

Uma Análise de Rankings de Escolas Brasileiras com Dados do SAEB[♦]

Ana Maria Paiva Franco

Professora - Universidade Federal de Uberlândia
Endereço para contato: Av. João Naves Ávila, 2121 - Uberlândia - MG - CEP: 38408-100
E-mail: anafranco@ie.ufu.br

Naercio Menezes-Filho

Professor (FEA/USP) e Instituto de Ensino e Pesquisa (INSPER)
Endereço para contato: Rua Quatá, 300 - CEP: 04546-042 - São Paulo
E-mail: naercioamf@insper.edu.br

Recebido em 26 de abril de 2010. Aceito em 24 de fevereiro de 2012.

Resumo

Os elaboradores de políticas educacionais têm dado enfoque às políticas de “responsabilização” como forma de melhorar a qualidade da educação. Com base no desempenho dos alunos em testes padronizados, constroem-se medidas para classificação das escolas e estabelecem-se metas para guiar políticas de incentivo, assistência e sanção às escolas, como no caso do Chile e dos EUA. Em alguns casos, pode ser difícil obter rankings confiáveis, devido à elevada estratificação social no sistema educacional e ao problema de reversão para a média, que causam ruído nas estimativas. Neste trabalho foram analisadas diferentes medidas utilizadas para se construir rankings de escolas a partir das avaliações feitas pelo SAEB na 4ª série do Ensino Fundamental. Uma das medidas analisadas parece atender aos requisitos de ser estável e, ao mesmo tempo, não refletir principalmente o status socioeconômico dos alunos.

Palavras-Chave

educação, desempenho, classificação, incentivos

Abstract

Educational policymakers have been focusing on accountability policies as a way to improve the educational quality. Based on the students' performance in standardized tests, measures for schools' classification are built and goals are established in order to guide incentive policies, assistance and sanctions on schools, as is being done in Chile and USA. In some cases, it may be difficult to obtain trustful school's rankings due to a high social stratification in the educational system and the mean reversion problem that can cause noise on the estimates. In this work different measures used to construct school rankings were analyzed using SAEB's data for the 4th grade of Primary Education.

[♦] A autora Ana Maria agradece o apoio financeiro do CNPq para a realização deste trabalho, que é parte de sua tese de doutorado.

One of the analyzed measures seems to fulfill the requirement of being stable and, at the same time, not to reflect mainly the students' socioeconomic status.

Keywords

education, performance, rankings, incentives

JEL Classification

I20, I21, I28

1. Introdução

Existe uma preocupação crescente com a “responsabilidade” atribuída à escola e aos educadores no processo de aprendizado. O Chile e os Estados Unidos são exemplos de países em que o desempenho das escolas em testes padronizados é utilizado pelo Governo para nortear a alocação de recursos, recompensas, sanções e assistência às escolas. Também o Brasil caminha no sentido de desenhar mecanismos de incentivos que tomam como base os resultados que emergem de avaliações em testes padronizados para nortear políticas de “responsabilização”. Recentemente, porém, alguns artigos chamaram a atenção para a existência de ruídos nas informações sobre qualidade das escolas a partir do desempenho dos alunos em testes padronizados, que podem prejudicar o correto entendimento do significado dos rankings e índices de desempenho (ver, por exemplo, Mcewan e Urquiola, 2005; Mizala *et al.*, 2007).

No Brasil há uma preocupação por parte da sociedade e dos elaboradores de políticas educacionais de melhorar a qualidade do ensino através da atribuição de “metas” educacionais a serem alcançadas pelas escolas. Isto exige que indicadores confiáveis de desempenho sirvam de parâmetro para as políticas de incentivo, quando o objetivo é premiar, punir ou auxiliar aquelas que são ou não capazes de atingir o desempenho esperado. Para a implementação de tais políticas, têm sido propostos alguns indicadores de qualidade educacional com base no desempenho dos alunos em exames padronizados, como o Sistema Nacional de Avaliação da Educação Básica (SAEB), a Prova Brasil e os resultados do programa internacional de avaliação de alunos conhecido como PISA (*Programme for International Student Assessment*).

O Índice de Desenvolvimento da Educação Básica (IDEB), proposto pelo Instituto Nacional de Pesquisas e Estudos Educacionais “Anísio

Teixeira” (INEP), é um exemplo de indicador de qualidade educacional obtido por uma combinação de informações sobre desempenho em exames padronizados (Prova Brasil ou SAEB, ao final da 4^a e 8^a séries do Ensino Fundamental e da 3^a série do Ensino Médio) e de rendimento escolar (nível de aprovação nas escolas). O IDEB pode servir para o Governo monitorar as redes de ensino e exigir contrapartidas de desempenho das escolas, em termos de um avanço esperado no indicador, quando da liberalização de recursos (para maiores detalhes, ver INEP, 2008 e Fernandes, 2008).

O Governo do Estado de São Paulo também tem buscado meios de aperfeiçoar o monitoramento das escolas estaduais paulistas. A partir da definição de uma métrica nos moldes do IDEB, foram estabelecidas metas para cada escola a serem atingidas pelas escolas ao longo dos anos. O sucesso das escolas no cumprimento das metas serve de parâmetro para a concessão de incentivos a todos os funcionários das escolas. A sociedade civil também tem se manifestado a respeito, como o movimento da sociedade civil denominado “Compromisso Todos Pela Educação”, que propõe valores de desempenho escolar acima dos quais devem estar 70% dos alunos brasileiros até 2022 (Menezes-Filho *et al.*, 2008).

É questão já amplamente conhecida por pesquisadores e estudiosos da área que os rankings puros (que tomam por base o desempenho médio dos alunos nos testes padronizados) refletem principalmente o status socioeconômico (SSE) dos alunos das escolas. Por isso, tal tipo de ranking não é adequado para nortear políticas de incentivos e responsabilização, pois, neste caso, diretores e professores estariam sendo recompensados (ou “punidos”) pelo tipo de aluno que servem, e não pelo quanto conseguem fazer com que aprendam mais. Também não é desejável que haja muita volatilidade nos rankings de uma avaliação para outra, uma vez que isso confunde os agentes e os fazem rever a cada período as suas escolhas e políticas.

