

VOLATILIDADE DOS RESULTADOS DE PROFICIÊNCIA E SEU IMPACTO SOBRE AS METAS DO IDEB NAS ESCOLAS PÚBLICAS DE MINAS GERAIS

Luís A. F. Pontes*

Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora - MG, Brasil

Tufi Machado Soares**

Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora - MG, Brasil

RESUMO: Este artigo tem dois objetivos principais: mensurar a volatilidade dos resultados das avaliações de Língua Portuguesa e Matemática nas escolas públicas mineiras; e estimar o possível impacto dessas flutuações nos resultados do cumprimento das metas do Ideb. Para tal, analisa-se um banco com dados provenientes da Prova Brasil e do Sistema Mineiro de Avaliação da Educação Básica (Simave). O método estatístico empregado é um modelo longitudinal linear hierárquico que tem a escola como nível de interesse. Entre as principais conclusões obtidas estão: o fato de que as flutuações das notas médias das escolas são bastante acentuadas, mesmo quando se leva em conta a sua tendência de crescimento; e a constatação de que o impacto sobre o cumprimento das metas do Ideb, devido a tais flutuações, pode ser considerável, a ponto de facilitar ou dificultar o sucesso das escolas quanto ao seu cumprimento.

Palavras-chave: Ideb. Regressão linear hierárquica. Volatilidade dos resultados educacionais.

VOLATILITY OF PROFICIENCY RESULTS AND ITS IMPACT ON IDEB GOALS IN THE PUBLIC SCHOOLS OF MINAS GERAIS

ABSTRACT: This paper has two main purposes. The first one is to obtain a volatility measure of Portuguese and Mathematics mean scores in Minas Gerais State public schools. The second one is to estimate the probable impact of these fluctuations on results of respective Ideb goals. The analysis is based on data obtained from Prova Brasil and Simave (*Sistema Mineiro de Avaliação da Educação Básica*) tests. The used statistical technique is a longitudinal hierarchical model with the schools as the level of interest. One of the main conclusions obtained is the fact that fluctuation in school mean scores is considerable, even

<http://dx.doi.org/10.1590/0102-4698153262>

Elocation-id - e153262

*Doutor em Educação pela Universidade Federal de Juiz de Fora (UFJF), analista de dados sênior do Centro de Políticas Públicas e Avaliação da Educação (CAEd). E-mail: <luis@caed.ufjf.br>.

**Doutor em Engenharia Elétrica pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Professor do Departamento de Estatística da Universidade Federal de Juiz de Fora (UFJF), Coordenador de Pesquisa do Centro de Políticas Públicas e Avaliação da Educação (CAEd). <E-mail: tufi@caed.ufjf.br>.

when one takes into account its growth rates. Another relevant conclusion is the fact that this fluctuation can have a considerable impact on the success or failure of a given school in reaching its respective Ideb goals.

Keywords: Ideb. Linear hierarchical regression. Volatility in school results.

INTRODUÇÃO: O IDEB E A VOLATILIDADE DOS RESULTADOS EDUCACIONAIS

O Índice de Desenvolvimento da Educação Básica (Ideb) foi criado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), do Ministério da Educação do Brasil. Seus primeiros resultados foram divulgados no ano de 2005, a partir de quando novos valores têm sido calculados a cada dois anos. Segundo o próprio Inep (BRASIL, s/d.a, p. 1), trata-se de:

[...] um indicador de qualidade educacional que combina informações de desempenho em exames padronizados (Prova Brasil ou Saeb) – obtido pelos estudantes ao final das etapas de ensino (4ª e 8ª séries do ensino fundamental e 3ª série do ensino médio) – com informações sobre rendimento escolar (aprovação).

Desde o seu surgimento, o Ideb vem adquirindo grande relevância no cenário educacional brasileiro, sendo amplamente utilizado como medida de qualidade educacional a ser considerada para diversos propósitos relevantes, tais como “o gerenciamento das escolas e redes de ensino, o planejamento escolar, a prestação de contas, a informação para a opinião pública, etc.” (PONTES, 2015, p. 8). E, nessa mesma linha, Soares e Xavier (2013, p. 904) também observam que

O Ideb tornou-se a forma privilegiada e frequentemente a única de se analisar a qualidade da educação básica brasileira e, por isso, tem tido grande influência no debate educacional no país. Sua introdução colocou no centro desse debate a ideia de que hoje os sistemas educacionais brasileiros devem ser avaliados não apenas pelos seus processos de ensino e gestão, mas principalmente pelo aprendizado e trajetória escolar dos alunos.

Um ponto de grande interesse acerca do Ideb é o comportamento desse índice ao longo da sua série histórica para diferentes unidades de interesse, como escolas, municípios e estados, além do país como um todo. E, nesse sentido, segundo observa Pontes (2015), o Ideb experimentou considerável crescimento, notadamente no 5º ano do Ensino Fundamental e na rede pública, muito embora se perceba em avaliações mais recentes – como as de 2011 e 2013 – uma desaceleração, ou mesmo estagnação, dessas medidas, que se torna ainda mais preocupante na rede particular e no Ensino Médio.

Outro ponto de grande relevância, particularmente para o presente estudo, é a questão das metas do Ideb, que foram oficialmente

traçadas para diversas unidades educacionais de interesse, segundo um modelo logístico de crescimento (BRASIL, s/d.b). E, nesse ponto, uma questão tratada com especial detalhe no presente texto é a associação entre possíveis problemas na determinação dessas metas e a volatilidade das medidas educacionais de desempenho.

Para se atingir esse propósito, utilizar-se-á uma base de dados com a série histórica de Língua Portuguesa e Matemática do Programa de Avaliação da Educação Básica (Proeb) do Sistema Mineiro de Avaliação da Educação (Simave), a qual cobre todas as suas edições entre os anos 2000 e 2011, bem como as três primeiras edições da Prova Brasil, correspondentes aos anos de 2005, 2007 e 2009. Porém, antes de se detalharem as características desse banco de dados, será abordada a questão da volatilidade dos resultados educacionais, tema central do presente trabalho.

