\hat{\eta}) + 2n_p \ln(N) \quad (15),$$

onde $L(\hat{\eta})$ representa o valor da verossimilhança marginal e n_p corresponde ao número de parâmetros envolvidos no modelo (JONES, 1993). A estrutura de covariância que, em conjunto com a curva de crescimento, gerar o menor valor para AIC e BIC comporá o modelo escolhido.

TABELA 4 - Log-verossimilhança associada a cada estrutura de covariância e curva de crescimento com critérios AIC e BIC

MODELO	CURVA	-2LOG-VEROS.	NP	AIC	BIC
Uniforme	Linear	499.598,23	4	499.606,23	499.658,98
	Quadrática	499.589,40	5	499.599,40	499.665,34
	Logística	499.533,42	5	499.543,42	499.609,36
	Gompertz	499.532,24	6	499.544,24	499.623,38
AR(1)	Linear	499.464,50	4	499.472,50	499.525,25
	Quadrática	499.460,56	5	499.470,56	499.536,51
	Logística	499.389,83	5	499.399,83	499.465,78
	Gompertz	499.389,51	6	499.401,51	499.480,65

Fonte: Elaboração dos autores.

A Tabela 4 apresenta os valores de AIC e BIC para as duas estruturas – uniforme e AR(1). Nota-se, de forma geral, que a estrutura AR(1) foi a que melhor se ajustou aos dados, com valores de AIC e BIC predominantemente menores para cada curva. E, para essa estrutura, verifica-se que a curva de crescimento logística foi claramente a que melhor modela o comportamento das proficiências médias do grupo ao longo dos seis instantes de avaliação, com estimativas dos parâmetros da curva de crescimento dadas por $\hat{\alpha}_0 = 1,8115$, $\hat{\alpha}_1 = 0,0697$, e $\hat{\alpha}_2 = 21,2647$, e da estrutura de covariância $\hat{\rho} = 0,9000$ e $\hat{\sigma}^2 = 1,1690$. As estimativas médias geradas com esta curva de crescimento para os tempos adotados são dadas por $\hat{\mu}_4 = 0,4180$, $\hat{\mu}_{11} = 0,5947$, $\hat{\mu}_{23} = 0,9605$, $\hat{\mu}_{35} = 1,3092$, $\hat{\mu}_{47} = 1,5534$ e $\hat{\mu}_{59} = 1,6899$.

Nota-se que na escala de proficiência $N(0,1)$ adotada (distribuição normal com média zero e desvio padrão 1), a proficiência média em Matemática do último nível do estudo, que é 12 no sistema educacional atual, deverá estar próxima da estimativa da assíntota superior do modelo logístico, ou seja, $\hat{\mu}_{59} = 1,6899$, tendo como grupo referência o conjunto de alunos do 4º ano que realizaram a avaliação em abril de 1999. No entanto, essa é uma extrapolação bastante dilatada, visto que o último nível avaliado foi o 9º ano, de forma que precisa ser interpretada com muita cautela.

Ainda com base na estimativa de α_2 , verifica-se que o período em que ocorre a maior taxa de crescimento da habilidade está um pouco além do 21º mês, ou seja, possivelmente já em outubro da 5ª série/6º ano escolar, a partir do qual a taxa de crescimento começa a diminuir.

Para a estrutura de correlação AR(1), obteve-se uma estimativa $\hat{\rho} = 0,9000$ para a correlação entre testes adjuntos. Essa informação garante que as respostas foram muito coerentes ao longo dos testes, ou seja, alunos com proficiências altas em um teste também as tiveram em testes próximos, assim como alunos com baixa proficiência em um teste também apresentaram tal desempenho nos testes adjuntos, sem melhorias significativas.

CONSIDERAÇÕES GERAIS

Neste trabalho foram propostos modelos para o comportamento das habilidades médias em Matemática de um mesmo grupo de indivíduos acompanhados ao longo do tempo (estudo longitudinal), por meio da estimação dos parâmetros de curvas de crescimento. Foram apresentadas as principais curvas constantes da literatura estatística: linear, quadrática, logística e Gompertz, mas há outras que também podem ser ajustadas (JONES, 1993).