Alguns estudos para o caso do Chile apontam novas questões quanto à confiabilidade dos resultados obtidos com diferentes medidas utilizadas para a construção de rankings de escolas, inclusive as que contam com controles para o status socioeconômico (SSE) dos alunos. Suas advertências serviram de estímulo para que neste trabalho se buscasse averiguar se os resultados encontrados para o Chile se estendem para o caso do sistema educacional brasileiro. É preciso

saber até que ponto os rankings de escolas que resultam de avaliações como o SAEB ou Prova Brasil, após ser feito ou não o controle por características socioeconômicas dos alunos, são confiáveis no sentido de servirem de parâmetro para pais, educadores e formuladores de política sobre quais escolas realmente agregam mais valor ao aprendizado.

Com este propósito, neste artigo construímos cinco medidas de desempenho para ordenar escolas, tendo como base o estudo para o caso chileno (Mizala *et al.*, 2007). Observamos quais dessas medidas não refletem majoritariamente o nível socioeconômico dos alunos e, ao mesmo tempo, não resultam em alterações muito acentuadas nas ordenações de um ano para o outro (o que aqui se chama de volatilidade). O que podemos observar é que, entre as cinco medidas testadas, apenas uma delas apresenta essas duas características simultaneamente e, portanto, poderia servir de referência para políticas de monitoramento da qualidade da educação nas escolas.

Além desta introdução, o trabalho está organizado da seguinte forma: a seção 2 discorre sobre os desafios de se mensurar a qualidade das escolas empiricamente; a seção 3 apresenta os dados utilizados neste trabalho e algumas estatísticas descritivas; a seção 4 apresenta a metodologia utilizada; a seção 5 traz os resultados, enquanto a seção 6 apresenta os comentários finais.

2. Desafios para Mensurar a Qualidade das Escolas

Para se estimar melhor a qualidade ou o valor adicionado pelas escolas, geralmente acompanha-se o progresso do aluno ao longo de vários anos, e utilizam-se informações sobre seu status socioeconômico (SSE) e outras informações disponíveis com respeito, por exemplo, à escola em que estuda.

Os desafios neste campo de investigação são muitos. Primeiramente, a maioria dos países em desenvolvimento não conta com um sistema de avaliação da qualidade na educação que permita acompanhar o progresso de cada aluno ao longo dos anos de escola. Além disso, muitas das variáveis de famílias, alunos e mesmo escolas não são observadas pelos pesquisadores. Esta omissão tende a viesar as estimativas do efeito-escola. Outro problema constatado, por exemplo,

por Mcewan e Urquiola (2005), é que o controle por SSE, principalmente num sistema de escolas com elevada estratificação, vem ao custo de introduzir muita volatilidade nos rankings de notas das escolas.

Segundo Mizala *et al.* (2007) existe, pelo menos no caso do Chile, um *trade-off* entre construir rankings de escolas que não espelhem majoritariamente a condição socioeconômica de seus alunos e, por outro lado, não apresentem muita volatilidade de um ano ao outro. Rankings que espelham majoritariamente o SSE de alunos não são adequados para se nortear políticas de incentivos e responsabilização, pois, neste caso, diretores e professores estariam sendo recompensados (ou “punidos”) pelo tipo de aluno que servem, e não pelo quanto conseguem fazer com que aprendam mais. Uma alta volatilidade nos rankings de uma avaliação a outra também não é desejável, uma vez que confunde os agentes e os fazem rever a cada período as suas escolhas e políticas. Por exemplo, qualquer mecanismo de incentivo que os elaboradores de política de educação programem, tomando como referência rankings voláteis, acabam tornando-se ineficazes ou inócuos, pois logo os agentes deixam de responder a recompensas e punições cujas distribuições se aproximem daquelas geradas por um processo de “loteria”.

A volatilidade pode ocorrer devido a eventos em apenas um momento do tempo, como o advento de doenças que atingem toda escola, distração de alunos e responsáveis no dia de realização das provas e assim por diante. Outra razão para volatilidade é a variação nas amostras. Cada corte de alunos que entra na escola é como uma amostra aleatória de uma população local e, portanto, a média de desempenho da escola vai variar de acordo com o grupo específico de alunos que começou a estudar num determinado ano. A sua variância, por sua vez, dependerá da variabilidade de desempenho na população da qual a escola recebe os alunos, e também do número de alunos avaliados (ver Chay *et al.*, 2005, p. 1242-1244).

3. Dados e Estatísticas Descritivas

Os dados utilizados nas análises que se seguem provêm do SAEB. O SAEB é um levantamento bianual realizado pelo INEP desde 1993, que avalia os alunos da 4^a e 8^a séries do Ensino Fundamental, e da 3^a série do Ensino Médio, nas disciplinas de Matemática e Língua Portuguesa. Seus levantamentos também trazem informações socioeconômicas sobre os alunos e suas famílias, professores e diretores, além de informações sobre a infraestrutura nas escolas. Os levantamentos do SAEB são amostras representativas da população de escolas no Brasil. A subamostra aqui utilizada se restringe a 480 escolas que se repetem nas avaliações de 1999, 2001 e 2003, formando um painel de escolas balanceado de três anos.¹ Serão utilizados os resultados das provas de Português e Matemática dos alunos na 4^a série do Ensino Fundamental para construirmos os rankings de escolas. Utilizamos dados para a 4^a série, pois a evasão de alunos ainda é baixa nesse momento da vida escolar, o que evita problemas de seleção de alunos pelo nível socioeconômico. Os anos utilizados são os que tinham dados disponíveis para análise com painel de escolas no momento de elaboração desse artigo. O número de alunos avaliados em 1999, 2001 e 2003 é de 5.819, 13.629 e 13.114, respectivamente. A Tabela 1 traz a evolução do número de matrículas na 4^a série nas 480 escolas avaliadas no SAEB nos anos de 1999, 2001 e 2003.²

Tabela 1 - Matrículas na 4^a série e sua Evolução, por Escola, na Amostra de 480 Escolas Avaliadas pelo SAEB em 1999, 2001 e 2003

Variável		Média	Desvio-Padrão	Mínimo	Máximo	Observações
	Total	84,28	69,40	1,00	589,00	N = 28.432
Matrículas na 4 ^a . Série	Between		50,39	5,00	459,79	n = 480
	Within		29,28	-47,62	231,66	

Fonte: Censo escolar.