Muito embora os sistemas de avaliação em larga escala venham se expandindo consideravelmente no mundo moderno, uma das críticas que se fazem a eles é não raro os testes nos quais eles se baseiam poderem ser influenciados por diversos e significativos fatores que não apenas a qualidade do ensino oferecido pelas escolas. Por exemplo, segundo Kane *et al.* (2002), um problema estatístico diretamente associado a essas imprecisões decorre da chamada volatilidade dos resultados das avaliações. Os autores a ela se referem assim:

Essa volatilidade pode prejudicar seriamente os sistemas educacionais de responsabilização. A partir do momento em que as notas obtidas nos testes são capazes de ocasionar recompensas ou sanções, os profissionais nas escolas passam a correr um grande risco de serem punidos ou recompensados por resultados que, na verdade, estão fora de seu controle. Além do mais, a partir do momento em que essas classificações de desempenho são utilizadas para identificar as melhores práticas em educação, virtualmente qualquer filosofia [pedagógica] poderá um dia vir a ser endossada, o que só serve para aumentar a confusão acerca dos méritos de diferentes estratégias de reforma educacional. (KANE *et al.*, 2002, p. 236)

Ainda segundo esses autores, as fontes de volatilidade costumam provir de uma de duas origens. Uma delas deriva do fato de que, nas avaliações transversais, as mais comumente utilizadas nos Estados Unidos – contexto do artigo mencionado – e também no Brasil, se costuma avaliar uma mesma série escolar durante vários anos seguidos, fazendo com que, a cada edição do teste, haja, para uma mesma escola, nova coorte de estudantes, diferente tanto das que vieram antes dela quanto das que virão depois. E, ainda que certas características escolares, como o índice socioeconômico médio do seu respectivo alunado, não costumem oscilar muito de um ano para o outro, há, por outro lado, variações consideráveis que podem

ocorrer numa escola em curto espaço de tempo. Um simples exemplo disso é a possível deterioração do clima de aprendizado dentro de sala de aula decorrente do surgimento de uma liderança “negativa” entre os alunos, algo que, por sua vez, pode diminuir o resultado médio de proficiência de uma escola, particularmente nos casos em que essa turma na qual surge tal problema é uma das poucas, ou mesmo a única, existentes naquela escola, como é, de fato, muitas vezes o caso.

A tudo isso se juntaria, ainda, uma segunda fonte de variação, esta de caráter aleatório e praticamente incontrolável: um leque virtualmente infinito de possibilidades de ocorrência de eventos capazes de introduzir distúrbios ou alterações que podem impactar consideravelmente os resultados de um único teste. Exemplos disso são, digamos: barulho incontrolável e inadvertido no dia da prova, como o de uma britadeira funcionando na rua; surto de gripe; estabelecimento de uma relação extremamente cordial – ou, ao contrário, bastante inamistosa – entre os examinados e seus professores ou aplicadores do seu teste; etc.

Diante disso, é aconselhável que as decisões sobre o desempenho de dada escola considerem os seus respectivos resultados obtidos ao longo de certo período de tempo, e não apenas o resultado da última avaliação, tendo em vista que, ao longo de uma série histórica, uma escola específica deve apresentar tendência aproximadamente previsível de evolução do nível de aprendizado de seus alunos. Ainda assim, podem ocorrer também flutuações de desempenho, aparentemente aleatórias, em torno de seus respectivos valores médios.

Ponto particularmente preocupante ocorre quando, por exemplo, os resultados de um único ano específico são utilizados para políticas de grandes consequências para as escolas, como é precisamente o caso dos resultados da Prova Brasil de 2005, que serviram de base para que o Inep traçasse metas individuais de desempenho do Ideb para cada escola considerada na prova nacional daquele ano.

Para tratar do tema o restante deste artigo divide-se em quatro seções. Na primeira delas, discorre-se sobre os bancos de dados, as variáveis e o modelo estatístico utilizados no presente estudo. Na segunda, obtém-se, por meio de regressão longitudinal linear hierárquica, estimativa do tamanho dessa volatilidade nas escolas públicas mineiras avaliadas pelo Proeb. Na terceira, introduz-se o tema das metas do Ideb e investigam-se as possíveis consequências de seu cumprimento, em associação com a volatilidade apresentada anteriormente. Por último, apresentam-se algumas conclusões mais relevantes do presente estudo, tanto do ponto de vista dos resultados estatísticos quanto das políticas públicas a eles associadas.

OS RESULTADOS DA 8ª SÉRIE/9º ANO DO ENSINO FUNDAMENTAL

O presente estudo considerou apenas os resultados da 8ª série/9º ano do Ensino Fundamental (EF), para o qual se apresentam, a seguir, algumas estatísticas descritivas de proficiência divididas por ano de edição do Proeb/Simave. Os alunos avaliados no Proeb são classificados em três níveis de proficiência: Baixo (nível 1), Intermediário (nível 2) e Recomendado (nível 3).

TABELA 1 – Resultados de Língua Portuguesa para o 9º ano do EF por ano de edição do Proeb (2000-2011)

EDIÇÃO	Média	DP	% nível 1	% nível 2	% nível 3	N
2000	238,3	44,5	20,1	59,1	20,8	192.654
2002	241,9	39,2	12,3	69,1	18,6	213.986
2006	242,2	48,0	19,2	55,2	25,6	230.654
2007	239,7	46,4	20,1	57,0	23,0	233.786
2008	249,0	41,9	12,8	59,9	27,3	235.033
2009	249,9	44,2	13,5	57,0	29,6	241.204
2010	253,8	44,9	11,9	54,9	33,2	266.184
2011	252,1	46,0	13,8	53,7	32,5	257.181

Fonte: Elaboração própria.

TABELA 2 – Resultados de Matemática para o 9º ano do EF por ano de edição do Proeb (2000-2011)

EDIÇÃO	Média	DP	% nível 1	% nível 2	% nível 3	N
2000	229,1	46,6	47,4	45,8	6,8	194.990
2003	241,9	49,2	37,0	51,0	12,0	200.875
2006	245,5	51,2	33,4	52,5	14,1	224.140
2007	249,5	50,3	31,4	52,6	16,0	230.257
2008	253,9	50,3	28,7	53,2	18,1	230.164
2009	258,4	48,6	23,7	57,0	19,2	239.466
2010	266,4	47,6	18,8	56,7	24,4	230.986
2011	262,3	46,5	20,4	59,2	20,4	253.018

Fonte: Elaboração própria.

Os resultados indicam uma melhoria constante dos níveis de proficiência dos alunos participantes nas sucessivas edições do Proeb.