Também foi abordado o fato de haver uma possível dependência entre os resultados nos vários instantes de avaliação, sendo propostas estruturas de covariância para modelar tal dependência, particularmente a diagonal, uniforme, autorregressiva de ordem 1 e estrutura de bandas. Optou-se por não incluir o caso geral, que seria a matriz não estruturada, porque nos casos longitudinais espera-se existir alguma estrutura na matriz de dependências (ANDRADE; TAVARES, 2006). Além disso, o estudo não foi absolutamente equiespaçado, uma vez que o primeiro teste (*baseline*) foi feito em período diferente dos demais, o que é normal neste tipo de estudo. Ainda que a curva de crescimento adotada na aplicação tenha incorporado os tempos referentes aos meses dos exames (4, 11, 23, 35, 47 e 59), a estrutura de covariância não incorporou tal informação, considerando que o impacto seria insignificante. Para estudos em que os intervalos entre os testes sejam muito distintos, as matrizes de covariância deverão ser apropriadas para tal situação, se necessário. Por exemplo, na AR(1) a correlação entre as habilidades nos tempos t_1 e t_2 pode ser escrita como $\rho^{|t_2 - t_1|}$.

Observou-se descritivamente que, em termos da dependência entre os escores dos estudantes, as estruturas teóricas que mais se aproximaram da matriz de covariância dos escores foram a uniforme e a AR(1). No entanto, a correlação na estrutura verificada dos escores decaiu com o tempo, aproximando-se da matriz AR(1), o que seria um fato esperado na prática, mas com uma diminuição um pouco mais lenta do que a AR(1). Embora a relação entre o escore e a proficiência não seja linear, e por isso as estimativas das correlações entre escores não se refletem nas correlações entre

proficiências, é de se esperar algum comportamento similar no sentido, embora não em grau.

Consideramos conhecidos os parâmetros dos itens envolvidos no estudo, pois o foco do trabalho foi a abordagem de curvas de crescimento com dependência. Foram feitas aplicações a dados simulados e os resultados foram bastante promissores, com um processo de estimação realmente satisfatório, para os parâmetros tanto das estruturas de dependência como das curvas de crescimento, mas esses resultados foram omitidos neste artigo para torná-lo mais direto e aplicado. Outras etapas de simulação considerando outras curvas de crescimento e outras estruturas de covariância ainda serão realizadas. Uma extensão em que também serão calibrados os parâmetros dos itens será desenvolvida em trabalhos futuros. Também é de interesse o caso multidimensional (BARBETTA et al., 2014), em que essas curvas unidimensionais passarão a evoluir no plano ou em maior dimensão, podendo englobar o caso multivariado, no qual são consideradas conjuntamente as áreas de Linguagem e Matemática, por exemplo. Em todos os casos, os códigos atuais serão traduzidos para o programa R e haverá disponibilização dos respectivos pacotes para uso dos pesquisadores interessados.

Foi feita uma aplicação a dados reais, resultantes da pesquisa “Avaliação do desempenho: fatores associados” do projeto Fundescola, aplicada a alunos da 4ª à 8ª série do ensino fundamental, no período de 1999 a 2003. Utilizou-se apenas o conjunto de indivíduos que participou de todas as condições de avaliação, totalizando 1.987 alunos. Os itens foram previamente calibrados, de forma que seus parâmetros já eram conhecidos e foram obtidos da calibração do PDE-Escola. Relativamente aos resultados da análise propriamente, verifica-se um bom ajuste da curva de crescimento logística associada à estrutura autorregressiva de ordem 1. Outros estudos podem ser planejados com itens na escala Saeb (250,50) com grupos distintos.

Com essa metodologia, é possível obter estimativas de habilidades médias em tempo ou níveis não incluídos no estudo, bem como estratos tais como escolas, redes, cidades,

estados e federação, podendo-se ainda comparar as taxas de crescimento, o ponto de maior desenvolvimento dos estudantes em várias esferas e as previsões a médio e longo prazos.