¹ Na Tabela A1 do Apêndice encontram-se as estatísticas descritivas das variáveis utilizadas nesta análise para a amostra do SAEB completa e para a subamostra de 480 escolas. A diferença entre as estatísticas descritivas das variáveis nos dois bancos de dados não é significativa, o que faz do painel de 480 uma amostra representativa da população de escolas do Brasil.

² Em alguns anos não há informações sobre o número de matrículas de algumas escolas no Censo Escolar, provavelmente devido a erros na computação dos dados. Em 1999, 30,35% das escolas da amostra não têm dados sobre matrículas na 4^a série (matrículas são iguais à zero), em 2001 essa porcentagem é de 1,80%, e em 2003 é de 16,15%. No total, 12,68% da amostra de escolas, considerando-se os três anos, não apresentam dados sobre matrículas. Estas escolas, porém, foram mantidas na análise, uma vez que o número de matrículas não foi utilizado para a construção dos rankings.

Na Figura 1 pode-se ter uma ideia de quanto varia o desempenho médio dos alunos nas provas de Matemática e Português de acordo com o tamanho da escola, medido pelo total de alunos matriculados na 4ª série. Percebe-se uma maior variabilidade na pontuação média das escolas menores, de acordo com o que foi observado para caso do Chile por Mcewan *et al.* (2007).

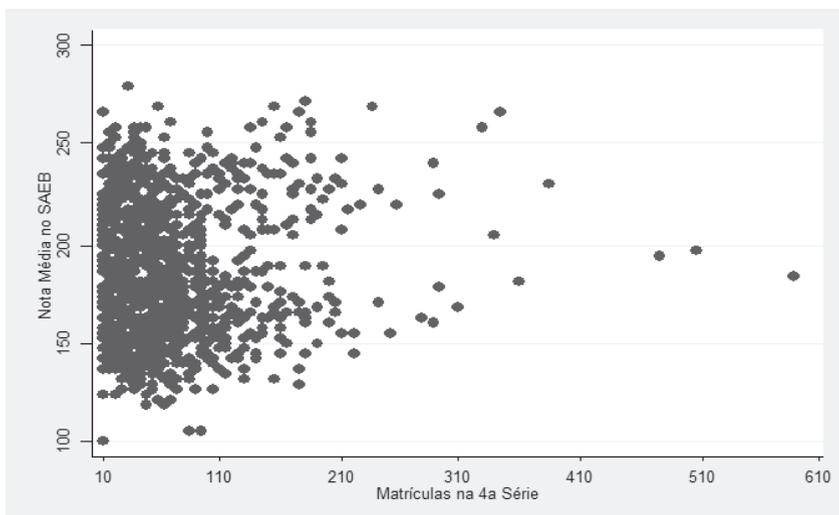


Figura 1 - Desempenho da escola e tamanho da escola, medidos pelas notas em matemática e português no SAEB e número de matrículas na 4ª série.

Fonte: Cálculos a partir das notas em matemática em 480 escolas que se repetem nos anos de 1999, 2001 e 2003 do SAEB e apresentam dados sobre matrículas.

Mas é quando se considera a diferença na proficiência média obtida em diferentes anos que se pode observar uma variância mais acentuada no caso das escolas pequenas. A Figura 2 mostra a relação entre a diferença na proficiência média das escolas (em Matemática e Português), medida entre os anos 1999-2001, e o tamanho da escola medido pelo número de alunos matriculados na 4ª série.³

³ Para o cálculo da diferença na proficiência, foram utilizados os dados apenas das 480 escolas que se repetem nos anos 1999, 2001 e 2003. Os gráficos das diferenças 2003-2001 e 2003-1999 não foram reportados por apresentarem padrão muito similar ao observado na Figura 2.

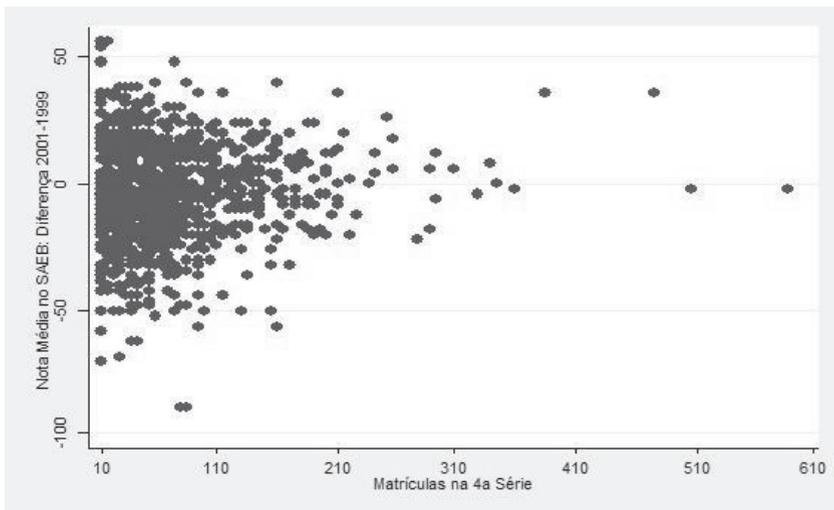


Figura 2 - Diferenças nas notas médias das escolas (2001-1999) e tamanho da escola, medidos pelas notas em matemática e português no SAEB e número de matrículas na 4ª série.

Fonte: Cálculos a partir das notas em matemática em 480 escolas que se repetem nos anos de 1999, 2001 e 2003 do SAEB e apresentam dados sobre matrículas.

As Figuras 1 e 2 sugerem indícios de ocorrência do fenômeno de “reversão para a média” para o qual Mizala *et al.* (2007), Chay *et al.* (*Op. cit*) e Mcewan *et al.* (*Op. cit*) chamam a atenção no Chile. Nas escolas menores, haveria a tendência dos resultados de desempenho variarem muito de um momento de avaliação para outro. Escolas que se saem muito mal em um primeiro momento tendem a apresentar, numa avaliação seguinte, uma recuperação, enquanto que as que se saem bem em seguida tendem a apresentar um resultado bem pior. Os autores demonstram que este efeito é maior quanto menor o número de alunos avaliados.⁴ Outro fator que contribui para a volatilidade nos rankings de escolas é a elevada estratificação por renda nas mesmas. A Figura 3 mostra a relação entre desempenho médio da escola nas avaliações de Português e Matemática na 4ª série e a proporção de mães com ensino superior na escola, uma variável *proxy* para indicar um elevado SSE dos alunos da escola que fizeram a avaliação. Nota-se claramente uma relação positiva, o que é forte indício de elevada estratificação social no sistema educacional brasileiro.