Perguntas iniciais do estudo multinível

Apresentamos, a seguir, algumas perguntas fundamentais que nos dão um ponto de partida para a realização deste estudo.

- 1) Há diferenças estatisticamente significativas de rendimento médio entre as escolas públicas de Minas Gerais no 9º ano do Ensino Fundamental, particularmente no ano mais recente de sua edição utilizado neste estudo (2011)?
- 2) Haverá, para as escolas do Simave, uma tendência significativa de variação das médias de proficiência ao longo do tempo? Em caso afirmativo, seria possível estimar seu valor?
- 3) No caso de haver tendências significativas de variações das médias de proficiência, as escolas do Simave se diferem significativamente entre si quanto a essas taxas de crescimento? Em outras palavras, haverá escolas cujo desempenho está crescendo mais rapidamente do que o de outras?

Em relação a essas perguntas, e acerca da metodologia que se empregará para se tentar respondê-las, cabe dizer que aqui se está interessado nos resultados “brutos” das escolas, ou seja, na evolução das suas respectivas médias de desempenho oficialmente informadas pelo Simave (e também pela Prova Brasil). Ora, essas médias de desempenho, por sua vez, são calculadas sem se adotarem quaisquer controles estatísticos para possíveis variáveis de interesse associadas ao desempenho educacional, de que é exemplo o índice socioeconômico (ISE) médio das escolas – como se vê, por exemplo, em BOWLES *et al.* (2008, p. 91), ou no seu respectivo tamanho, conforme mensurado pelo seu número de alunos (KANE *et al.*, 2002, p. 240).

Portanto, a adoção desses controles, embora adequada – ou mesmo necessária – para se avaliar com maior precisão o desempenho das escolas envolvidas no presente estudo, não foi aqui realizada devido ao fato de que o propósito da presente análise, como se disse, é especificamente considerar o comportamento *observado* dessas mesmas unidades escolares, conforme retratado nos boletins oficiais de resultados.

As variáveis

As seguintes variáveis, especificadas por níveis, serão utilizadas no trabalho:

Nível 1:

- 1) MÉDIA: proficiência média da escola em Língua Portuguesa ou Matemática mensurada em sua escala original (Saeb).¹

2) EDIÇÃO: ano da edição do Simave.

3) N: número de alunos avaliados por escola numa disciplina (LP ou MAT).

Nível 2:

1) REDE: rede de ensino (estadual ou municipal).

O modelo

Na modelagem do problema, utilizou-se um modelo de crescimento linear das médias que bem se adéqua à evolução do desempenho das escolas, conforme amiúde se adota na literatura internacional sobre estudos longitudinais hierárquicos (BRYK; RAUDENBUSH, 2002; SINGER; WILLET, 2003), e no qual o primeiro nível é o tempo (medido em anos e correspondente às diversas edições do Proeb) e o segundo nível é a escola. Esse modelo pode ser expresso pelas seguintes equações:

Nível 1:

$$Y_{it} = \pi_{0i} + \pi_{1i} a_{it} + e_{it}$$

onde:

Y_{it} : proficiência média (em Matemática ou Língua Portuguesa) da escola i no instante t .

π_{0i} : proficiência final média (ou seja, no ano 2011) da escola i .

π_{1i} : taxa de crescimento da escola i ao longo do período considerado.

a_{it} : momento da avaliação da escola i , onde t pode assumir valores correspondentes aos anos das diversas edições do Proeb, entre 2000 e 2011.

e_{it} : erro idiossincrático da média da escola i no momento t , sendo que os erros são independentes e possuem distribuição normal com variância comum σ^2 .

O que a última equação quer dizer é que a média Y_{it} de proficiência de uma escola i num instante t sofre variação temporal que pode ser descrita ao menos razoavelmente por uma linha reta cujo coeficiente linear π_{0i} é a média (prevista pelo modelo) da escola em 2011, e o coeficiente angular π_{1i} (ou inclinação da reta) corresponde à taxa média anual de variação da proficiência média da escola.

O coeficiente linear π_{0i} passou, no modelo, a representar o resultado médio da escola em 2011 mediante recodificação da variável EDIÇÃO, pela qual se obteve variável transformada de modo a ter o zero correspondente a 2011. Teria sido, aparentemente, mais natural adotar o zero, ou início da série, no ano de 2000, que correspondeu à primeira edição do Proeb. Entretanto, tal procedimento teria o inconveniente de não representar dado real para uma grande quantidade de escolas que não tomaram parte da primeira avaliação, sobretudo

devido ao fato de que muitas escolas cobertas pelo Proeb atualmente são municipais e, portanto, pertencentes a uma rede que não havia sido avaliada nas primeiras provas desse sistema (as redes municipais somente começaram a participar do Proeb a partir de 2006).

Quanto à inclinação da reta de regressão π_{1i} , ela representa a taxa de variação anual das médias de cada escola. Assim sendo, o modelo adotado permite que cada escola tenha a sua própria equação, cujos parâmetros foram calculados com base em processo de centralização das variáveis nos seus respectivos grupos, e não na grande média amostral. Essa centralização nos grupos é a mais indicada nesse caso, visto que se está justamente interessado em detectar possíveis diferenças entre as escolas quanto às suas respectivas taxas de evolução da proficiência.

Nível 2:

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

onde:

β_{00} : grande média das proficiências médias finais em 2011.

r_{0i} : erro idiossincrático da proficiência média final da escola i em 2011.

β_{10} : grande média das taxas de crescimento.

r_{1i} : erro idiossincrático da taxa de crescimento da escola i .

Resultados

Para o modelo linear especificado, os resultados encontrados foram os seguintes:

TABELA 3 – Estimativas dos efeitos aleatórios – Modelo linear, nulo no nível 2 – 9º ano do EF – Simave

Efeito aleatório	LP			MAT		
	DP	Componente da Variância	valor-p	DP	Componente da Variância	valor-p
<i>INTERCEPTO, r_0</i>	1.444.410	20.863.195	<0.001	1.680.818	28.251.502	<0.001
<i>INCLINAÇÃO, r_1</i>	116.504	135.733	<0.001	151.079	228.249	<0.001
<i>nível-1, e</i>	1.166.717	13.612.281		1.116.696	12.470.105	

Fonte: Elaboração própria.