TABELA 5 - Estimativas dos parâmetros dos itens (PDE-Escola)

ITEM	TESTE	A	B	C	ITEM	TESTE	A	B	C	ITEM	TESTE	A	B	C
101	1	0.78	0.84	0.35	229	2	1.23	-0.55	0.21	440	4	1.44	2.48	0.26
102	1	0.68	0.22	0.22	230	2 e 3	1.46	1.48	0.27	501	5 e 6	0.58	0.53	0.38
103	1	1.26	0.37	0.35	233	2	1.05	0.34	0.24	502	5	0.93	1.83	0.17
104	1 e 2	1.12	0.54	0.34	234	2	1.29	0.06	0.1	503	5	0.57	1.53	0.23
105	1	0.49	-2.57	0.15	236		1.16	-0.16	0.13	504	5	0.78	2.27	0.5
106	1	0.57	-0.7	0.08	240		0.88	-0.1	0.17	506	5	1.56	2.14	0.17
108	1	0.49	-0.83	0.16	302	3	0.49	1.83	0.15	507	5	1.18	2.6	0.32
109	1, 2 e 3	1.17	0.01	0.31	303	3, 4, 5 e 6	0.57	-0.29	0.02	508	5	0.98	2.77	0.23
110	1 e 2	1.07	0.25	0.15	308	3	0.96	1.67	0.19	511	5 e 6	0.84	2.93	0.36
111	1	0.7	-0.8	0.09	309	3, 4 e 5	0.63	0.12	0.11	513	5	0.78	1.33	0.29
112	1	0.43	-0.84	0.19	311	3	0.5	1.7	0.15	514	5	0.68	2.03	0.25
114	1, 2 e 3	0.75	0.39	0.17	312	3, 4 e 5	1.01	1.01	0.3	515	5 e 6	1.24	3.17	0.37
115	1	0.45	-0.4	0.2	315	3	1.06	1.46	0.24	517	5	1	1.57	0.22
116	1	0.76	-0.62	0.12	316	3	1.1	-0.07	0.23	518	5	1.5	3.27	0.18
117	1	0.66	0.52	0.07	317	3	1.05	1.04	0.21	519	5	0.54	2.11	0.19
118	1	0.64	0.02	0.14	318	3	1.07	1.44	0.25	521	5	0.93	1.4	0.38
119	1, 2, 3 e 4	0.8	0.57	0.29	321	3	0.52	0.93	0.12	522	5	1.08	2.89	0.28
120	1	0.55	0.03	0.07	323	3	0.44	0.42	0.08	523	5	1.61	2.12	0.15
121	1	0.77	-1.03	0.06	324	3 e 4	0.91	0.37	0.17	526	5 e 6	1.06	2.5	0.24
122	1 e 2	1.08	0.66	0.11	326	3	0.61	0.45	0.08	527	5 e 6	0.45	3.86	0.25
123	1	0.56	0.17	0.22	327	3	1.42	1.91	0.22	528	5	2.2	2.73	0.19
124	1, 2 e 3	1.1	0.62	0.16	329	3	1.17	2.19	0.28	530	5	1.23	2.69	0.18
125	1	0.97	-0.33	0.15	330	3	1.34	1.32	0.26	531	5	1.28	2.47	0.19
126	1 e 2	0.95	0.12	0.11	331	3	0.71	0.03	0.08	532	5 e 6	0.98	1.63	0.45
127	1	0.79	0.6	0.1	332	3	1.1	1.05	0.13	534	5 e 6	0.94	3.69	0.39
128	1, 2 e 3	1.36	0.67	0.21	334	3 e 4	0.72	0.27	0.12	535	5	1.1	1.99	0.19
129	1	1.04	0.32	0.08	335	3	0.91	0.86	0.13	536	5	0.52	2.86	0.19

(continua)

(continuação)

TABELA 5 - Estimativas dos parâmetros dos itens (PDE-Escola)