⁴ Ver, em especial, Chay *et al.* (*Op cit*, p. 1243).

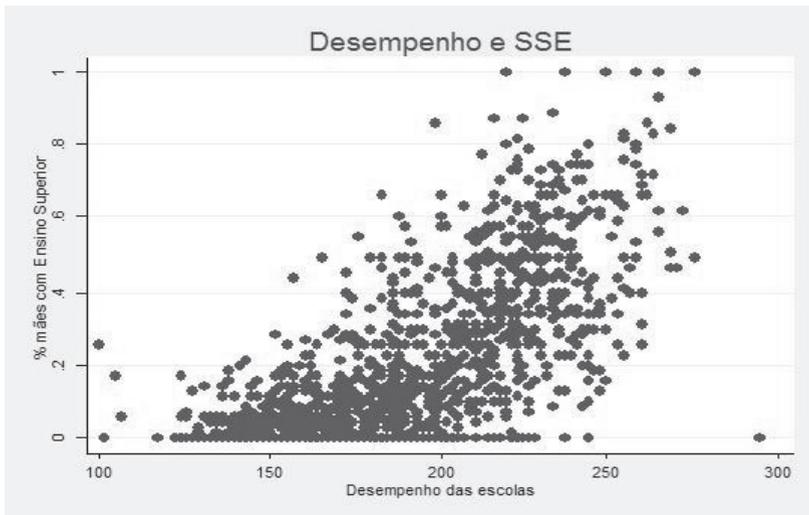


Figura 3 - Desempenho nas provas de matemática e português do SAEB e proporção de mães com ensino superior. Média para escolas, alunos da 4ª série.

Fonte: Cálculos a partir das 480 escolas que se repetem nos anos de 1999, 2001 e 2003 do SAEB.

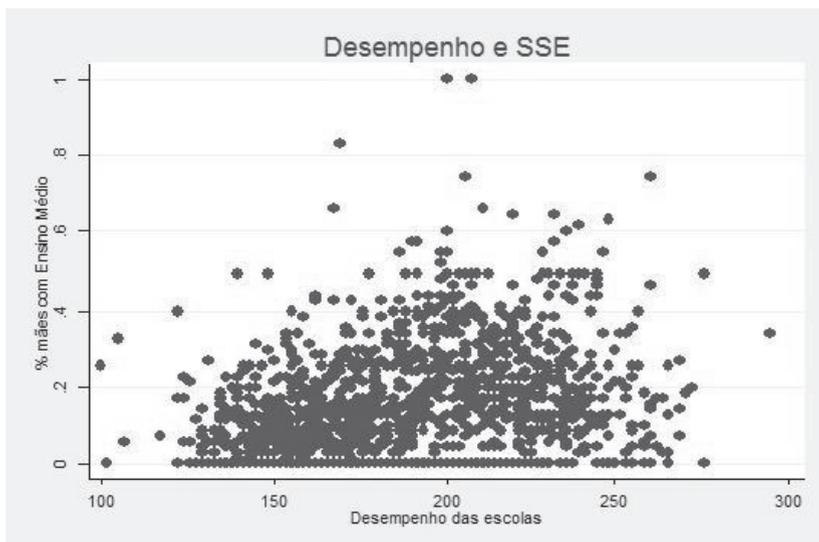


Figura 4 - Desempenho nas provas de matemática e português do SAEB e proporção de mães com ensino médio. Média para escolas, alunos da 4ª série.

Fonte: cálculos a partir das 480 escolas que se repetem nos anos de 1999, 2001 e 2003 do SAEB

A Figura 4, por sua vez, mostra a relação entre o desempenho médio da escola e a proporção de mães com Ensino Médio. A relação também se mostra positiva, embora não tão acentuada e clara como no caso da Figura 3.

Os Quadros 1, 2 e 3 abaixo mostram o ranking das 10 melhores escolas que se repetem no painel 1999-2003, sem controles para quaisquer características dos alunos, escolas, diretores, professores ou turmas.

Quadro 1 - Ranking das 10 Melhores Escolas de 1999

10 melhores de 1999	Pontuação média no SAEB 1999 (Matemática e Português)	Desvio-Padrão	Alunos avaliados da 4ª série	Média de alunos por sala de 4ª série	Matriculas 4ª série em 1999	Ranking 2001	Ranking 2003
1	293,55	26,40	6	8	8	84	6
2	279,50	59,24	12	20	33	38	90
3	275,85	20,59	4	8	14	105	9
4	275,38	43,18	31	11	22	6	11
5	265,08	34,01	68	34	344	3	3
6	264,69	33,81	26	18	54	5	43
7	260,49	41,87	63	33	185	27	47
8	259,98	42,10	10	19	66	8	44
9	259,65	32,47	58	34	143	1	1
10	258,35	23,79	3	10	10	31	299

Fonte: Cálculos a partir dos dados de 480 escolas que se repetem no SAEB 1999, 2001 e 2003.

No Quadro 1 pode-se perceber que há maior volatilidade nos resultados das escolas menores, com menos de 100 alunos matriculados na 4ª série. O comportamento é tão mais volátil quanto menor o tamanho da escola e menor o número de alunos avaliados pelo SAEB. O mesmo padrão se observa nos Quadros 2 e 3, que trazem os mesmos resultados das 10 primeiras escolas classificadas em 2001 e 2003, respectivamente.

Quadro 2 - Ranking das 10 Melhores Escolas de 2001

10 melhores de 2001	Proficiência média 2001 (Matemática e Português)	Desvio-Padrão	Alunos avaliados da 4ª série	Média de alunos por sala de 4ª série	Matrículas 4ª série em 2001	Ranking 1999	Ranking 2003
1	270,04	32,47	58	32	178	9	1
2	268,45	36,96	64	39	154	12	4
3	263,72	34,01	68	32	-	5	3
4	259,05	44,51	18	14	23	20	15
5	258,86	33,81	26	14	40	6	43
6	255,20	43,18	31	23	36	4	11
7	254,98	50,82	18	9	-	51	478
8	254,82	42,48	18	18	18	8	44
9	254,46	35,30	35	19	96	31	62
10	253,19	41,27	36	29	63	28	31

Fonte: Cálculos a partir dos dados de 480 escolas que se repetem no SAEB 1999, 2001 e 2003.

