

DEPARTAMENTO DE ECONOMIA
PUC-RIO

TEXTO PARA DISCUSSÃO
Nº. 455

QUALIDADE E EQÜIDADE NA EDUCAÇÃO FUNDAMENTAL
BRASILEIRA

ÂNGELA ALBERNAZ
FRANCISCO H.G. FERREIRA
CRESO FRANCO

MAIO 2002

Qualidade e Equidade na Educação Fundamental Brasileira

Ângela Albernaz¹
Francisco H.G. Ferreira¹
Creso Franco²

Palavras Chave: Qualidade da Educação; Funções de Produção de Educação; Modelos Hierárquicos Lineares

Código JEL: I21

Resumo: Este artigo estima uma função de produção educacional para o ensino fundamental brasileiro, com base nos dados do Sistema de Avaliação da Educação Básica (SAEB) para as 8^{as} séries em 1999. Utilizando modelos hierárquicos lineares, de forma a evitar vieses de agregação, a estimação sugere que cerca de oitenta por cento da variância em desempenho médio entre as escolas deve-se a diferenças na composição sócio-econômica de seus alunos. Como em outros estudos internacionais, a rede a que a escola pertence está relacionada de forma importante com o desempenho de seus alunos. Não obstante, e em contraste com a maior parte da literatura internacional, outras variáveis escolares, tais como a escolaridade dos professores e a qualidade da infraestrutura física, também afetam o desempenho dos estudantes. Alguns fatores positivos para a eficácia média das escolas reduzem a sua "equidade".

¹ Departamento de Economia, PUC-Rio.

² Departamento de Educação, PUC-Rio.

Os autores agradecem ao Prof. Naércio Menezes Filho por suas excelentes sugestões. É claro que continuamos sendo os únicos responsáveis pelos erros que porventura permaneçam no artigo.

1. Introdução

O sistema educacional de um país é um conjunto de instituições fundamental para o seu desenvolvimento sócio-econômico. Desde Arrow (1962) e Uzawa (1965), passando por Lucas (1988) e Becker, Murphy & Tamura (1990), a acumulação de capital humano em geral, e a educação em particular, ocupam posição central em nossa visão sobre o processo de crescimento econômico. A educação também exerce um papel crucial na determinação da distribuição dos frutos desse crescimento, como sabemos desde Mincer (1958), passando por Tinbergen (1975). No Brasil, em particular, a importância da educação em explicar nosso alto nível de desigualdade vem sendo apontada desde Langoni (1973). Ver também Barros, Henriques e Mendonça (2000).

É natural, portanto, que uma melhor compreensão dos determinantes do desempenho educacional dos alunos brasileiros seja de grande interesse acadêmico e de política pública. Em particular, interessa quantificar os efeitos dos vários insumos educacionais - tais como a escolaridade e a experiência dos professores, bem como vários aspectos da infraestrutura escolar - sobre o aprendizado. O objetivo deste trabalho é contribuir para esta maior compreensão, através da estimação de uma função de produção educacional para o ensino fundamental brasileiro.

A "função de produção educacional" é o nome dado por economistas à relação existente entre uma série de "insumos" ao processo educacional e o seu "produto". Desde a publicação do Relatório Coleman sobre a qualidade das escolas públicas e privadas nos Estados Unidos (Coleman et. al.,1966), desenvolveu-se uma literatura considerável sobre o tema, tanto em economia como na área da educação. A forma geral da função de produção educacional é dada pela equação 1:

$$(1) \quad y = F(c, m, g, p, s)$$

Em (1), y denota o desempenho dos alunos, que pode depender de uma série de fatores, agrupados em cinco categorias: características pessoais do aluno (c), tais como raça e gênero; características de suas famílias (m), tais como renda ou outra medida de seu nível sócio-econômico; características de seus colegas na escola (g); características dos seus professores (p), tais como escolaridade, salário ou experiência; e outras características escolares (s).³

³ Ver Hanushek (1986) para uma resenha clássica desta literatura.