Neste ponto, cabe lembrar que a resposta à primeira pergunta é fornecida na linha do efeito aleatório do intercepto (r_0), onde se informa que são extremamente baixos os valores p para a variância dos resultados do teste no nível das escolas. Em outras palavras, há uma variância não nula para as diferenças entre as proficiências médias das escolas públicas

de Minas Gerais, tanto para os resultados de Língua Portuguesa quanto para os de Matemática, o que implica algumas delas terem se saído significativamente melhor do que outras no teste de 2011.

Já a resposta para a segunda pergunta pode ser fornecida pelos resultados dos parâmetros fixos da regressão multinível, que são os seguintes:

TABELA 4 – Estimativas dos efeitos fixos – Modelo linear, nulo no nível 2 – 9º ano do EF – Simave

<i>Efeito fixo</i>	LÍNGUA PORTUGUESA		MATEMÁTICA	
	Coefficiente	<i>valor-p</i>	Coefficiente	<i>valor-p</i>
<i>Para INTERCEPTO1, π_0</i>				
<i>INTERCEPTO, β_{00}</i>	243.569.308	< 0.001	251.758.780	< 0.001
<i>Para INCLINAÇÃO, π_1</i>				
<i>INTERCEPTO2, β_{10}</i>	1.730.964	< 0.001	3.252.343	< 0.001

Fonte: Elaboração própria.

Observa-se, na última linha da tabela, para Língua Portuguesa, que a média das inclinações vale aproximadamente 1,7 e que esse coeficiente é significativamente diferente de zero, devido ao seu baixíssimo valor *p*. Dessa forma, conclui-se que as escolas no Simave estão, em média, acrescentando quase dois pontos à sua respectiva média esperada de proficiência em Língua Portuguesa a cada edição anual do teste. Por sua vez, a grande média estadual corresponde, segundo esse modelo, à aproximadamente 243 pontos na escala de proficiência do Saeb. Em Matemática, a cada ano, as escolas ganham, em média, 3,2 pontos. Entretanto, vale lembrar que os valores de Língua Portuguesa e Matemática estão expressos em escalas de proficiência diferentes, uma para cada disciplina, de modo que os números, ao menos na forma como aqui estão, não são diretamente comparáveis entre si.

Por sua vez, a resposta para a terceira pergunta pode ser encontrada também na tabela das variâncias apresentada (TABELA 3). É possível ver, na linha do parâmetro de inclinação, que a resposta a essa pergunta é positiva, tanto para Língua Portuguesa quanto para Matemática: as escolas do Simave diferem-se umas das outras quanto à taxa de evolução de suas respectivas médias de Língua Portuguesa, visto que, novamente, é aproximadamente nulo o valor *p* para a variância dessas taxas no nível das escolas.

Assim, conclui-se que as escolas diferiram significativamente em suas médias de proficiência em 2011 – esses resultados, de fato, podem ser observados para todos os anos. Além disso, conclui-se que as escolas crescem, em média, nos seus níveis de proficiências;

contudo, esse crescimento não é o mesmo para todas as escolas: pode haver aquelas que não apresentam crescimento ou mesmo decrescem ao longo do tempo.

ANÁLISE DA VOLATILIDADE DOS RESULTADOS DAS ESCOLAS AO LONGO DAS EDIÇÕES DO PROEB

Ao se construir o modelo multinível apresentado na seção anterior, foi possível obter os resíduos do modelo, que correspondem às distâncias entre os valores previstos e os respectivos valores reais. Dessa forma, no nível 1 (o do tempo), calcularam-se: resíduos indicando a distância entre a nota média obtida pela escola i no tempo t , e o seu respectivo valor previsto pelo modelo linear com a variável REDE modelando o nível 2, da escola. Nesse segundo nível, por sua vez, os resíduos corresponderam à diferença entre as médias das escolas na avaliação do Simave de 2011 (a edição mais recente da série) e à grande média das escolas nesse mesmo teste.

Um resíduo pode ser positivo, nulo ou negativo. É positivo quando o valor real se encontra acima do valor previsto. Nesse caso, pode-se dizer que o modelo subestimou a nota média da escola, prevendo para ela valor menor do que o que realmente se obteve. Inversamente, um resíduo pode ser negativo quando o valor real se encontra abaixo do valor previsto. Nesse caso, houve superestimação da média da escola. E, por fim, naturalmente, o resíduo é zero quando os valores real e previsto da média escolar são coincidentes, um caso de ajuste perfeito do modelo. Cabe dizer que, nos modelos empregados, há a suposição de que os resíduos se concentrem ao redor da linha de previsão segundo uma distribuição normal e de variância constante ao longo de sua própria escala. Esses são, respectivamente, os pressupostos da normalidade e da homocedasticidade herdados da teoria clássica da testagem, e observados nos modelos lineares hierárquicos, não no nível global da amostra, mas no nível de cada unidade de interesse (nesse caso, no nível de cada escola).²

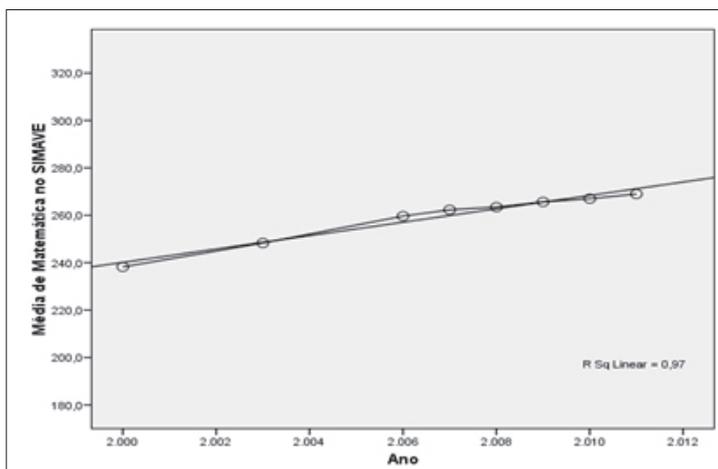
O desvio-padrão dos resíduos

A fim de sintetizar uma estimativa da oscilação de cada escola em torno da sua média prevista, foi calculado o desvio-padrão dos resíduos de cada uma delas. Dessa forma, tem-se medida que, *grosso modo*, significa a quantidade de pontos na escala de proficiência em que a escola em média varia, para mais ou para menos, já descontados os efeitos da sua evolução temporal e também da rede de ensino à qual ela pertence.

