

# Desigualdades acadêmicas induzidas pelo contexto escolar

Cibele Comini César\*  
José Francisco Soares\*\*

*O desempenho acadêmico é influenciado por diversos fatores, usualmente classificados em dois grandes grupos: alunos (gênero, condição socioeconômica etc.) e ambiente escolar (rede de ensino, formação do corpo docente, capacidade de liderança do diretor etc.). Além destes fatores, deve-se considerar o efeito da turma sobre o desempenho de um aluno específico, conhecido na literatura por efeito contextual ou efeito de pares. Este artigo avalia como o efeito socioeconômico dos pares deve ser considerado em um estudo do desempenho acadêmico dos alunos. Em geral, a introdução do fator é feita por meio da inclusão da média da variável calculada em cada turma. Utilizando os resultados do teste de Matemática da 8ª série do SAEB-99, mostra-se que a inclusão da média não é suficiente para captar todo o efeito dos pares.*

## Introdução

O desempenho acadêmico dos alunos do ensino fundamental e médio é influenciado tanto pelas suas características pessoais quanto pelas características das escolas freqüentadas. Desde a publicação do Relatório Coleman, está claramente estabelecido que os fatores individuais explicam uma porcentagem muito maior da variação observada entre as proficiências dos alunos do que os fatores escolares. No entanto, o poder explicativo destes últimos fatores é suficientemente alto para alterar a trajetória escolar dos alunos. Assim sendo, os fatores escolares têm sido amplamente estudados na literatura e são também o objeto deste artigo.

Holland (1986) classifica os fatores escolares em duas categorias: tratamentos e atributos. A primeira categoria é composta por aqueles fatores associados às políticas e práticas internas da escola. Como

exemplo, podemos citar método de ensino, a relação aluno/professor, o currículo utilizado, a formação do corpo docente, a capacidade de liderança pedagógica e administrativa do diretor, entre outros. Há uma grande literatura sobre o que é denominado de "efetividade da escola". Veja, por exemplo, Sammons (1999).

O segundo grupo de fatores é constituído por fatores associados ao contexto social, espacial e demográfico no qual a escola está inserida mas que não estão sob o controle imediato desta.

A introdução no modelo de variáveis do contexto da escola é especialmente importante nos estudos realizados com dados do Brasil, uma vez que não se dispõe de informação sobre o nível do aluno no início do processo. Segundo Thomas e Mortimore (1996), quando o nível de desempenho prévio do aluno não está disponível, os fatores de contexto são mais úteis e podem ser vistos como medidas

---

\* Doutora em Demografia, professora adjunta do Departamento de Estatística e pesquisadora do Centro de Desenvolvimento e Planejamento Regional (Cedeplar) da Universidade Federal de Minas Gerais (UFMG).

\*\* Ph.D. em Estatística, professor titular do Departamento de Estatística da UFMG e coordenador geral do Grupo de Avaliação e Medidas Educacionais (GAME).

aproximadas, porém adequadas, do nível do aluno na entrada.

Esta distinção é importante na discussão do *efeito de escola*. O fato de um aluno freqüentar uma determinada escola tem um efeito específico sobre o seu desempenho. Raudenbush e Willms (1995) definem dois tipos de efeitos de escola, A e B, motivados pela distinção entre os dois tipos de fatores escolares descritos acima.

O efeito Tipo A é a diferença entre o desempenho real do aluno e o que ele teria caso freqüentasse uma escola típica. Para a determinação do efeito Tipo A é necessário o ajuste de um modelo estatístico que inclua como variáveis de controle a caracterização sociodemográfica e cultural dos alunos. Não se deve controlar, neste caso, pelas variáveis de contexto. Raudenbush e Willms (1995) enfatizam que os pais tendem a mandar seus filhos para uma escola que, na sua percepção, produz o maior efeito Tipo A, não interessando se a eficácia da escola é decorrente da excelência dos professores, ou de sua composição discente favorável, ou da influência benéfica do contexto econômico e social no qual a escola está inserida.

O efeito Tipo B mede apenas o efeito das práticas da escola, incluindo-se aí aspectos administrativos, pedagógicos, curriculares, ou seja, os tratamentos, na linguagem de Holland (1986). Neste efeito não está incluída a influência do contexto. O efeito Tipo B é, então, a diferença entre o desempenho real do aluno e o desempenho que se esperaria que ele tivesse caso freqüentasse uma escola com contexto idêntico. Para a determinação do efeito Tipo B é, pois, necessário usar um modelo estatístico que controle pelas características do aluno e pelas variáveis de contexto no qual a escola está inserida.

Os atributos ou variáveis de contexto podem ser decorrentes meramente do ambiente em que a escola está inserida, ou seja, do contexto econômico e social da comunidade em que a escola está localizada, ou definidos a partir da composição de seu corpo discente. Este último é chamado *efeito de pares*. Segundo Goldstein (2001, p. 90), "as crianças também aprendem com seus

colegas, e as características e realizações de outros alunos na escola irão influenciar o desenvolvimento de cada criança em particular."

Há diversas evidências empíricas de que o fato de o aluno estar em uma sala de aula com alunos com alta motivação, por exemplo, exerce uma influência no desempenho deste aluno, independente do efeito da sua própria motivação. Assim, um aluno com uma dada capacidade acadêmica poderá ter um melhor desempenho em uma classe formada basicamente por bons alunos do que em uma turma com alunos com capacidade acadêmica menor que a sua, independente de outras características da escola (Aitkin e Longford, 1986).

O efeito de pares, embora seja encontrado com alta freqüência e não deva ser ignorado, não tem uma interpretação direta e clara, como observa Raudenbush (1989). Pode ser resultante da interação do indivíduo com o grupo, ficando aquele sujeito aos efeitos normativos dos seus pares. Pode ser o reflexo das práticas instrucionais e dos recursos alocados a esse contexto em função de sua composição. Pode, ainda, estar atuando como *proxyp* para características do aluno não medidas ou medidas com erro.