ITEM	TESTE	A	B	C	ITEM	TESTE	A	B	C	ITEM	TESTE	A	B	C
130	1	0.43	0.46	0.11	336	3	0.7	1.29	0.21	537	5 e 6	1.44	2.77	0.21
131	1	0.42	1.16	0.02	337	3	1.21	1.43	0.13	538	5	1.04	1.38	0.12
132	1	1.94	1	0.35	338	3	0.84	0.75	0.2	539	5	0.91	2.33	0.12
133	1	0.75	1.97	0.13	339	3	1.26	1.29	0.19	540	5	2.21	2.78	0.2
134	1 e 2	0.77	-0.99	0.07	340	3	0.95	1.63	0.2	602	6	0.55	2.78	0.34
135	1	1.07	-0.09	0.05	401	4	0.74	-0.21	0.17	603	6	0.52	4.22	0.13
136	1	0.78	0.8	0.25	403	4 e 5	1.06	1.7	0.23	607	6	0.41	1.19	0.14
201	2	1.37	2.35	0.2	406	4 e 5	1.02	1.9	0.17	608	6	0.82	2.93	0.19
203	2	0.74	0.74	0.17	408	4	0.95	2.38	0.22	613	6	0.43	2.66	0.12
204	2 e 3	0.53	-0.63	0.11	409	4	1.78	2.41	0.16	616	6	1	2.07	0.26
207	2	1.24	0.72	0.18	418	4	0.72	1.97	0.29	618	6	0.69	3.49	0.21
208	2	1.55	1.44	0.1	419	4	1.38	2.8	0.14	619	6	0.64	-0.61	0.1
209	2	0.87	-0.27	0.14	420	4 e 5	1.42	1.63	0.15	621	6	0.54	-0.07	0.08
210	2	0.78	0.37	0.17	422	4 e 5	1.01	1.82	0.29	622	6	0.93	3.14	0.31
212	2	1	1.85	0.26	423	4 e 5	0.7	2.05	0.18	623	6	1.08	2.5	0.31
213	2	0.81	0.53	0.32	424	4	1.08	2.27	0.22	626	6	1.13	2.07	0.2
214	2	1.49	1.56	0.14	425	4	0.74	1.82	0.2	627	6	0.72	3.61	0.24
215	2	1.51	1.47	0.22	427	4	1.75	2.06	0.21	628	6	0.72	4.59	0.28
216	2	0.87	-1.05	0.05	428	4	0.87	0.67	0.05	629	6	1.04	2.81	0.29
218	2 e 3	0.97	-0.16	0.17	430	4	2.17	2.74	0.17	630	6	0.81	4.12	0.3
219	2, 3 e 4	0.93	0.88	0.27	431	4	1.44	2.37	0.17	632	6	0.6	2.04	0.13
220	2	1.41	1.5	0.18	432	4, 5 e 6	1.29	1.96	0.27	633	6	0.45	4.74	0.28
221	2	0.71	0.56	0.17	433	4	0.83	1.23	0.18	634	6	0.4	3.46	0.22
222	2	0.89	-0.82	0.05	434	4 e 5	1.06	2.14	0.14	635	6	1.16	4.26	0.2
223	2 e 3	1.5	0.87	0.19	435	4	1.29	2.97	0.29	636	6	0.85	2.02	0.33
225	2	0.97	-0.96	0.12	436	4	1.2	1.84	0.31	638	6	1.01	3.77	0.26
226	2	0.63	0.25	0.16	437	4	1.98	2.81	0.12	639	6	0.87	2.39	0.24
227	2	1.15	1.3	0.19	438	4	1.89	2.35	0.26	640	6	0.69	0.1	0.11
228	2	2	1.1	0.25	439	4	0.67	3.59	0.21					

Fonte: Elaboração dos autores.

REFERÊNCIAS

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. *Teoria da resposta ao item: conceitos e aplicações*. São Paulo: Associação Brasileira de Estatística, 2000.

BAKER, F. B.; KIM, S. H. (Ed.). *Item response theory: parameter estimation techniques*. 2. ed. Boca Raton, Flórida: CRC, 2004.

BARBETTA, P. A.; TREVISAN, L. M.; TAVARES, H.; MACEDO AZEVEDO, T. C. A. de. Aplicação da teoria da resposta ao item uni e multidimensional. *Estudos em Avaliação Educacional*, São Paulo, v. 25, n. 57, p. 280-302, 2014.