O desvio-padrão dos resíduos, nesse caso, é uma medida natural da precisão das estimativas. Menores valores de desvios-padrão

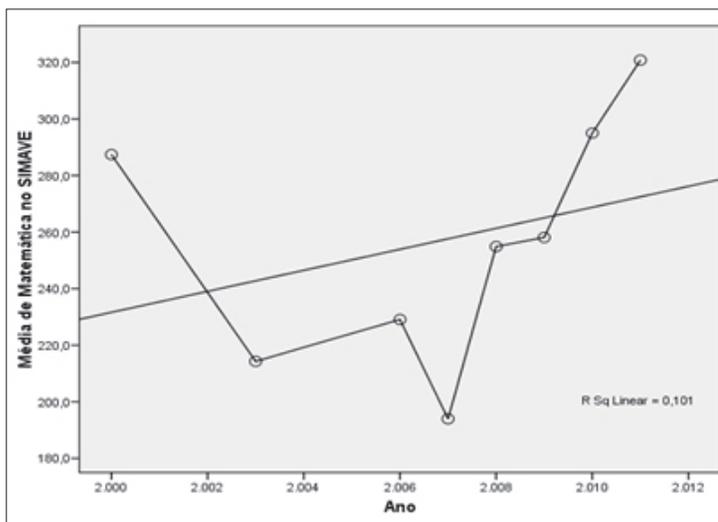
indicam a existência de menores resíduos e, conseqüentemente, de maior precisão do modelo. Também se pode dizer que os maiores desvios-padrão indicam maior volatilidade dos resultados. Os três gráficos a seguir mostram o exemplo de três escolas com baixa e alta volatilidade de resultados.

GRÁFICO 1 – Escola com baixa volatilidade nas médias de Matemática e um pequeno desvio no início da série temporal (Simave – 2000/2011 – 9º ano do EF)



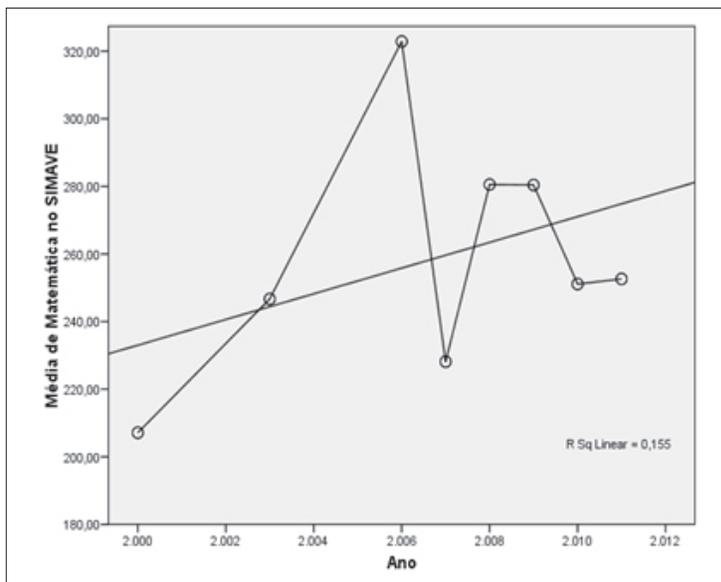
Fonte: Elaboração própria.

GRÁFICO 2 – Escola com alta volatilidade nas médias de Matemática e um acentuado desvio positivo no início da série temporal (Simave – 2000/2011 – 9º ano do EF)



Fonte: Elaboração própria.

GRÁFICO 3 – Escola com alta volatilidade nas médias de Matemática e um acentuado desvio negativo no início da série temporal (Simave – 2000/2011 – 9º ano do EF)



Fonte: Elaboração própria.

Os dados descritivos a seguir referem-se somente às escolas que, no decorrer da história do Simave, foram testadas, ao menos, três vezes.

TABELA 5 – Desvios-padrão dos resíduos das médias escolares de proficiência em Matemática no Simave

		MATEMÁTICA	LÍNGUA PORTUGUESA
<i>N</i>	<i>Válidos</i>	4.249	4.252
	<i>Faltantes</i>	0	0
<i>Média</i>		14,7	12,0
<i>DP</i>		7,6	6,3
<i>Mínimo</i>		1,6	0,8
<i>Máximo</i>		68,2	78,0
<i>Percentis</i>	25	9,7	7,8
	50	13,0	10,4
	75	17,7	14,5

Fonte: Elaboração própria.

Os dados mostram que as escolas públicas mineiras, no 9º do EF, variam aleatoriamente, em média, 12 pontos para mais ou para menos na escala de Língua Portuguesa em relação à sua respectiva média prevista pela evolução temporal de seus próprios resultados. Em Matemática, essa variação média chega a quase 15 pontos da escala de proficiência.

Uma categorização de escolas quanto ao tamanho

A observação da distribuição das escolas quanto ao número de alunos avaliados permite-nos, para fins de descrição e análise dos dados, dividir convencionalmente essa escala em três níveis, que são os apresentados na Tabela 6.

TABELA 6 – Tipo de escola quanto ao número médio anual de alunos avaliados no Simave, no período 2000-2011 – 9º do EF

Escola	Nº médio de alunos avaliados	N	%	% cumul.
<i>Pequena</i>	Até 40	1.750	41,2	41,2
<i>Média</i>	Entre 40 e 80	1.297	30,5	71,7
<i>Grande</i>	Acima de 80	1.205	28,3	100
<i>Total</i>		4.252	100	

Fonte: Elaboração própria.

Cabe investigar como a volatilidade dos resultados se apresenta para os diferentes tamanhos de escolas. Para isso, para cada um dos três grupos de escolas – pequenas, médias e grandes –, obteve-se a média dos seus respectivos desvios-padrão dos resíduos, que correspondem à medida de volatilidade aqui considerada. Os resultados foram:

TABELA 7 – Média dos desvios-padrão dos resíduos por tamanho de escola no Simave, no período 2000-2011 – 9º do EF

Escola	Média dos desvios-padrão dos resíduos	
	MATEMÁTICA	LÍNGUA PORTUGUESA
<i>Pequena</i>	18,4	15,6
<i>Média</i>	11,8	9,5
<i>Grande</i>	12,1	9,3

Fonte: Elaboração própria.