O efeito dos pares tem sido considerado na literatura mediante a inclusão de uma única variável explicativa, definida pela média da capacidade prévia dos alunos da turma ou escola freqüentada pelos alunos. Entretanto, o *efeito de pares* pode ser observado também com relação à média de outras variáveis que caracterizam o grupo no qual o aluno está inserido, bem como através de outras medidas de síntese da variável, que não a média. Por exemplo, pode-se considerar o nível socioeconômico médio dos alunos, porcentagem de alunos provenientes de minorias étnicas, fração do corpo discente do sexo feminino, variabilidade do desempenho prévio dos alunos medida através do desvio padrão, do intervalo interquartilico, do coeficiente de variação etc. (Lee e Bryk, 1989; Longford, 1989; Plewis, 1989; Raudenbush e Willms, 1995).

A existência de fatores medidos em níveis distintos – aluno, escola, comunidade – atuando sobre a variável resposta – desempenho do aluno – indica a necessidade da utilização dos modelos de regressão multiníveis ou hierárquicos (Bryk e Raudenbush, 1992; Goldstein, 1995; Snijders e Bosker, 1999). Esta metodologia é adequada para a análise de amostras que exibem uma estrutura de variabilidade complexa, em que as fontes de variabilidade se encontram aninhadas.

O uso da metodologia justifica-se não somente por resolver o problema da “unidade de análise”, uma vez que permite a incorporação de fatores medidos nos diversos níveis, mas também pelo procedimento amostral utilizado, que induz uma estrutura hierárquica nos dados. Em decorrência do processo de amostragem, é gerada uma correlação entre as observações.

Nos modelos multiníveis ou hierárquicos prevalece a idéia de que devem ser ajustados modelos em cada um dos contextos ou grupos. Estes modelos têm a mesma variável resposta e as mesmas variáveis explicativas, mas diferentes coeficientes de regressão. Esses são os modelos de nível 1, uma vez que modelam a resposta observada. Os modelos de nível 1 definidos em cada grupo são relacionados através de um modelo de segundo nível ou de nível 2, em que os coeficientes de regressão do primeiro nível são as variáveis respostas de um modelo de regressão nas variáveis explicativas de segundo nível (Kreft e De Leeuw, 1998). O procedimento generaliza-se para um número maior de níveis.

Neste artigo, argumenta-se que o efeito socioeconômico dos pares deve ser considerado em todos os estudos de explicação do desempenho acadêmico e no cálculo do efeito Tipo B. Isto deve ser feito mediante a inclusão, como variável de controle, não só de um indicador do nível socioeconômico típico do grupo, mas também de uma medida de dispersão dos índices socioeconômicos dos alunos. Argumenta-se, ainda, que há necessidade de inclusão da interação entre o valor do

índice individual e o valor do índice no grupo.

Estas hipóteses são avaliadas utilizando os dados de uma amostra de alunos brasileiros da 8ª série, cuja proficiência em Matemática foi medida no âmbito do SAEB-1999.

## Os dados

O Sistema de Avaliação da Educação Básica (SAEB) utiliza uma amostra probabilística representativa de todas e cada uma das unidades da Federação brasileira – os estados e o Distrito Federal. O sistema foi implantado em 1990 e é coordenado pelo INEP, contando com a participação e o apoio das Secretarias Estaduais de Educação. Segundo Franco (2001), “o objetivo declarado do SAEB é gerar e organizar informações sobre a qualidade, a equidade e a eficiência da educação nacional, de forma a permitir o monitoramento das políticas brasileiras, fornecendo subsídios para a melhoria da qualidade, equidade e eficiência da educação no Brasil.” Já foram realizadas seis avaliações do SAEB, em 1990, 1993, 1995, 1997, 1999 e 2001.

Em 1999, os alunos foram testados em Língua Portuguesa, Matemática, Ciências, História e Geografia. Cada aluno participante da amostra foi avaliado em apenas uma disciplina. Para atender a seu objetivo, o SAEB-99 utilizou cinco instrumentos de coleta: os questionários aplicados aos alunos, aos professores, aos diretores e às escolas amostrados, além do teste de habilidade específica aplicado aos alunos. Maiores informações sobre o SAEB-99 podem ser encontradas em Franco (2001) e em Soares *et al.* (2001b).

Este artigo utiliza os dados referentes aos alunos da 8ª série do ensino fundamental que fizeram a prova de Matemática. A amostra inicial continha 17.890 alunos distribuídos em 2.570 escolas. Destes, tiveram de ser excluídos os alunos e as respectivas escolas para os quais não foi possível localizar todos os questionários aplicados, bem como aquelas escolas em que menos de quatro alunos

foram submetidos à prova de Matemática. A segunda restrição advém do tamanho de amostra mínimo para que se possa calcular as medidas representativas de *efeito de pares* avaliadas no estudo. A Tabela 1 apresenta a distribuição espacial da amostra estudada.

Foram ajustados modelos explicativos de três níveis: aluno, escola e estado. As variáveis explicativas consideradas na análise estão descritas na Tabela 2.