Quadro 3- Ranking das 10 Melhores Escolas de 2003

10 melhores de 2003	Proficiência média (Português e Matemática)	Desvio - Padrão	Alunos avaliados da 4ª série	Média de alunos por sala de 4ª série	Matrículas 4ª série em 2003	Ranking 1999	Ranking 2001
1	272,48	32,47	58	31	-	9	1
2	267,93	38,12	30	31	57	50	47
3	267,85	34,01	68	32	235	5	3
4	265,13	36,96	64	35	173	12	2
5	265,07	30,16	13	14	10	15	18
6	262,30	24,41	6	9	-	1	84
7	261,99	41,54	20	20	28	25	12
8	260,98	33,20	6	6	8	57	322
9	259,34	20,59	4	4	4	3	105
10	258,44	38,08	26	10	166	63	76

Fonte: Cálculos a partir dos dados de 480 escolas que se repetem no SAEB 1999, 2001 e 2003.

Esses resultados mostram que nas escolas com menos de 50 alunos a variabilidade nos rankings de um ano para outro é muito elevada. Isso prejudica sensivelmente a utilização desses rankings como instrumento para informar a sociedade sobre a qualidade das escolas e premiar os professores com base nos resultados dos alunos.

4. Metodologia

Para aprofundar a análise de quanto as medidas usuais utilizadas para construção de rankings de escola espelham principalmente o status socioeconômico ou, por outro lado, implicam a volatilidade demasiada de um ano para outro (ambos efeitos indesejados, caso estes rankings sirvam de parâmetros para a tomada de decisão dos agentes), serão calculadas cinco medidas de desempenho com base

nos exames do SAEB de Matemática para a 4ª série, seguindo-se a abordagem dada por Mizala *et al.* (2007). Quatro dessas medidas são obtidas após o ajuste de um modelo de regressão, como descrito no Quadro 4. A quinta medida é, para cada escola em um determinado ano, o acréscimo na média de notas SAEB /escola em relação ao ano anterior.

Quadro 4 - Descrição de 4 Medidas de Desempenho de Escolas Obtidas por Meio de Regressões

*Medida	Covariáveis	Resposta	Medida utilizada para a escola j
1	Dummies de escolas	Nota SAEB/aluno no ano	Estimativa do coeficiente da <i>dummy</i> de escola j
2	Dummies de escolas e SSE dos alunos	Nota SAEB/aluno no ano	Estimativa do coeficiente da <i>dummy</i> de escola j
3	Intercepto e SSE dos alunos	Nota SAEB/aluno no ano	Média dos resíduos relativos à escola j
4	Intercepto e média SSE /escola	Nota média SAEB/escola no ano	j-ésimo resíduo.

* Número da equação que descreve o modelo como aparece no texto.

A **primeira** destas medidas é a mais simples, pois utiliza dados *cross section* das notas de alunos, e as escolas são ordenadas de acordo com suas médias. Esta medida aparece em praticamente todos os modelos de responsabilização. Num contexto de regressão, ela pode ser representada por uma constante específica da escola j, no modelo de regressão dado por:

$$y_{ij} = \alpha_j + u_{ij} \quad i=1, \dots, n_j ; j=1, \dots, 480 \quad (1)$$

Em (1) y_{ij} é a nota do aluno i na escola j, α_j é o efeito da escola j e u_{ij} o erro idiossincrático. A Equação (1) pode ser escrita também na forma vetorial como

$$y = \mathbf{D}\boldsymbol{\alpha} + \mathbf{u} \quad (1')$$

sendo y : vetor $N \times 1$ das observações da nota de cada aluno de cada escola

$\mathbf{N} = n_1 + n_2 + \dots + n_{480}$

\mathbf{D} : matriz $N \times 480$ contendo 1-s e 0-s

\mathbf{u} : vetor de erros aleatórios $N \times 1$

$\boldsymbol{\alpha}$: vetor 480×1 de coeficientes de interceptos específicos de escolas.

Seja \bar{y}_j = nota média dos alunos da escola j e \mathbf{y} o vetor 480×1 destas médias para todas as escolas. A estimativa $\hat{\mathbf{a}}$ de \mathbf{a} é o vetor $\bar{\mathbf{y}}$ e espelha fortemente a composição do status socioeconômico dos alunos. Isto pode ser observado na Tabela 2 abaixo, em que foi realizada a regressão da nota média da escola em variáveis observáveis que capturam o status socioeconômico médio dos alunos que compõem a escola e suas famílias. Na ausência de informações sobre a renda das famílias, foram incluídas *dummies* para escolaridade da mãe e se o aluno tem ou não computador. Sendo $\bar{\mathbf{X}}$ a matriz 480×8 de médias por escola de 8 variáveis *dummies* que espelham o SSE dos alunos (sexo, educação da mãe com 6 níveis, e posse de computador pelo aluno), $\boldsymbol{\gamma}$ um vetor 8×1 e \mathbf{v} um vetor 480×8 de erros idiossincráticos, foi ajustado o seguinte modelo:

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}\hat{\mathbf{a}} + \mathbf{v} \quad (2)$$

A Tabela 2 traz o R^2 resultante da estimação por MQO do Modelo (2) para cada ano.

Tabela 2 - Valores de R^2 em Ajustes por MQO do Modelo (2) de Regressão Linear*

	1999 (4ª Série)	2001 (4ª série)	2003 (4ª série)
R^2	64,3%	71,8%	71,3%

Fonte: Cálculos a partir de 480 escolas que se repetem no SAEB 1999, 2001 e 2003. Regressores: valores médios por escola do gênero do aluno, educação da mãe (6 níveis) e presença de computador em casa.

É possível observar que, mesmo na ausência da medida de renda da família, as variáveis disponíveis que espelham o status socioeconômico (SSE) dos alunos explicam entre 64% e 71% dos resultados do desempenho médio observado da escola. Portanto, o ranking que emerge de \mathbf{a} no Modelo (1), tem a desvantagem de refletir majoritariamente o SSE dos alunos que compõem a escola e, por isso, de não captar adequadamente o “efeito escola”, o que seria desejável para guiar as decisões dos elaboradores de políticas educacionais e famílias.