Observa-se, com os resultados apresentados, que, em Língua Portuguesa, as escolas pequenas no Simave têm uma variação aleatória média superior a 15 pontos da escala de proficiência em relação à sua trajetória de crescimento prevista pelo modelo. Já entre as escolas médias e grandes o efeito é aproximadamente o mesmo, em média, correspondendo a cerca de nove pontos na escala. Em Matemática, ocorre algo semelhante, no sentido em que as menores escolas apresentam as maiores flutuações – 18 pontos em média –, ao passo que as escolas de porte médio e grande tiveram flutuação média próxima de 12 pontos da sua respectiva escala.

Questões a serem levantadas sobre a volatilidade encontrada

Vimos que, mesmo levando em conta a evolução temporal das médias das escolas – por meio de modelagem linear hierárquica longitudinal que considera a evolução de cada escola individualmente, e não apenas tendência geral de crescimento –, ainda assim, os resíduos que se observam nas médias escolares são consideravelmente elevados, sendo da ordem de uma dezena de pontos em média, em qualquer das escalas de proficiências consideradas (Matemática e Língua Portuguesa). E tal fato é, como já se devia esperar, mais acentuado nas escolas menores (com 40 alunos ou menos, em média).

Em relação às escolas de porte médio e grande, é possível tirar duas outras conclusões: o efeito da volatilidade ainda existe para elas e está longe de ser desprezível; e, ao menos em geral, não há grandes diferenças quanto a esse efeito entre as escolas médias e as grandes. Na próxima seção, serão discutidas as implicações dessa volatilidade no estabelecimento das metas do Ideb.

A VOLATILIDADE DOS RESULTADOS E AS METAS DO IDEB

As metas do Ideb são determinadas com base num modelo de regressão logística que leva em conta os resultados de fluxo e proficiência das escolas no ano 2005, tendo essa proficiência média sido mensurada pela Prova Brasil do mesmo ano. Segundo esse modelo, às escolas que apresentaram desempenho médio menor em 2005 foram apresentadas metas menores para 2007, devido ao fato de que, nesse caso, a curva de regressão “passava”, em 2005, por um ponto mais baixo quando comparado às escolas que, na mesma edição do teste, tiveram resultados melhores (BRASIL, s/d.b).

Para realizar o estudo desta seção, que associa as metas das escolas no Ideb às flutuações em relação aos seus resultados previstos em 2005, foi montado novo banco de dados que parcialmente

correspondeu ao mesmo utilizado nas análises descritas até aqui. Entretanto, nele foram inseridos os resultados das escolas na Prova Brasil de 2005 (lembrando que, neste ano, não houve avaliação do PROEB em nenhuma das duas disciplinas consideradas), além dos resultados, também das escolas na Prova Brasil de 2007 e 2009, que substituíram os resultados do Proeb para esses mesmos anos. Essa mescla de resultados, a rigor, não representou grande problema, devido ao fato de que os testes do Proeb têm seus resultados calibrados, precisamente, segundo os parâmetros das escalas da Prova Brasil. Por outro lado, a introdução das notas da Prova Brasil permitiu que se traçasse uma linha evolutiva da média das escolas levando em conta todos os resultados já disponíveis para essa avaliação, acrescidos dos dados das mesmas escolas recolhidos por sistema de avaliação praticamente análogo ao federal.

Cabe ressaltar que essa mescla de resultados do Proeb com a Prova Brasil resultou na grande vantagem de se aumentar significativamente, para uma mesma escola, o número de observações (ou seja, de testes), visto que tanto a primeira quanto a segunda encontram-se em mesma escala. E, assim, foi possível aplicar a medida da volatilidade das médias escolares (obtida em grande parte com os resultados do Proeb) no problema da determinação das metas do Ideb, que se faz a partir dos dados da Prova Brasil.

Com tudo isso, foi possível rodar o mesmo modelo longitudinal de crescimento linear, e com a variável REDE no nível 2 – das escolas –, o que possibilitou a obtenção dos resíduos das escolas em cada edição da Prova Brasil ou do Proeb. Mais particularmente, nesse caso, nosso interesse se concentrou nos resíduos das médias escolares na Prova Brasil de 2005, devido ao papel decisivo que as médias das escolas no referido ano tiveram sobre o cálculo das suas respectivas metas do Ideb. O raciocínio por trás desse procedimento é simples e pode ser expresso por meio da seguinte pergunta:

- Considerando que as escolas públicas mineiras apresentam desvios aleatórios significativos entre suas médias reais e as suas respectivas médias previstas pelo modelo de crescimento longitudinal, haverá associação significativa entre esses desvios observados em 2005 – ano de fixação das metas do Ideb – e o fato de as escolas atingirem ou não suas metas do Ideb em anos posteriores?

O cumprimento das metas do Ideb em 2007 e 2009

A tabela a seguir mostra o percentual das escolas consideradas na amostra que se encontraram abaixo de suas metas para o Ideb nos anos de 2007 e 2009.

TABELA 8 – Escolas públicas de Minas Gerais quanto ao cumprimento da meta do Ideb em 2007 e 2009

Ano		Abaixo da meta	Na meta ou acima	Totais válidos	Faltantes
2007	<i>N</i>	1.856	818	2.674	31
	<i>% vál.</i>	69,4	30,6	100	
2009	<i>N</i>	1990	608	2.598	107
	<i>% vál.</i>	76,6	23,4	100	
<i>Ambos</i>	<i>N</i>	2213	370	2.583	122
	<i>% vál.</i>	85,7	14,3	100	

Fonte: Elaboração própria.

Observa-se, portanto, que cerca de 30% das escolas públicas mineiras deixaram de atingir a meta do Ideb para 2007, no 9º do EF, e que tal proporção reduziu-se para pouco menos de um quarto na avaliação seguinte – a de 2009. Por outro lado, aproximadamente uma em cada sete escolas não atingiu a meta em nenhuma das duas vezes.