O nível socioeconômico dos alunos (NSE) foi avaliado pelo Critério Brasil de Posição Econômica, medida de poder de

TABELA 1  
Distribuição da amostra efetivamente estudada

Unidade da Federação	Nº de Escolas	Nº de Alunos	Unidade da Federação	Nº de Escolas	Nº de Alunos
Rondônia	25	197	Sergipe	68	555
Acre	33	246	Bahia	85	769
Amazonas	72	651	Minas Gerais	89	737
Roraima	33	252	Espírito Santo	91	742
Pará	79	712	Rio de Janeiro	74	646
Amapá	27	230	São Paulo	67	668
Tocantins	27	184	Paraná	87	716
Maranhão	86	737	Santa Catarina	104	772
Piauí	86	648	Rio Grande do Sul	68	502
Ceará	92	708	Mato Grosso do Sul	80	578
Rio Grande do Norte	77	629	Mato Grosso	85	566
Paraíba	91	709	Goiás	94	747
Pernambuco	79	702	Distrito Federal	48	469
Alagoas	57	486			

**TOTAL: 1.904 escolas e 15.558 alunos**

TABELA 2  
Variáveis explicativas incluídas na análise

Variável	Descrição	Tipo de Medida
<b>nível aluno</b>		
SEXO	Sexo do aluno	Variável indicadora: 0: masculino 1: feminino
COR	Cor declarada pelo aluno	Variável indicadora: 1: não branco 0: branco
TRABALHA	Informação sobre trabalho do aluno	Variável indicadora: 1: trabalha 0: não trabalha
FILHO	Informação sobre filhos do aluno	Variável indicadora: 1: tem filho 0: não tem filho
NSE	Nível socioeconômico do aluno	Variável contínua determinada segundo critério ABIPEME
<b>nível escola</b>		
REDE	Rede de ensino	Variável indicadora 1: Particular 0: Pública
MÉDIA	Média de NSE na escola	
DP	Desvio Padrão de NSE na escola	

compra utilizada pela Associação Brasileira de Anunciantes. Este índice é a soma de pontos atribuídos, por meio de tabela padrão, à posse de itens de conforto doméstico, além de considerar a escolaridade do chefe de família e a contratação de serviços de empregados domésticos. Maiores detalhes sobre a variável podem ser encontrados em Soares *et al.* (2001b).

As variáveis explicativas do aluno SEXO, COR, TRABALHA e FILHO foram incluídas no modelo sem centralização. A variável NSE foi centralizada no nível socioeconômico médio da escola<sup>1</sup>. Em vista desta opção de centralização, os modelos ajustados estarão assumindo que as variáveis não ajustadas pela média têm um efeito entre as escolas que é igual ao efeito dentro da escola. Quanto à variável NSE, nos modelos em que é incluída entre as variáveis explicativas, seu efeito entre as escolas é dado pelo seu coeficiente específico. Para os modelos em que a variável não é incluída explicitamente está sendo feita a suposição de que não há efeito do nível socioeconômico na variação entre escolas.

É importante ressaltar que, embora em pequeno número, as variáveis explicativas do aluno foram suficientes para descrever a heterogeneidade existente na amostra com relação ao perfil do corpo discente.

Como alternativa para a média do NSE dos alunos de uma escola foi considerada também a mediana, e para o desvio padrão, o intervalo interquartil. No entanto, os resultados obtidos não se distinguem substantivamente daqueles obtidos para a média e o desvio padrão. Diante disso, esses resultados não são relatados neste trabalho. A semelhança entre os resultados pode ser explicada pelo fato de que, na grande parte das escolas consideradas, a média era muito próxima da mediana. Se a distribuição do nível socioeconômico dentro das escolas tivesse um perfil distinto,

apresentando maior assimetria ou pontos discrepantes, os resultados poderiam ser distintos.

## Metodologia

Considerando o objetivo do presente artigo, a análise dos dados foi feita utilizando-se de modelos de regressão hierárquica.

Nos dados do SAEB, as fontes de variabilidade são o aluno, a escola e o estado. A variabilidade devida ao aluno provém do fato de os alunos de uma mesma escola não terem todos o mesmo desempenho. Mesmo após o ajuste pelas variáveis explicativas definidas ao nível do aluno, restam diferenciais não explicados. Toda a variabilidade remanescente entre os alunos de uma mesma escola e não captada pelas variáveis explicativas de aluno fica incorporada em um primeiro termo de erro.

A segunda fonte de variabilidade são as escolas, já que elas diferem em eficiência. Ou seja, alunos semelhantes em escolas distintas têm desempenhos distintos. Mesmo que se modele a variabilidade entre as escolas, a partir de variáveis explicativas definidas neste nível, ainda restará uma parte da variação que não pode ser explicada. Esta variação é captada em um segundo termo de erro.

A incorporação da terceira fonte de variabilidade, relacionada ao estado, decorre do fato de que o ambiente em que a escola está inserida exerce influência no aprendizado e no desempenho dos alunos no teste. Assim, escolas semelhantes em todas as suas características medidas e não medidas, com alunos semelhantes, mas localizadas em estados distintos, não apresentam, necessariamente, o mesmo resultado em um teste padronizado. Este efeito do estado no desempenho do aluno é incorporado no modelo por meio de um termo de erro que sintetiza todos os

<sup>1</sup> Nos modelos hierárquicos os coeficientes do modelo de nível 1 são variáveis respostas dos modelos de níveis superiores. Daí ser importante que estes tenham interpretação. Nos modelos de regressão o intercepto não tem interpretação quando a variável explicativa não assume o valor zero. Para evitar esta dificuldade, é importante centralizar as variáveis explicativas quando se usa a metodologia de modelos hierárquicos. Para uma discussão sobre centralização ver Kreft e De Leeuw (1998).

possíveis fatores explicativos desta variabilidade.

Os modelos hierárquicos, por permitirem a incorporação ao modelo de mais de um termo de erro, tornam possível quantificar a contribuição de cada um deles para a variação total. Mais ainda, através da introdução de variáveis explicativas específicas a cada um dos níveis é possível explicar grande parte da variação presente nos dados.

Em termos matemáticos, cada modelo é escrito por meio de um conjunto de equações representando cada um dos níveis. Por exemplo, para um modelo em que não há incorporação de variáveis explicativas em nenhum dos níveis, chamado modelo nulo, temos o seguinte conjunto de equações:

$$NOTA_{ijk} = \pi_{0jk} + \varepsilon_{ijk} \tag{1}$$

$$\pi_{0jk} = \beta_{00k} + u_{0jk} \tag{2}$$

$$\beta_{00k} = \gamma_{000} + v_{00k} \tag{3}$$

As equações (1), (2) e (3) representam os modelos de nível 1, 2 e 3, respectivamente.