BROOKE, N.; SILVA FERNANDES, N. da; MIRANDA, I. P. H. de; SOARES, T. M. Modelagem do crescimento da aprendizagem nos anos iniciais com dados longitudinais da pesquisa GERES. *Educação e Pesquisa*, São Paulo, v. 40, n. 1, p. 77-94, 2014.

CARNOY, M. et al. How schools and students respond to school improvement programs: the case of Brazil's PDE. *Economics of Education Review*, v. 27, n. 1, p. 22-38, 2008.

DOORNIK, J. A.; OOMS, M. *Introduction to Ox: an object-oriented matrix language*. London: Timberlake Consultants Ltd and Oxford, 2007.

FRANCO, C.; BROOKE, N.; ALVES, F. Estudo longitudinal sobre qualidade e equidade no ensino fundamental brasileiro: GERES 2005. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 16, n. 61, p. 625-638, 2008.

GOLDSTEIN, H.; HUIQI, P.; RATH, T.; HILL, N. *The use of value-added information in judging school performance*. London: Institute of Education, University of London, 2000.

JONES, R. H. *Longitudinal data with serial correlation: a state-space approach*. London: Chapman and Hall, 1993.

MOREIRA JR., F. Aplicações da Teoria de Resposta ao Item (TRI) no Brasil. *Rev. Bras. Biom.*, São Paulo, v. 28, n. 4, p. 137-170, 2010.

RAUDENBUSH, S. W.; FOTIU, R. P.; CHEONG, Y. F. Inequality of access to educational resources: a national report card for eighth grade math. *Educational Evaluation and Policy Analysis*, Thousand Oaks, CA, v. 20, n. 4, p. 253-268, 1998.

SAKAI, M. H.; FERREIRA FILHO, O. F.; MATSUO, T. Avaliação do crescimento cognitivo do estudante de medicina: aplicação do teste de equalização no teste de progresso. *Revista Brasileira de Educação Média*, Londrina, v. 35, n. 4, p. 493-501, 2011.

SCHAFER, W. D.; LISSITZ, R. W.; ZHU, X.; ZHANG, Y.; HOU, X.; LI, Y. *Using student growth models for evaluating teachers and schools*. Technical report. College Park, MD: Marces, 2012.

SMITH, R. L.; YEN, W. M. Models for evaluating grade-to-grade growth. In: LISSITZ, R. W. (Ed.). *Longitudinal and value added modeling of student performance*. Maple Grove, MN: JAM Press, 2006. p. 82-94.

TAVARES, H. R. *Teoria da resposta ao item para dados longitudinais*. 2001. 118 f. Tese (Doutorado em Estatística) – Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, 2001.

TAVARES, H. R.; ANDRADE, D. F. Item Response Theory for longitudinal data: item and population ability parameters estimation. *Test*, v. 15, n. 1, p. 97-123, 2006.

VAN DER LINDEN, W. J.; HAMBLETON, R. K. (Ed.). *Handbook of modern item response theory*. New York: Springer Science & Business Media, 2013.

ZIMOWSKI, M. F.; MURAKI, E.; MISLEVY, R. J.; BOCK, R. D. *Bilog-Mg 3 for Windows*. Chicago: Scientific Software International, 2003.

HELITON RIBEIRO TAVARES

Professor Associado da Faculdade de Estatística e do Programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará (UFPA), Belém, Pará, Brasil.
Pesquisador Associado da Fundação Vunesp, São Paulo, São Paulo, Brasil
heliton@ufpa.com

ADRIANA MORAES DE CARVALHO

Mestre pelo Programa de Pós-Graduação em Matemática e Estatística (PPGME) da Universidade Federal do Pará (UFPA), Belém, Pará, Brasil. Analista quantitativo da Fundação Vunesp, São Paulo, São Paulo, Brasil
adriana.moraesc@gmail.com

WALTER LANA LEITE

Professor Associado do College of Education da University of Florida, Gainesville, Florida, Estados Unidos
walter.leite@coe.ufl.edu

Recebido em: FEVEREIRO 2016

Aprovado para publicação em: MARÇO 2016