Uma **segunda** medida de desempenho das escolas é a obtida a partir do ajuste de uma regressão das notas SAEB dos indivíduos que leva em conta o efeito específico de escola e a influência de variáveis SSE individuais. Essa regressão pode ser expressa por:

$$y_{ij} = \delta_j + X_{ij}\beta + u_{ij} \quad (3)$$

$$i=1, \dots, n_j ; j=1, \dots, 480$$

Em (3) y_{ij} é a nota do aluno i na escola j , δ_j é o intercepto específico da escola j X_{ij} é um vetor linha $1 \times p$ com as características de SSE do aluno i na escola j , β é um vetor $p \times 1$ e u_{ij} é o erro idiossincrático. Seja d o vetor 480×1 de estimativas dos interceptos específicos δ_j de escolas no Modelo (3). O vetor d é a **segunda medida** que pode ser usada para fazer o ranking de escolas.

Para avaliar se a segunda medida difere muito da primeira, pode-se verificar se há ganho no R^2 da regressão quando se acrescentam as variáveis de controle SSE ao modelo de regressão. Usando a mesma simbologia usada no Modelo (3), pode-se ajustar um modelo sem efeitos específicos de escolas dado por:

$$y_{ij} = X_{ij}\beta + u_{ij} \quad (4)$$

$$i=1, \dots, n_j ; j=1, \dots, 480$$

A Tabela 3 mostra os valores de R^2 obtidos nos ajustes dos Modelos (1), (4) e (3), respectivamente.

Tabela 3 - Valores de R^2 em Ajustes por MQO de 3 Modelos de Regressão Linear da Variável Resposta “nota SAEB/aluno (Português e Matemática)”

	1999	2001	2003
Modelo (1): apenas <i>dummies</i> de escolas	5.819	13.629	13.114
N	30,99%	33,92%	47,04%
R^2			
Modelo (4): apenas controles de SSE dos alunos ^a	5.819	12.246	13.064
N	18,77%	24,20%	22,25%
R^2			
Modelo (3): <i>dummies</i> de escolas e controles de SSE dos alunos	5.819	12.026	13.064
N	33,01%	36,72	47,56%
R^2			

Fonte: Cálculos a partir dos SAEB de 1999, 2001 e 2003.

a) No Modelo (4) os controles de SSE são 6 *dummies* para nível de educação da mãe, *dummy* de sexo do aluno e *dummy* para posse de computador pelo aluno.

Os resultados na Tabela 3 mostram que o Modelo (1), que tem apenas *dummies* de escola como regressoras, explica entre 30,99% e 47,04% das variações nas notas. No Modelo (4), observa-se que as características de SSE dos alunos explicam entre 18,77% e 24,20% da variação das notas. Comparando o Modelo (3) com o Modelo (1), observa-se que, uma vez controlada a regressão com *dummies* de escolas, as características de SSE adicionam muito pouca informação para a variação das notas (de 33,01% a 47,56%).

Isto indica que há uma forte correlação entre as estimativas (a e d) dos interceptos nas Equações (1) e (3), precisamente de 81,95% em 1999, 86,96% em 2001 e 99,90% em 2003, resultado que Mizala *et al.* (2007) também encontraram para o caso do Chile. Segundo esses autores, isto sugere que o sistema apresenta uma forte estratificação social, e que a escolha de onde se matricular já transmite muito das características de SSE observáveis do aluno e, possivelmente, também das características não observáveis que contribuem para o aprendizado.

A terceira medida é construída a partir dos resíduos que emergem da regressão da seguinte equação:

$$y_{ij} = \alpha_0 + X_{ij}\beta + v_{ij} \quad (5)$$

$$i=1, \dots, n_j ; j=1, \dots, 480$$

Em (5) y_{ij} é a nota do aluno i na escola j , α_0 é um intercepto comum a todas as escolas, X_{ij} é um vetor linha $1 \times p$ com as características de SSE do aluno i na escola j , β é um vetor $p \times 1$ e v_{ij} é o erro idiossincrático.

Segundo Mizala *et al.* (2007), as escolas podem ser “rankeadas” de acordo com seu resíduo médio \bar{v}_j . Esta medida seria menos influenciada por características de SSE observáveis, uma vez que se faz o controle por elas em X_{ij} . Por outro lado, espera-se que ela seja mais volátil, pois não contém mais o efeito-escola específico que, como mostra a comparação das regressões (1) e (3), capta muito bem a influência da escola.

A **quarta** medida de desempenho da escola j é o j -ésimo resíduo do seguinte modelo de regressão:

$$\bar{y}_j = \alpha_0 + \bar{\mathbf{X}}_j \boldsymbol{\beta} + v_j \quad (6)$$

$j = 1 \dots 480$

Em (6) \bar{y}_j é a nota média dos alunos da escola j , α_0 é um intercepto comum, $\bar{\mathbf{X}}_j$ a matriz 1×8 de médias da escola j de 8 variáveis *dummies* que espelham o SSE dos alunos (sexo, educação da mãe com 6 níveis, e posse de computador pelo aluno), $\boldsymbol{\beta}$ é um vetor 8×1 e v_j é o erro idiossincrático. Os resíduos v_j resultantes do ajuste do Modelo (6) são utilizados para montar o ranking.