Estes modelos específicos para cada um dos três níveis podem ser combinados em um modelo único, representado pela equação (4):

$$NOTA_{ijk} = \gamma_{000} + v_{00k} + u_{0jk} + \varepsilon_{ijk} \tag{4}$$

O modelo nulo assume que a nota de um aluno específico é determinada por uma média geral,  $\gamma_{000}$ , que seria a média das notas de todos os alunos brasileiros, acrescida de três diferenciais. O primeiro decorre do fato de o aluno residir no estado  $k - v_{00k}$ . O segundo diferencial é específico para a escola que o aluno frequenta - a  $j$ -ésima escola do  $k$ -ésimo estado,  $u_{0jk}$ . O terceiro diferencial reflete as características do próprio aluno,  $\varepsilon_{ijk}$ . No modelo nulo não é feito nenhum esforço para determinar quais são os fatores relacionados a estes diferenciais. Os efeitos de todos os fatores são sintetizados nos termos de erro, um para cada nível.

Em geral, quando se utiliza a metodologia de modelos hierárquicos,

deseja-se explicar parte da variação observada nos diversos níveis da análise. Isto se faz mediante a inclusão de variáveis explicativas específicas de cada um dos níveis, o que leva a modelos mais elaborados. Para cada variável introduzida no nível 1 existe uma equação de nível 2. Estas equações podem ou não incluir variáveis explicativas de nível 2, bem como o termo aleatório. Da mesma forma, para cada variável explicativa introduzida no nível 2 existe uma equação de nível 3, que pode ou não incluir variáveis explicativas e o termo aleatório.

Para exemplificar, o modelo mais completo utilizado neste trabalho pode ser escrito pelo conjunto de equações (5) que segue.

$$NOTA_{ijk} = \pi_{0jk} + \pi_{1jk}SEXO + \pi_{2jk}COR + \pi_{3jk}TRABALHA + \pi_{4jk}FILHO + \pi_{5jk}NSE + \varepsilon_{ijk}$$

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}MEDIA + \beta_{02k}DP + u_{0jk}$$

$$\pi_{5jk} = \beta_{50k} + \beta_{51k}MEDIA + u_{5jk} \tag{5}$$

$$\pi_{pjk} = \beta_{p0k}, \text{ para } p \neq 0 \text{ e } 5$$

$$\beta_{00k} = \gamma_{000} + v_{00k}$$

$$\beta_{pqk} = \gamma_{pq0} \text{ para } p, q \neq 0$$

Este conjunto de equações nos diz que o desempenho de um aluno em particular é determinado a partir do desempenho médio da escola em que ele estuda - a  $j$ -ésima escola do  $k$ -ésimo estado. Esta média é modificada pelas características próprias do aluno: SEXO, COR, TRABALHA, FILHO e NSE. Existem ainda outras características do aluno que afetam seu desempenho e que não foram incluídas na análise, daí o termo de erro aleatório  $\varepsilon_{ijk}$ .

O modelo supõe que o efeito de NSE não é o mesmo em todas as escolas. Ele varia segundo o nível socioeconômico médio dos alunos que as frequentam, ou seja, existe uma interação entre o nível socioeconômico médio da escola e o nível socioeconômico do aluno. Existem, ainda, outras variáveis que modificam o efeito de NSE, mas que não foram incorporadas no modelo. O efeito destas variáveis omitidas fica sintetizado no termo aleatório  $u_{5jk}$ . As

QUADRO 1  
Definição dos modelos ajustados

Modelo	Variáveis introduzidas
Nulo	-
1	SEXO, COR, TRABALHA, FILHO, NSE centrado na média
2	MODELO 1 acrescido de PARTICULAR
3	MODELO 2 acrescido de MÉDIA
4	MODELO 3 acrescido de DP
5	MODELO 4 acrescido de NSE*MÉDIA

demais variáveis explicativas do modelo de aluno têm o mesmo efeito em todas as escolas

O desempenho médio da escola é parcialmente explicado pela média e pelo desvio padrão do nível socioeconômico dos alunos que a freqüentam. Resta ainda uma variabilidade no desempenho médio da escola que é devida a variáveis não consideradas na análise. O efeito destas variáveis não consideradas está sintetizado no termo de erro  $U_{ojk}$ .

Finalmente, resta um componente aleatório relacionado ao estado em que a escola está localizada, que é representado pelo termo  $v_{ook}$ .

Para os termos de erro  $v_{pqa}$ ,  $U_{pjk}$  e  $\epsilon_{jjk}$  assume-se distribuição normal centrada no zero, variâncias constantes em cada nível e independência entre os erros de níveis distintos. Termos de erros do mesmo nível podem ser correlacionados. Os modelos ajustados neste trabalho supõem não haver correlação entre os termos de erro do segundo nível<sup>2</sup>. Maiores detalhes sobre o uso de modelos hierárquicos em dados do SAEB podem ser encontrados em Soares *et al.* (2001a) e Barbosa e Fernandes (2001).

Neste estudo, primeiramente foi ajustado um modelo nulo – que não inclui variáveis explicativas – para se quantificar a contribuição de cada um dos níveis para a variação total das notas. Em seguida, ajustou-se o Modelo 1, que inclui somente as variáveis explicativas definidas no nível

do aluno. Este modelo propicia a estimativa do efeito Tipo A. A este modelo acrescentou-se a variável explicativa PARTICULAR, gerando o Modelo 2. Este serviu de base para o estudo da influência do efeito de pares. O Quadro 1 apresenta a definição dos modelos ajustados.