No caso brasileiro, a análise econômica desta relação entre insumos e produto do processo educacional sofria, tradicionalmente, de um sério problema, referente à mensuração da variável produto (y). Até a publicação dos primeiros resultados dos exames padronizados do Sistema de Acompanhamento da Educação Básica (SAEB), em 1997, a única variável disponível para medir desempenho educacional era o número de séries completas pelo aluno (anos de escolaridade). Ainda que estudos baseados nesta variável tenham contribuído para uma compreensão inicial da problemática educacional brasileira (ver Barros, Henriques e Santos, 1999), os mesmos sempre reconheceram as severas limitações da variável "anos de escolaridade" como medida de aprendizado.⁴

A disponibilização dos micro-dados do SAEB⁵ tornou possível, pela primeira vez no Brasil, a investigação dos determinantes de uma medida de desempenho escolar baseada em rendimentos de alunos em testes padronizados de conhecimento. Além de incluir informações sobre o rendimento dos alunos nestes testes, a base de dados do SAEB inclui ainda informação acerca de todas as outras categorias de determinantes incluídos na equação (1), constituindo assim um recurso potencialmente muito valioso para o estudo da eficácia, eficiência e equidade da escola brasileira. Estudos na área de educação - como Fletcher (1997), Barbosa e Fernandes (2001) e Franco, Mandarinó e Ortigão (2001) - têm confirmado a importância desta base de dados.

Neste artigo, nós estimamos uma função de produção educacional para o Brasil, aplicando modelos hierárquicos lineares aos micro-dados do SAEB de 1999. Ao contrário de qualquer estudo anterior de que tenhamos conhecimento, investigamos a contribuição de diferentes variáveis escolares (e de professores) tanto sobre uma medida de eficácia, quanto para uma medida de equidade da prática educacional brasileira.

Entre os resultados encontrados, destacamos três. Primeiro, como em outros países, a variância em desempenho entre as escolas brasileiras deve-se principalmente a diferenças no nível sócio-econômico médio de seus alunos, refletindo um importante efeito de seleção da clientela. Segundo, uma vez controlado este efeito, diferenças na quantidade e qualidade dos insumos escolares ainda respondem por uma parcela significativa da diferença de desempenho entre as escolas. Ao contrário de resultados encontrados para vários outros países, tanto a

⁴ O principal problema com esta variável é que ela não inclui qualquer informação sobre a qualidade de cada ano de estudo.

⁵ Assim como de seu equivalente para o ensino superior (o "Provão").

qualidade dos professores quanto a qualidade da infraestrutura física das escolas afetam o rendimento de forma significativa.⁶ Terceiro, mesmo controlando para todos os fatores acima, no Brasil, o desempenho médio da escola particular supera o da escola pública.

O trabalho está organizado em quatro seções além desta introdução. Na próxima seção, apresentamos os dados e a metodologia utilizada. Na terceira seção, discutem-se os resultados para o modelo estimado sem as características escolares. A quarta seção apresenta o modelo completo. A quinta seção apresenta nossas conclusões.

2. Metodologia e Dados

O presente artigo utiliza a base de dados do SAEB de 1999, uma amostra representativa nacional de alunos e de escolas do país. Em particular, utilizamos a amostra da 8ª série, que inclui informações sobre o desempenho em Ciências, Geografia, História, Português e Matemática de 89.671 alunos pertencentes a 2.588 escolas. O tamanho médio da amostra é de aproximadamente 35 alunos por escola. Este sistema de avaliação oferece uma série de informações que vai desde o desempenho dos alunos, medido pela nota obtida no teste, até as características dos próprios alunos, dos seus professores e das escolas onde estes estudam. A disponibilidade desses resultados nos incentiva a avaliar quais as características dos alunos, de seus professores e da escola onde eles estudam que favorecem um maior desempenho dos mesmos.

A Tabela 1 apresenta as médias e os desvios-padrão das variáveis de rendimento escolar e nível sócio-econômico. O Apêndice 1 contém estatísticas descritivas para as variáveis escolares. Os erros-padrão reportados levam em consideração o desenho amostral com conglomerados e estratificação.⁷ A média de Matemática foi escolhida como a média base para todas as matérias. Portanto, os valores reportados para as demais disciplinas representam a variação das médias dessas disciplinas em relação à média de Matemática.

⁶ A pouca relevância das variáveis educacionais tem sido uma característica polêmica, mas persistente, da literatura internacional sobre funções de produção de educação. De acordo com a tabela 3.23.1, p. 303 do Relatório de Coleman, menos de 2% da variância total do desempenho dos alunos (brancos e negros) é atribuída a características escolares, enquanto que menos de 4% é atribuída a características dos professores. Numa resenha mais recente, Hanushek (1989) apresenta um resumo das estimativas dos coeficientes dos gastos escolares sobre o desempenho dos alunos em 187 estudos, chegando à conclusão de que a única variável cuja relevância para o aprendizado dos alunos parece ser realmente robusta é a experiência do professor.