A divisão das escolas segundo o erro médio em 2005: diferentes pontos de partida para o estabelecimento das metas

Conforme vimos, a distribuição das médias reais de proficiência das escolas é grande o suficiente para permitir a ocorrência relativamente frequente de grandes desvios em relação aos valores previstos. Entretanto, em relação especificamente ao Ideb, cabe comentar que os desvios de Matemática e de Língua Portuguesa não podem ser analisados separadamente, mas sim em conjunto, visto que ambos entram simultaneamente no cálculo das variações desse índice. Uma proposta aparentemente válida seria a de se tirar a média dos desvios nessas duas disciplinas. Assim, por exemplo, se uma escola teve determinado resíduo positivo na nota de Língua Portuguesa e resíduo negativo em Matemática, o efeito médio dessas variações sobre o Ideb seria atenuado, pois tais flutuações estariam ocorrendo em sentidos opostos. Entretanto, cabe dizer que esse caso *não* é o mais frequente. O que se percebe, em geral, é uma elevada correlação *positiva* entre os resultados das duas disciplinas, de modo que as escolas que obtêm médias maiores em Matemática também obtêm, em geral, médias maiores em Língua Portuguesa,

e algo semelhante se passa com as escolas de desempenho mais baixo. Portanto, poder-se-ia usar a média simples dos dois resíduos para exprimir o resíduo médio de cada escola. Esse, de fato, foi o procedimento aqui adotado, com somente uma modificação: no nosso caso, calculou-se, na verdade, a *média padronizada dos resíduos padronizados*, tendo a padronização dos resíduos sido feita devido ao fato de que as escalas das duas disciplinas são diferentes entre si.

A seguir, a população de escolas é dividida em quatro grupos de mesmo tamanho e dispostos em ordem ascendente de resíduos. Dessa forma, os grupos 1 e 2, chamados respectivamente de “Negativo Acentuado” e “Negativo”, correspondem às escolas que tiraram nota média real abaixo daquela prevista para elas em 2005. “Negativo Acentuado” corresponde, assim, ao primeiro quarto da distribuição. Nos referidos grupos, portanto, as escolas ficaram muito abaixo do que deveriam segundo o modelo de crescimento aqui proposto. Por outro lado, há os grupos 3 e 4, que correspondem respectivamente aos resíduos “Positivo” e “Positivo Acentuado”, indicativos de escolas cujas médias em 2005 ficaram acima dos seus respectivos valores previstos, correspondendo o último grupo à parte superior da distribuição de resíduos.

TABELA 9 – Tipos de escolas públicas mineiras quanto aos resíduos de sua média calculados em relação aos valores previstos de suas respectivas médias na Prova Brasil de 2005

Resíduo	Limites do desvio médio	Número de escolas
<i>Negativo Acentuado</i>	Mínimo e 1º quartil	676
<i>Negativo</i>	1º e 2º quartis	677
<i>Positivo</i>	2º e 3º quartis	676
<i>Positivo Acentuado</i>	3º quartil e máximo	676

Fonte: Elaboração própria.

A associação entre os resíduos e o cumprimento das metas (I)

A tabela a seguir indica o percentual de escolas que não conseguiram atingir suas respectivas metas do Ideb associado ao tipo de resíduo de desempenho. Há colunas para os anos de 2007 e 2009, separadamente, e para o caso do não cumprimento verificado em ambos os anos.

TABELA 10 – Percentual das escolas que não atingiram a meta do Ideb em associação ao tipo de resíduo de suas médias na Prova Brasil de 2005

Resíduo	2007 (%)	2009 (%)	Ambos os anos (%)
<i>Negativo Acentuado</i>	14,2	9,1	3,0
<i>Negativo</i>	22,1	14,3	7,2
<i>Positivo</i>	32,3	24,5	12,5
<i>Positivo Acentuado</i>	53,7	45,5	34,5

Fonte: Elaboração própria.

Observa-se, por exemplo, que, se do ponto de vista geral, 30 % das escolas haviam falhado em atingir suas metas em 2007, ocorreram, por outro lado, acentuadas diferenças de desempenho observadas quando essa informação é cruzada com o desvio sofrido pela média da escola em 2005, ano de estabelecimento das metas. Das escolas que tiveram resíduo negativo acentuado, o percentual das que não atingiram a meta em 2007 foi consideravelmente baixo quando comparado com o dos demais grupos: cerca de uma em cada sete escolas desse grupo não alcançou suas metas em 2007. Entretanto, na mesma edição da Prova Brasil, o grupo das escolas que mais havia crescido em 2005, em relação às suas médias previstas, foi aquele em que se verificaram taxas altíssimas de não cumprimento das metas: mais da metade das escolas desse grupo não atingiram o Ideb para elas estipulado em 2007.

Algo semelhante, embora um pouco menos acentuado em termos de percentuais de falhas em atingir o Ideb, se passou em 2009 – e também no caso do não cumprimento das metas em ambos os anos. Em todos esses casos, os dados parecem indicar evidências muito fortes de que uma parte considerável do sucesso das escolas em alcançar as metas do Ideb *não* tem dependido apenas de variáveis sob o controle da escola, mas, devido a flutuações estatísticas ocorridas no momento em que as metas foram estabelecidas, são em parte condicionadas ao tamanho das escolas.

A associação entre os resíduos e o cumprimento das metas (II)

Finalizando a análise, é possível precisar ainda mais o impacto do resíduo das médias observadas sobre o resultado do Ideb por meio de regressão logística, onde a variável dependente é o logaritmo da chance de cada escola *não* alcançar a meta para o Ideb e a variável independente é, novamente, a média dos resíduos em ambas as disciplinas em 2005 por escola. A tabela a seguir resume os principais resultados dessa regressão.

TABELA 11 – Efeito do resíduo das médias da escola sobre a chance de ela não conseguir atingir seu Ideb

ANO	COEF.	B	Valor-p	Exp (B)
2007	INCLIN.	0,829144	,000	2,291357
	INTCPT	-0,93694	,000	0,391824
2009	INCLIN.	0,936902	,000	2,552063
	INTCPT	-1,39309	,000	0,248306
Ambos	INCLIN.	1,191061	,000	3,29057
	INTCPT	-2,21628	,000	0,109014

Fonte: Elaboração própria.