Os modelos 2 a 5 fornecem estimativas alternativas para o efeito Tipo B. O objetivo da análise é verificar se a incorporação de variáveis medindo o efeito do nível socioeconômico dos pares aumenta de forma significativa a capacidade de explicação do modelo, em relação ao modelo que contém a informação do nível socioeconômico individual e da rede na qual a escola está inserida.

A princípio, considerou-se que o efeito de pares, bem como o efeito da rede (medido pela covariável PARTICULAR), poderia modificar tanto a média da escola quanto o efeito do nível socioeconômico do aluno. Como somente a interação com a média foi significativa, os outros modelos não são apresentados.

## Resultados

Os resultados são apresentados em duas etapas. Primeiramente, ajustam-se modelos contendo as variáveis de aluno e de rede. Em seguida são apresentados e comparados os modelos que incorporam a média e/ou o desvio padrão.

O ajuste do modelo nulo forneceu os resultados apresentados na Tabela 3.

<sup>2</sup> A decisão por considerar os termos de erros não correlacionados foi tomada a partir dos resultados obtidos nas análises. Com isso, diminui-se o número de parâmetros de nível 2 a serem estimados.

**TABELA 3**  
Resultado do ajuste do modelo nulo

<b>Parâmetro</b>	<b>Estimativa</b>	<b>(Erro padrão)</b>
<b>Parte Fixa</b>		
Intercepto	251,14	(2,805)
<b>Parte Aleatória</b>		
Entre estados	193,681	(57,340)
Entre escolas	919,106	(36,931)
Entre alunos	1508,790	(18,260)
Varição total	2621,57	

**TABELA 4**  
Resultado do ajuste do modelo que considera variáveis explicativas do aluno

<b>Parâmetro</b>	<b>Estimativa</b>	<b>(Erro Padrão)</b>
<b>Parte Fixa</b>		
Intercepto	263,330	(2,698)
SEXO	-13,636	(0,658)
COR	-3,748	(0,697)
TRABALHA	-8,955	(0,848)
FILHO	-13,503	(1,749)
NSE	0,314	(0,079)
<b>Parte Aleatória</b>		
Entre estados	170,379	(50,964)
Entre escolas/intercepto	826,468	(33,717)
Entre escolas/NSE	0,613	(0,311)
Entre alunos	1445,967	(18,304)
-2LL	160708,8	

Ao considerar somente a informação sobre a escola em que o aluno estuda e o estado em que ela está localizada, temos que a variabilidade total observada nos dados é de 2621,570. Esta variação está dividida entre os três níveis, sendo que 7,4% é devida à diferenciação entre os estados, 35,1% é decorrente da variabilidade entre as escolas de um mesmo estado e os restantes 57,5% são devidos a diferenças entre os alunos de uma mesma escola.

A Tabela 4 fornece o resultado do ajuste do modelo contendo variáveis explicativas medidas no nível do aluno. O modelo considera que o intercepto varia entre as escolas e entre os estados e que o coeficiente relativo a NSE varia entre as escolas. Em outras palavras, alunos semelhantes em estados distintos apresentam desempenho médio diferente. Ademais, alunos semelhantes de um mesmo estado têm um desempenho médio que difere com relação à escola em que

estudam. Além disto, o efeito do nível socioeconômico do aluno sobre o desempenho é diferenciado por escola.

A incorporação ao modelo de variáveis medidas no nível do aluno acarretou uma redução na variabilidade em todos os níveis, porém menos intensa no nível do aluno que nos demais níveis. Para este modelo e os que seguem, não é mais possível determinar um valor único para a variabilidade total, uma vez que esta depende do valor assumido pela variável NSE. Como esta variável toma valores entre 0 e 34, temos que a variação total pode ficar entre 2442,814 e 2463,656.

Ao Modelo 1 foi acrescentada a variável explicativa PARTICULAR. Os resultados obtidos são apresentados na Tabela 5.

A variável explicativa PARTICULAR apresenta um coeficiente da ordem de 48, indicando uma grande vantagem dos alunos que freqüentam a escola particular em relação aos da rede pública.

TABELA 5  
Resultado do ajuste do modelo que considera variáveis explicativas do aluno e PARTICULAR

Parâmetro	Estimativa	(Erro Padrão)
<b>Parte Fixa</b>		
Intercepto	246,802	(2,276)
SEXO	-13,056	(0,650)
COR	-3,820	(0,687)
TRABALHA	-6,834	(0,834)
FILHO	-11,983	(1,722)
NSE	0,326	(0,079)
PARTICULAR	48,298	(1,160)
<b>Parte Aleatória</b>		
Entre estados	116,909	(34,311)
Entre escolas/intercepto	337,140	(14,488)
Entre escolas/NSE	0,638	(0,312)
Entre alunos	1443,509	(18,252)
-2LL		159468,9

Considerando que a nota média dos alunos do sexo masculino, que se autotransformam como brancos, que não trabalham, não têm filhos e têm nível socioeconômico igual à média da escola que frequentam é de 246,8 pontos, o valor 48 representa um acréscimo da ordem de 19%.

Deve-se ter cuidado na interpretação deste efeito. Não se pode interpretar a diferença entre os desempenhos médios das duas redes como resultante apenas de diferenças nas práticas pedagógicas, de gerenciamento ou outras características próprias de cada uma. Estão inseridos neste valor todas as diferenças entre as redes no que tange ao seu corpo discente e aspectos contextuais. Temos que 65,3% dos alunos da amostra são provenientes da escola pública. Entretanto, eles perfazem 72% entre os que se declaram não-brancos, 91,8% entre os que têm filhos e 87% dos que trabalham. Como estas variáveis estão todas negativamente associadas com o desempenho, quanto maior a participação de cada um dos grupos em uma dada rede, maior é o decréscimo observado na média. Além disso, a variável NSE, que é positivamente relacionada com o desempenho, tem média 12,89 entre os alunos da escola pública, contra o valor 22,45 entre os alunos da escola particular.