⁷ A amostra do SAEB é complexa no sentido que ela envolve estratificação e conglomerados. Os estratos são constituídos pelos estados mais o distrito federal e o tipo de rede ao qual a escola pertence e os conglomerados são representados pelas escolas sorteadas para compor a amostra.

Tabela 1 Estatísticas Descritivas para as Provas Realizadas pelos alunos da 8ª série

	Variável	Média	Desvio Padrão
Variáveis ao Nível do Aluno			
Desempenho em Matemática	Y_{ij}	246.35	1.11
Desempenho em Ciências	$(dcie)_{ij}$	-2.34	0.92
Desempenho em Geografia	$(dgeo)_{ij}$	3.68	0.97
Desempenho em História	$(dhit)_{ij}$	3.67	0.85
Desempenho em Português	$(dport)_{ij}$	-13.5	0.93
Nível Sócio-Econômico	$(nse)_{ij}$	0.00	1.00
Variáveis ao Nível da Escola			
Nível Sócio Econômico Médio da Clientela	$(nsemédio)_j$	0.02	0.77

Como qualquer outra base de dados, o SAEB não é perfeito. Entre seus problemas, do ponto de vista da análise econômica, está a ausência de uma variável confiável de renda da família do aluno. Dada essa ausência, adotamos neste trabalho uma prática comum na análise educacional, que é a construção de uma variável "proxy" do nível sócio-econômico do aluno, através da extração do componente principal de nove variáveis derivadas de respostas a perguntas no questionário familiar do SAEB. O procedimento adotado para a extração deste componente principal está descrito no Apêndice 2. A variável "nível sócio-econômico médio da clientela" na Tabela 1 acima reflete a média global das médias escolares deste fator.

Um segundo problema com os dados do SAEB é o fato de que seu desenho não permite a criação de uma variável de *diferencial* dos testes de rendimento (por exemplo, da 7ª para a 8ª série).⁸ Tal estimativa - chamada de medida de valor agregado da educação no período de tempo relevante - é reconhecida teoricamente como a mais apropriada, pois elimina efeitos fixos dos alunos, e corresponde melhor ao conceito de ganho com a educação contemporânea do aluno. Ver Hanushek (1989). Na sua ausência, utilizamos a melhor variável disponível, que é o *nível* do teste de rendimento do aluno.

Passamos agora a uma breve apresentação da metodologia econométrica a ser utilizada na estimação da equação (1), com base nos dados do SAEB, que é a construção de um modelo hierárquico linear. Os modelos hierárquicos lineares, também conhecidos como modelos

multinível, são modelos de coeficientes aleatórios, apropriados para a análise de variáveis aleatórias cuja distribuições sobre a população não são independentes e idênticas, mas sim agrupadas de forma não aleatória em sub-grupos relevantes. Dados educacionais são um exemplo clássico, já que compõem-se de variáveis descrevendo os alunos, que estão agrupados em unidades maiores - as escolas - onde interagem entre si. Assim sendo, é demasiado restritivo impor que os coeficientes relacionando as variáveis independentes à variável dependente sejam constantes entre os sub-grupos (no caso educacional, entre as escolas), ou que os resíduos sejam i.i.d..

Os modelos hierárquicos lineares combinam a flexibilidade da estimação de equações separadas para cada escola com o fato de que a variação destes coeficientes entre as escolas pode ser estimado num segundo nível do modelo. O modelo hierárquico geral de dois níveis⁹ é representado pelas seguintes equações:

$$\text{Nível 1: } Y_{ij} = \mathbf{b}_{0j} + \sum_{q=1}^Q \mathbf{b}_{qj} X_{qij} + r_{ij}, \text{ onde: } r_{ij} \sim N(0, \mathbf{s}^2)$$

$$\text{Nível 2: } \mathbf{b}_{qj} = \mathbf{g}_{q0} + \sum_{s=1}^{S_q} \mathbf{g}_{qs} W_{sj} + u_{qj}, \text{ onde } q=0, \dots, Q$$

Assume-se que cada componente aleatório u_{qj} , $q=0, \dots, Q$, segue uma distribuição normal multivariada de média zero e $\text{var}(u_{qj}) = \mathbf{t}_{qq}$. Para qualquer par de coeficientes aleatórios q e q' , $\text{cov}(u_{qj}, u_{q'j}) = \mathbf{t}_{qq'}$. Os erros de nível 1 e de nível 2 são supostos independentes: $\text{cov}(r_{ij}, u_{qj}) = 0, \forall_{q,i,j}$. O algoritmo de estimação do modelo baseia-se na maximização da função de máxima verossimilhança conjunta dos dois níveis. As referências estatísticas clássicas para estes modelos são Bryk & Raudembush (1992) e Goldstein (1995).