Observa-se, pela Tabela 11, que todos os coeficientes da regressão são significativos. Uma informação particularmente relevante encontra-se na última coluna – Exp (B) – e, nas linhas das inclinações, significa a variação nas chances de uma escola *não* atingir o Ideb associada a um aumento de desvio-padrão na variável do desvio entre as notas reais e as previstas em 2005 na Prova Brasil. Dessa forma, percebe-se que, para cada desvio-padrão a mais que a escola avança quanto à volatilidade de seus resultados médios, suas chances de *não* atingir a meta do Ideb em 2007 aumentaram cerca de 2,3 vezes. Esse efeito ainda subiu ligeiramente em 2009 – para aproximadamente 2,6 – e chegou a cerca de 3,3 vezes, nos casos das escolas que não atingiram o Ideb em nenhuma das edições.

CONCLUSÕES

As escolas públicas de Minas Gerais apresentam considerável volatilidade em torno de suas tendências de evolução média, quando se consideram os seus resultados nos testes de proficiência em grande escala aplicados atualmente (Proeb/Simave e Prova Brasil), tanto na disciplina de Língua Portuguesa quanto de Matemática. Essa volatilidade é mais acentuada nas escolas pequenas; mas, mesmo nas grandes, é bastante apreciável. As escolas vêm, paulatinamente, aumentando suas médias de proficiência ao longo das edições desses testes. Mesmo quando se controla a média das escolas pela sua evolução temporal e pela rede à qual elas pertencem, sobra ainda grande quantidade de variação do resultado que se associa a fatores não explicados (não incluídos no modelo). A média dessa flutuação dos resultados das escolas alcança, em geral, dois dígitos nas escalas de proficiência de Língua Portuguesa

e Matemática utilizadas. Ao mesmo tempo, observa-se essa volatilidade ser maior nas escolas menores (com até 40 alunos avaliados por ano, em média), embora continue sendo significativa e aproximadamente igual para as escolas de porte médio e grande.

Um problema decorrente da excessiva volatilidade é que ela pode comprometer a validade do acompanhamento anual do desempenho das escolas, quando este é realizado, ao menos parcialmente, com base nas médias de proficiência alcançadas pela escola apenas nas últimas avaliações. Outro problema decorre do fato de que essa volatilidade pode ter consequências consideráveis e indesejadas nas situações em que se fixam metas de desempenho para sistemas ou unidades educacionais com base nos resultados de um ano específico de avaliação em grande escala. Nesse sentido, os erros ou resíduos presentes podem provocar considerável subestimação ou superestimação do desempenho médio de uma escola.

Decorre dos resultados aqui encontrados que o estabelecimento de metas precisa levar em consideração a volatilidade das médias de proficiências e uma janela temporal adequada, de tal forma que a meta não seja construída apenas com base na última avaliação. Proposta para o estabelecimento de metas que leve em consideração a trajetória de resultados da escola está sendo construída e será discutida em novo trabalho dos autores.

REFERÊNCIAS

- BOWLES, S. *et al.* Promessas quebradas: reforma da escolar em perspectiva. In: BROOKE, N.; SOARES, J. F. (Org.). **Pesquisa em eficácia escolar: origem e trajetórias**. Belo Horizonte: Editora UFMG, 2008. p. 91.
- BRASIL. Ministério da Educação – MEC. Instituto Nacional de Pesquisas Educacionais Anísio Teixeira – Inep. **Nota técnica n. 1: índice de desenvolvimento da educação básica – Ideb**. Brasília, s/d.a. Disponível em: <http://download.inep.gov.br/educacao_basica/portal_ideb/o_que_e_o_ideb/Nota_Tecnica_n1_concepcaoIDEB.pdf>. Acesso em: 18 set. 2010.
- BRASIL. Ministério da Educação – MEC. Instituto Nacional de Pesquisas Educacionais Anísio Teixeira – Inep. **Nota técnica n. 2: metodologia utilizada para o estabelecimento das metas intermediárias para a trajetória do Ideb no Brasil, Estados, Municípios e Escolas**. Brasília, s/d.b. Disponível em: <http://download.inep.gov.br/download/Ideb/Nota_Tecnica_n2_metas_intermediarias_IDEB.pdf>. Acesso em: 18 mar. 2012.
- BRYK, A. S.; RAUDENBUSH, S. W. **Hierarchical linear models: applications and data analysis methods**. Newbury Park: SAGE, 2002.
- KANE, T. *et al.* Volatility in school test scores: implications for test-based accountability systems. **Brookings Papers on Education Policy**, Washington D. C., n. 5, p. 235-283, 2002. Disponível em: <<http://www.jstor.org/discover/10.2307/20067246?sid=21105466072533&uid=3737664&uid=4&uid=2>>. Acesso em: 24 fev. 2017.

O'CONNELL, A. A.; MCCOACH, D. B. (Ed.) **Multilevel modeling of educational data**. Charlotte: Information Age Publishing, 2008.

PONTES, L. A. F. **Medidas de eficácia escolar no contexto das políticas brasileiras de responsabilização educacional**: o caso do Índice de Desenvolvimento da Educação Básica, o Ideb, em Minas Gerais. 2015. 242 f. Tese (Doutorado em Educação) – Faculdade de Educação da UFJE, Universidade Federal de Juiz de Fora, Juiz de Fora. 2015.

SINGER, J. D.; WILLETT, J. B. **Applied longitudinal data analysis**: modeling change and event occurrence. New York: Oxford University Press, 2003.

SOARES, J. F.; XAVIER, F. P. Pressupostos educacionais e estatísticos do Ideb. **Educação & Sociedade**, Campinas, v. 34, n. 124, p. 903-923, jul./set. 2013.

NOTAS

¹ Em geral, variáveis quantitativas contínuas, como a proficiência nos testes, costumam ser padronizadas para se rodar uma regressão, visto que tal procedimento tem a vantagem de unificar a interpretação dos coeficientes das diferentes variáveis. Entretanto, no presente caso, como há interesse específico pela própria interpretação das escalas do Simave, pensamos mais conveniente rodar a regressão nas escalas originais, pois isso favorecerá, assim esperamos, a compreensão dos resultados por parte de quem já esteja habituado a trabalhar com resultados do Simave e da Prova Brasil (dois testes que adotam as mesmas escalas de Matemática e Língua Portuguesa).

² Cf. O'CONNELL; MCCOACH, 2008.

Submetido: 14/08/2015

Aprovado: 19/01/2016

Contato:

Luís A. F. Pontes

Rua Francisco Vaz de Magalhães, nº 820 apt.202,

Juiz de Fora | MG | Brasil

CEP 36.033-340