Como resultado desta distribuição desproporcional dos alunos entre as redes, a inclusão da variável PARTICULAR no modelo resultou em uma redução dos efeitos das outras variáveis consideradas, à exceção da variável SEXO, que tem aproximadamente a mesma distribuição entre as redes.

Também os termos de variabilidade são muito afetados pela introdução da variável explicativa PARTICULAR, à exceção da variação entre os alunos. Como este termo faz a comparação intra-escola, não se espera que ele sofra alteração com a introdução de características das escolas. Com a nova variável, há uma redução da ordem de 30% na variação entre os estados e de 60% na variação entre as escolas. Estes dados indicam a segmentação do sistema escolar brasileiro.

O Modelo 2 forma a base para a análise da incorporação das medidas de *efeito de pares*. Na Tabela 6 são apresentados os modelos ajustados a partir do Modelo 2.

A incorporação da média no modelo contendo os controles individuais e o indicador de rede levou a um aumento do poder explicativo do modelo, refletido na redução da *deviance*. O aumento de uma unidade na média do nível socioeconômico da escola está associado a um ganho médio de 3,65 pontos na prova de Matemática. Como existe uma grande variação entre as

TABELA 6  
Resultado do ajuste dos modelos que incluem efeito de pares

Parâmetro	Estimativas (Erro Padrão)					
	Modelo 3		Modelo 4		Modelo 5	
<b>Parte Fixa</b>						
Intercepto	199,409	(2,480)	204,917	(2,650)	213,171	(3,717)
SEXO	-12,727	(0,646)	-12,759	(0,646)	-12,784	(0,646)
COR	-2,317	(0,682)	-2,269	(0,681)	-2,213	(0,682)
TRABALHA	-5,628	(0,826)	-5,569	(0,825)	-5,680	(0,826)
FILHO	-11,441	(1,708)	-11,548	(1,707)	-11,680	(1,708)
NSE	0,333	(0,079)	0,333	(0,079)	-0,233	(0,196)
PARTICULAR	14,881	(1,733)	13,686	(1,731)	13,391	(1,727)
MÉDIA	3,650	(0,152)	3,821	(0,154)	2,627	(0,407)
DP			-1,897	(0,154)	-1,623	(0,355)
MÉDIA em NSE					0,034	(0,011)
<b>Parte Aleatória</b>						
Entre estados	38,001	(12,106)	35,544	(11,419)	36,584	(11,680)
Entre escolas/intercepto	217,799	(13,460)	211,782	(13,253)	207,900	(13,122)
Entre escolas/NSE	0,641	(0,313)	0,642	(0,313)	0,686	(0,315)
Entre alunos	1444,988	(18,249)	1444,905	(18,247)	1444,923	(18,249)
-2LL	158983,3		158953,5		1588943,6	

escolas no que tange ao nível socioeconômico médio, com a variável assumindo valores no intervalo [3,06; 31,17], pode-se observar uma variação de 103 pontos entre os desempenhos médios das escolas, explicada pela variação na composição das escolas com relação ao nível socioeconômico de seus alunos. A introdução da variável não acarretou mudanças nos coeficientes das variáveis explicativas de aluno. A variável PARTICULAR teve seu efeito drasticamente reduzido, passando de 48,3 para 14,9. Esta redução é decorrente da grande diferenciação entre as escolas particular e pública com relação ao perfil de seus alunos, como já foi observado anteriormente. Com relação aos termos de variação, verifica-se uma grande redução na variação entre estados, que passa de 116,9 para 38,0, e na variação entre escolas, que passa de 337,1 para 217,8.

Os Modelos 4 e 5 comprovam a necessidade de se considerar o *efeito de pares* de uma forma mais ampla. O coeficiente negativo e significativo do desvio

padrão indica que a homogeneidade da escola em relação aos seus alunos é fator importante para o desempenho da instituição. A cada aumento de um ponto no valor do desvio padrão do nível socioeconômico da escola está associada uma redução de 1,90 pontos no Modelo 4 e 1,62 no Modelo 5. Na amostra analisada, o desvio padrão assume valores no intervalo [0,5; 11,4]. Não há alteração importante nos demais coeficientes quando se introduz o desvio padrão.

O Modelo 5 indica que existe interação entre o nível socioeconômico do aluno e o da escola. Quanto mais alto o nível socioeconômico médio da escola, maior o efeito do nível socioeconômico do aluno no seu desempenho, ou seja, menor a equidade. Ao introduzir o nível socioeconômico médio interagindo com o individual, tem-se que a estimativa do individual torna-se negativa, mas não significativamente diferente de zero. Lembrando que o coeficiente é considerado aleatório, o modelo indica que o nível socioeconômico do aluno tem um efeito no desempenho que

é distribuído segundo uma distribuição normal, com média não significativamente diferente de zero.

## Discussão

Todos os modelos apresentaram ganho em relação ao modelo imediatamente anterior. Daí pode-se concluir que a inclusão de somente uma medida de centro do nível socioeconômico não é suficiente para descrever o efeito de grupo. É também necessário inserir uma medida da variabilidade para que se tenha um quadro mais bem delineado do efeito de grupo, bem como considerar a interação entre o nível socioeconômico individual e o dos pares.

Este resultado, se comprovado em outras situações, mostra que, para situações em que os alunos estão separados em escolas pelo seu nível socioeconômico, o prejuízo em termos de desempenho não é apenas duplo, como argumentado por Willms (2000), mas triplo. O aluno perde porque seu nível socioeconômico é baixo, por não conviver com colegas de nível mais elevado e pela interação entre seu baixo nível socioeconômico e o do meio em que ele convive. Alguns exemplos ilustram a situação.