Por fim, a **quinta** e última medida analisada aqui é a que se obtém pela diferença no desempenho médio das escolas de um ano ($t = 0$) para o ano seguinte ($t = 1$). Esta é uma medida bastante utilizada para se tentar capturar o “valor adicionado” da escola.⁵ Levando-se em conta as notas médias nos testes por escolas e as características de SSE médias de seu corpo discente, esta medida pode ser representada pela diferença entre as Equações (7a) e (7b), dada em (7c)

$$\bar{y}_{j0} = \alpha_j + \bar{\mathbf{X}}_{j0} \boldsymbol{\beta} + \bar{v}_{j0} \quad (7a)$$

$$\bar{y}_{j1} = \alpha_j + \bar{\mathbf{X}}_{j1} \boldsymbol{\beta} + \bar{v}_{j1} \quad (7b)$$

$$\Delta \bar{y}_j = \bar{y}_{j1} - \bar{y}_{j0} = \bar{v}_{j1} - \bar{v}_{j0} \quad (7c)$$

$$j = 1 \dots 480$$

Nas Equações (7a) a (7c), \bar{y}_{jt} é a nota média dos alunos da escola j no tempo t , α_j é o efeito-escola específico da escola j e $t = 0, 1$. Se a composição socioeconômica média da escola é estável no tempo, e é bastante razoável se esperar que o seja, a diferença entre (7a) e (7b) elimina a sua influência, bem como o efeito específico da escola que não varia no tempo. Por isso, a volatilidade que pode emergir do ranking com a medida dada por (7c) dependerá da diferença

⁵ Para ser mais precisa, a mensuração correta do valor adicionado pela escola exige que se acompanhe o progresso de cada aluno ao longo dos anos. No Brasil, e também na maioria dos países em desenvolvimento, este tipo de informação ainda não está disponível.

$\Delta \bar{y}_j = \bar{v}_{j1} - \bar{v}_{j0}$, muito influenciável por choques ocorridos em apenas um dado momento que podem ou não perdurar no tempo.

5. Resultados

Nos Quadros 5 e 6 estão os resultados esperados para o caso do ranking de escolas ter caráter completamente estável, caso de volatilidade esperada nula (coluna 1), ou ter o comportamento de uma loteria, caso de volatilidade esperada máxima (coluna 2).⁶ Estas duas situações extremas servem de parâmetro para se comparar a volatilidade dos rankings obtidos pelas cinco diferentes medidas acima apresentadas.

Quadro 5 - Distribuições Teóricas de Freqüências Percentuais do “Número de vezes (K) em que uma escola aparece entre as 20% melhores em 3 anos” nos casos de “certeza” e “loteria”, Comparadas com Freqüências Percentuais “Número de escolas que aparecem K vezes/número total de escolas (480)” Segundo Várias Medidas

K	Dados anuais						Diferenças temporais		
	Modelos teóricos		Usando Medidas definidas por				Modelos teóricos		Usando diferenças
	Certeza (1)	Loteria (2)	Médias de escola (3)	Médias ajustadas (4)	Resíduos individuais (5)	Resíduos escolas (6)	Certeza (7)	Loteria (8)	Ganhos $\Delta \bar{y}_j$ (9)
0	80,0	51,2	72,08	78,75	61,5	58,3	80,0	64,0	73,9
1	0	38,4	6,88	1,04	22,7	25,8	0	32,0	12,5
2	0	9,6	10	1,04	10,2	13,3	20,0	4,0	13,6
3	20,0	0,8	11	18,54	5,6	2,5	-	-	-

Fonte: Cálculos próprios a partir de 480 escolas que se repetem nas amostras do SAEB 1999, 2001 e 2003. Rankings com base em medidas obtidas a partir dos dados de alunos da 4ª série em Matemática e Português.

⁶ A loteria é simulada encarando-se cada julgamento de escola como um ensaio de Bernoulli independente dos demais ensaios, com M sendo o número de ensaios e 0,20 a probabilidade de sucesso. Assim, a probabilidade de K sucessos é dada pela Binomial ($M=3, P=0,2$): $P(\text{Número de sucessos} = k) = M! / [k! (M-k)!] 0,20^k (0,80)^{M-k}$, $k=0,1,2,\dots,M$.

Quadro 6 - Distribuições Teóricas de Frequências Percentuais do “Número de vezes (K) em que uma escola aparece entre as 20% piores em 3 anos” nos casos de “certeza” e “loteria”, Comparadas com Frequências Percentuais “Número de escolas que aparecem K vezes/ número total de escolas (480)” Segundo Várias Medidas

K	Dados anuais						Diferenças temporais		
	Modelos teóricos		Usando Medidas definidas por				Modelos teóricos		Usando
	Certeza (1)	Loteria (2)	Médias (3)	Médias ajustadas (4)	Resíduos indivíduos (5)	Resíduos escolas (6)	Certeza (7)	Loteria (8)	Ganhos $\Delta \bar{y}_j$ (9)
0	80,0	51,2	65,83	77,92	62,9	56,46	80,0	64,0	71,1
1	0	38,4	15,63	2,50	19,2	29,17	0	32,0	18,1
2	0	9,6	11,25	1,25	12,9	12,29	20	4,0	10,8
3	20,0	0,8	7,29	18,33	5,00	2,1	-	-	-

Fonte: Cálculos próprios a partir de 480 escolas que se repetem nas amostras do SAEB 1999, 2001 e 2003. Rankings com base em medidas obtidas a partir dos dados de alunos da 4ª série em Matemática e Português.

O comportamento do ranking que emerge das médias “puras” das notas por escolas na Equação (1) pode ser visto nos Quadros 5 e 6. Como exemplo, no Quadro 5, coluna 3, verifica-se que 72,08% das escolas nunca apareceram entre as 20% melhores, nos três anos analisados. Esta proporção é muito próxima do caso de “certeza”, que seria de 80%, o que indica estabilidade na posição dessas escolas. Por outro lado, 11% das escolas aparecem nos três anos, segundo esta medida, entre as 20% melhores. Esses dados mostram que esta medida gera rankings com caráter certo, o que é bastante desejável. Mas, como visto na discussão da Tabela 2, o ranking definido pelas médias puras reflete o status socioeconômico dos alunos. Como Mizala *et al.* (2007) encontraram para o caso do Chile, aqui o ranking é estável, mas reflete fundamentalmente o SSE.

Quanto à volatilidade do ranking com base em d que surge da Equação (3), a coluna 4 do Quadro 5 mostra que a distribuição que emerge deste tipo de medida é ainda mais estável do que o resultado apresentado pelas médias puras. Este resultado é esperado, uma vez que o que se está medindo é o efeito específico não observado da escola que, no contexto de uma regressão, não varia no tempo. Porém, mais uma vez, embora a estabilidade seja desejável, esta medida tem fundamentalmente como base o SSE dos alunos das escolas,

como mostram as elevadas correlações entre d da Equação (3) e a da Equação (1) (ver seção 4).