Nas Tabelas 7 e 8 são apresentados os desempenhos esperados, segundo o Modelo 5, de alunos que têm o mesmo perfil, à exceção do nível socioeconômico, em algumas escolas selecionadas. As escolas foram escolhidas nos estados de Minas Gerais e Piauí. O objetivo é avaliar qual o impacto do nível socioeconômico do aluno no seu desempenho. O perfil considerado foi o do aluno do sexo feminino que se declara não-branco, trabalha e tem filho. A variável NSE, indicativa do nível socioeconômico, assumiu quatro valores, correspondentes aos percentis 1, 25, 75 e 99 da amostra de todos os alunos. Os valores são 3, 5, 11, 21 e 32. Foram escolhidas quatro escolas de Minas Gerais e quatro do Piauí, os estados que apresentaram o maior e o menor efeito de estado, respectivamente. Em cada um dos

estados, escolheu-se a escola pública e a escolar particular com menor e maior valor para a média de NSE.

Observa-se na Tabela 7 que para a ESCOLA 1, em que a média do nível socioeconômico é baixa – 7,8 –, não há muita diferença entre os desempenhos dos alunos das diversas faixas econômicas. Assim é que o escore esperado para o *aluno rico* é superior ao escore do *aluno pobre* em apenas 0,90 pontos, que corresponde a 0,43% do escore do *pobre*.

Com o aumento do nível socioeconômico médio da escola, há um distanciamento dos escores esperados para os alunos das diversas faixas econômicas. Comparando os desempenhos relativos aos alunos com NSE iguais aos percentis 1 e 99, observa-se que a diferença atinge 20,56 pontos para a escola de maior nível socioeconômico médio. Este aumento equivale a 7,21% do escore do aluno que ocupa o percentil 1 na classificação socioeconômica.

A mesma tendência de aumento é observada quando se comparam os alunos que ocupam os percentis 25 e 75. Se na escola que apresenta nível socioeconômico baixo a diferença é irrisória, representando 0,15% do escore relativo ao percentil 25, na escola de nível socioeconômico mais elevado a diferença é de 7,2 pontos, que corresponde a 2,56% do escore do aluno que ocupa o percentil 25.

A análise da Tabela 8 apresenta a mesma tendência de aumento da diferença com o aumento do nível socioeconômico da escola. É uma novidade em relação à tabela anterior: por conter informações de um estado mais pobre e com pior desempenho, evidencia que o escore esperado para um *aluno rico* pode ser inclusive inferior ao escore esperado do *aluno pobre*. Isto acontece se o nível socioeconômico da escola for muito baixo.

Comparando as Tabelas 7 e 8, percebe-se que o efeito em valor absoluto do nível socioeconômico da escola é o mesmo nos dois estados. Por exemplo, comparando a ESCOLA 2 de Minas Gerais com a ESCOLA 2 do Piauí, que têm aproximadamente o

TABELA 7  
Notas estimadas para candidatos com diferentes valores para NSE em escolas escolhidas. Minas Gerais

NSE	ESCOLA1	ESCOLA2	ESCOLA3	ESCOLA4
	$\bar{x} = 7,8$ DP=4,1 Pública	$\bar{x} = 17,2$ DP=4,1 Particular	$\bar{x} = 21,28$ DP=6,8 Pública	$\bar{x} = 28,1$ DP=3,0 Particular
3,5	210,10	244,78	227,59	276,78
11	210,33	247,43	231,25	282,19
21	210,65	250,95	236,12	289,40
32	211,00	254,83	241,48	297,34
Dif1*	0,90	10,05	13,89	20,56
Dif2**	0,32	3,88	4,87	7,21

\*Dif1 é a diferença entre os escores relativos aos dos alunos com NSE 3,5 e 32.

\*\*Dif2 é a diferença entre os escores relativos aos dos alunos com NSE 11 e 21.

TABELA 8  
Notas estimadas para candidatos com diferentes valores para NSE em escolas escolhidas. Piauí

NSE	ESCOLA1	ESCOLA2	ESCOLA3	ESCOLA4
	$\bar{x} = 5,6$ DP=3,0 Pública	$\bar{x} = 17,5$ DP=2,4 Particular	$\bar{x} = 18$ DP=6,0 Pública	$\bar{x} = 25,5$ DP=3,7 Particular
3,5	187,30	209,77	181,29	221,86
11	186,97	212,48	184,14	226,61
21	186,54	216,10	187,93	232,94
32	186,01	220,08	192,10	239,94
Dif1*	-1,29	10,31	10,81	18,08
Dif2**	-0,43	3,62	3,79	6,33

\*Dif1 é a diferença entre os escores relativos aos dos alunos com NSE 3,5 e 32.

\*\*Dif2 é a diferença entre os escores relativos aos dos alunos com NSE 11 e 21.

mesmo nível socioeconômico médio, a diferença entre os alunos dos percentis 1 e 99 é de aproximadamente 10 pontos nos dois estados. Entretanto, como o Piauí tem pior desempenho, a distância relativa entre os escores dos alunos é maior neste estado. Este fato fica mais bem evidenciado quando se compara a ESCOLA 4 de Minas Gerais com a ESCOLA 4 do Piauí. Na escola do Piauí, a diferença entre os escores dos alunos ocupantes dos percentis 1 e 99 é menor que em Minas Gerais. Entretanto, a diferença de 18,08 pontos observada no Piauí equivale a 8,1% da nota do aluno do percentil 1. Em Minas Gerais, a diferença é de 20,56, que equivale a 7,4% da nota do aluno do percentil 1.

Sintetizando, o *aluno rico* na *escola pobre* sofre o efeito deletério do ambiente. Seu escore é próximo ao do *aluno pobre*, podendo ser um pouco superior ou um pouco inferior, dependendo do grau de pobreza da escola. O *aluno pobre* na escola

rica é beneficiado pelo meio. Entretanto, o ganho do *aluno pobre* é muito inferior à perda do *aluno rico*, uma vez que a distância entre os escores dos *ricos* e dos *pobres* aumenta sistematicamente com o aumento da média.