A coluna 5 do Quadro 5 mostra que a distribuição da medida resultante do modelo de regressão da Equação (5) apresenta maior volatilidade que as demais vistas até aqui, e aproxima-se da distribuição de uma loteria.

A coluna 6 do Quadro 5 mostra a elevada volatilidade na distribuição que a medida resultante do modelo de regressão dado pela Equação (6) gera, apesar desta medida controlar bem pelo SSE e os efeitos específicos não observados das escolas.

Já o que se pode observar na coluna 9 do Quadro 5, com respeito à quinta medida obtida a partir da Equação (7c), é que a distribuição resultante da utilização das diferenças nas médias de notas de um ano para o outro apresenta um caráter bastante estável, o que é desejável, e ao mesmo tempo expurga de seus resultados os efeitos do SSE dos alunos.

Pode-se afirmar, portanto, que de todas as medidas aqui avaliadas, a medida dada por (7c) seria uma boa medida para servir de referência nas avaliações sobre a qualidade nas escolas, uma vez que controla os efeitos do SSE dos alunos e os efeitos específicos não observados das escolas e apresenta, ao mesmo tempo, estabilidade. Este resultado difere daquele encontrado em Mizala *et al.* (2007). Uma explicação, talvez, seja que aquele estudo contou com um período mais longo de observações (8 anos) e também com dados de natureza censitária de uma região específica do Chile. De acordo com os seus resultados, a distribuição do ranking com base nas diferenças se aproximava claramente ao de uma distribuição de loteria. No caso do Brasil, porém, dado que o SAEB é uma amostra representativa da população, o uso da diferença nas médias apresentou as características desejáveis para uma medida de qualidade das escolas: não é influenciada pelo SSE e variáveis escolas específicas não observadas, e também não apresenta no tempo uma volatilidade próxima à de um resultado de loteria.

6. Comentários Finais

Os resultados acima alertam para o cuidado que os agentes econômicos devem ter ao tomarem como base os rankings de escolas, construídos a partir de exames de avaliação educacional como o SAEB, para fazerem inferências sobre a qualidade nas escolas ou para o desenho de incentivos e sanções num contexto de “políticas de responsabilização”.

A elevada estratificação do sistema, por um lado, e o processo de reversão para a média que tende a ocorrer nos resultados de um momento no tempo para o outro, geram ruídos nas medidas de desempenho utilizadas para a construção de rankings, de forma que estes possam apresentar resultados influenciados principalmente pelo status socioeconômico dos alunos ou uma volatilidade grande de um período de avaliação para o outro.

Das cinco medidas estudadas neste trabalho, apenas aquela que leva em conta a diferença nas médias (Equação 7c) resultou num ranking com as desejáveis características de não ser fortemente determinado pelo SSE dos alunos e pelas variáveis não observáveis das escolas e, ao mesmo tempo, apresentar baixa volatilidade.

No Brasil, como é crescente a preocupação com a qualidade do ensino público, e caminhamos para políticas nos moldes de “responsabilização”, os resultados aqui encontrados são bastante importantes para serem levados em conta na definição de índices de desempenho e de metas a serem atingidas pelas escolas.

Referências

- Chay, Kenneth Y., McEwan, P., Urquiola, M. The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review*, v.4, n.95, p. 1237-1258, 2005.
- Fernandes, Reynaldo. *Índice de desenvolvimento da educação básica (IDEB)*. Texto para discussão n.26, 2007. Disponível em: <<http://www.publicações.inep.gov.br>>. Acesso em: 15/07/2008.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP. Nota técnica: índice de desenvolvimento da educação básica. Disponível em: <<http://www.inep.gov.br>>. Acesso em 15/07/2008.
- McEwan Patrick J., Urquiola, M, Vegas, E. School choice, stratification and information on school performance: lessons from Chile. Preliminary draft for presentation at the 16th Panel of Economic. *Journal of LACEA*, 2007.
- Menezes-Filho, Naercio A., Soares, F. *Definição de metas de desempenho para as escolas estaduais paulistas*. Mimeo. 2008.

Mizala, Alejandra, Romaguera, P., Urquiola, M. Tradeoffs in the generation of school quality information. *Journal of Development Economics*. n. 84, p.61-75, 2007.

Apêndice

Tabela A1 - Estatísticas Descritivas das Variáveis da Amostra do SAEB e do Painel de 480 Escolas

Variáveis	1999		2001		2003	
	Amostra SAEB	Amostra 480 escolas	Amostra SAEB	Amostra 480 escolas	Amostra SAEB	Amostra 480 escolas
Escolas	4.108	480	4.048	480	3.267	480
Proficiência Média	178,21 (44,58)	189,50 (47,66)	179,27 (49,13)	186,05 (51,35)	183,34 (47,46)	189,15 (50,09)
Alunos	42.294	5.819	108.733	13.629	87.254	13.114
Alunos do sexo masculino	0,506 (0,500)	0,504 (0,500)	0,509 (0,499)	0,513 (0,450)	0,500 (0,500)	0,493 (0,500)
Aluno tem computador	0,195 (0,397)	0,270 (0,444)	0,250 (0,433)	0,311 (0,463)	0,201 (0,401)	0,266 (0,442)
<u>Estudo da mãe:</u>						
Não estudou	0,085 (0,279)	0,056 (0,231)	0,075 (0,264)	0,061 (0,240)	0,033 (0,179)	0,030 (0,170)
Até 4ª série	0,236 (0,425)	0,192 (0,394)	0,232 (0,422)	0,220 (0,414)	0,222 (0,415)	0,197 (0,398)
Até 8ª série	0,166 (0,372)	0,162 (0,369)	0,170 (0,376)	0,153 (0,360)	0,146 (0,353)	0,127 (0,333)
Ensino Médio	0,083 (0,276)	0,095 (0,294)	0,152 (0,359)	0,157 (0,364)	0,186 (0,389)	0,200 (0,400)
Faculdade	0,142 (0,349)	0,209 (0,407)	0,133 (0,339)	0,163 (0,369)	0,158 (0,364)	0,206 (0,404)
Aluno não soube responder	0,279 (0,449)	0,276 (0,447)	0,287 (0,452)	0,291 (0,454)	0,257 (0,437)	0,242 (0,428)

Fonte: Cálculos próprios a partir dos dados do SAEB de 1999, 2001 e 2003. Nota: Desvios-padrão em parênteses.