Estes resultados mostram por que, em democracias mais estáveis, o ensino público não permite a criação de escolas com grande concentração socioeconômica. A única solução para não prejudicar todos os alunos é fazer com que as diferentes escolas tenham, aproximadamente, o mesmo nível médio socioeconômico e pequena variabilidade.

Não se está sugerindo que a simples transposição de *alunos pobres* para *escolas ricas* levará a um sistema escolar mais justo. Foi feito um exercício de simulação supondo que as equações de regressão encontradas se aplicam à situação simulada. Como não se dispõe de dados sobre desempenho nas situações simuladas, não há como verificar

a plausibilidade da suposição. A solução para a busca de maior equidade pode ser mais complexa, alterando não somente o perfil dos alunos intra-escola, mas também introduzindo mudanças no processo ensino-aprendizagem que levem a uma maior taxa de retorno para os alunos de menor habilidade.

### Referências bibliográficas

AITKIN, Murray e LONGFORD, Nicholas T. Statistical modelling issues in school effectiveness studies. **Journal of the Royal Statistical Society, Series A**, v. 149, p. 419-461, 1986.

BARBOSA, Maria Eugênia Ferrão e FERNANDES, Cristiano. A escola brasileira faz diferença? Uma investigação dos efeitos da escola na proficiência em Matemática dos alunos da 4ª série. In: FRANCO, Creso (org.). **Avaliação, ciclos e promoção na educação**. Porto Alegre: Artmed Editora, 2001. 162 p.

BRYK, Anthony S. e RAUDENBUSH, Stephen W. **Hierarchical linear models**. Newbury Park: Sage, 1992. 265 p.

FRANCO, Creso. Iniciativas recentes de avaliação da qualidade da educação no Brasil. In: FRANCO, Creso (org.). **Avaliação, ciclos e promoção na educação**. Porto Alegre: Artmed Editora, 2001. 162 pág.

GOLDSTEIN, Harvey. **Multilevel statistical models**. 2ª ed. Londres: Edward Arnold, 1995. 178p.

\_\_\_\_\_. Modelos da realidade: novas abordagens para a compreensão de processos educacionais. In: FRANCO, C. (org.). **Avaliação, ciclos e promoção na educação**. Porto Alegre: Artmed Editora, 2001. 162 p.

HOLLAND, Paul W. Statistics and casual inference. **Journal of the American Statistical Association**, v. 81, n. 306, p. 945-960, 1986.

KREFT, Ita e DE LEEUW, Jan. **Introducing multilevel modeling**. Londres: Sage, 1998. 149 p.

Finalmente, a análise evidenciou a necessidade de se considerar o *efeito de pares* de forma mais ampla do que simplesmente introduzir uma medida de desempenho prévio ou de nível socioeconômico médio. A incorporação de modelos como o 5 deve ser padrão em estudos realizados para o cálculo do efeito Tipo B.

LEE, Valerie E. e BRYK, Anthony S. A multilevel model of the social distribution of high school achievement. **Sociology of Education**, v. 62, p. 172-192, 1989.

LONGFORD, Nicholas. Contextual effects and group means. **Multilevel Modelling Newsletter**, v. 1(3), p. 5 e 11, 1989.

PLEWIS, Ian. Comment on “‘Centering’ predictors in multilevel analysis: choices and consequences”. **Multilevel Modelling Newsletter**, v. 1(3), p. 6 e 11, 1989.

RAUDENBUSH, Stephen W. “Centering” predictors in multilevel analysis: choices and consequences. **Multilevel Modelling Newsletter**, v. 1(2), p. 10-12, 1989.

RAUDENBUSH, Stephen W. e WILLMS, J. Douglas. The estimation of school effects. **Journal of Educational and Behavioral Statistics**, v. 20(4), p. 307-335, 1995.

SAMMONS, Pam. **School effectiveness: coming of age in the twenty-first century**. Lisse: Swets & Zeitlinger Publishers, 1999.

SNIJDERS, Tom A. B. e BOSKER, Roel J. **Multilevel analysis: an introduction to basic and advanced multilevel modelling**. Londres: Sage Publications, 1999. 266 p.

SOARES, José Francisco, CÉSAR, Cibele Comini e MAMBRINI, Juliana. Determinantes de desempenho dos alunos do ensino básico brasileiro: evidências do SAEB de 1977. In: FRANCO, Creso (org.). **Avaliação, ciclos e promoção na educação**. Porto Alegre: Artmed Editora, 2001a. 162 p.

SOARES, José Francisco, BATISTA, José Rodrigues, ALVES, Maria Teresa Gonzaga e TEIXEIRA, Ana Cristina Azeredo. **Fatores explicativos do desempenho em Língua**

**Portuguesa e Matemática:** a evidência do SAEB-99. Mimeo, 2001b. 132 p.

THOMAS, Sally e MORTIMORE, Peter. Comparison of value-added models for secondary-school effectiveness. **Research Papers in Education**, 11(1), p. 5-33, 1996.

WILLMS, J. Douglas. **Monitoring school performance:** a guide to educators. Londres: The Falmer Press, 1992. 183 p.

\_\_\_\_\_. Monitoring school performance for "standards-based reform". **Evaluation & Research in Education**, 14 (3-4), p. 237-253, 2000.

### Abstract

Usually, studies of academic achievement consider two broad classes of factors: student factors (gender, socio-economics status, etc.) and school factors (sector, pupil-teacher ratio, teacher training, etc.). In addition to these factors it is important to consider the effect of the whole class upon the achievement of a specific student, known as *compositional*, *contextual* or *peer effects*. Although the *peer effect* have been represented by the group means, it is not obvious that it is the best way of introducing it. The article evaluates alternatives forms of inclusion of the *peer effect* in a model of academic achievement. Using the results of the Mathematics test of SAEB-99, it is showed that group means are not enough to capture all the effects of the peers.

Enviado para publicação em 10/10/2001.