No caso específico da estimação de funções de produção educacional, a aplicação destes modelos, além de estatisticamente mais recomendável, tem efeitos potencialmente sérios sobre

⁸ No Brasil, a única base de dados compatível com a estimação de valor agregado é a base do SARESP. A mesma não se encontra disponível para uso dos pesquisadores.

uma das conclusões mais importantes da literatura, a saber: a pouca importância relativa das variáveis educacionais vis-à-vis as variáveis familiares na determinação do desempenho escolar. Diversos estudos que utilizam mínimos quadrados ordinários (MQO) apontam a existência de uma forte relação entre as características dos alunos e de suas famílias, e o desempenho escolar (Summers e Wolfe, 1977; Hanushek 1989; Mizala e Romaguera, 1988; Mizala, Reinaga e Romaguera, 1999; Barros, Mendonça e Santos, 1999), mas concluem que não há uma relação robusta entre as variáveis escolares e as características dos professores, por um lado, e o rendimento dos alunos por outro (Hanushek, 1989).

Há alguma evidência, porém, de que este resultado pode dever-se, pelo menos em parte, ao tratamento inadequado dispensado pelas estimações por MQO à natureza hierárquica dos dados educacionais. Algumas das variáveis que descrevem a escola são variáveis agregadas dos estudantes que pertencem a esta escola. Como exemplo, pode-se mencionar: o nível sócio econômico médio dos estudantes da escola, a escolaridade média dos professores da escola, o número de livros presentes na biblioteca, e assim por diante.

Para ilustrar as possíveis consequências de conduzir a análise em um único nível, sem considerar a estrutura dos dados, Bidwell e Kasarda (1980) propõem um modelo teórico, onde o nível sócio-econômico dos alunos (X_1) e os recursos escolares (X_2) dessem contribuições da mesma magnitude na explicação do desempenho obtido pelos alunos (Y)¹⁰. A partir do modelo teórico, os autores simulam uma “pseudo base de dados” de 2500 alunos e atribuem a cada aluno valores individuais para o nível sócio econômico, assim como valores para os recursos escolares. Os principais resultados das simulações daquele estudo - que encontram-se reproduzidos na Figura 1 - indicam um viés sistemático de subestimação do efeito da variável agregada. No caso dos estudos citados anteriormente, este resultado corresponderia à subestimação das variáveis ligadas às escolas e aos professores.

A introdução de modelos hierárquicos lineares (HLM), em substituição à estimação por MQO, poderia alterar tais resultados. “Pela sua própria natureza, perguntas a respeito do efeito das escolas sobre o desempenho dos alunos recaem em uma investigação de relações hierárquicas. Tal investigação envolve a procura de associações estatísticas entre os fatores

⁹ Quando os dados permitem, é freqüente a utilização de modelos hierárquicos de três níveis (alunos, turmas, escolas) para estimar este tipo. Como o número de alunos por turma na base do SAEB é demasiado pequeno para permitir esta desagregação, optou-se por trabalhar somente com dois níveis.

¹⁰ As variâncias de X_1 , X_2 e Y assumiram valores tipicamente reportados em pesquisas educacionais.

escolares e as variáveis ao nível do aluno. Os recentes avanços na teoria estatística dos modelos hierárquicos lineares - HLM, fornecem ferramentas apropriadas para modelar efeitos entre as escolas e dentro das mesmas”. (Bryk e Raudenbush, 1986).

Tabela 2.1: Resultados das Análises

Agrupamento Aleatório	Parcialmente por X_1
i. $X_1 \xrightarrow{0,221} Y$ $X_2 \xrightarrow{0,213} Y$	i. $X_1 \xrightarrow{0,221} Y$ $X_2 \xrightarrow{0,213} Y$
ii. $X_1 \xrightarrow{0,262} Y$ $\bar{X}_2 \xrightarrow{0,038} Y$	ii. $X_1 \xrightarrow{0,262} Y$ $\bar{X}_2 \xrightarrow{0,014} Y$
iii. $\bar{X}_1 \xrightarrow{0,031} Y$ $X_2 \xrightarrow{0,256} Y$	iii. $\bar{X}_1 \xrightarrow{0,103} Y$ $X_2 \xrightarrow{0,248} Y$
iv. $\bar{X}_1 \xrightarrow{0,032} Y$ $\bar{X}_2 \xrightarrow{0,039} Y$	iv. $\bar{X}_1 \xrightarrow{0,116} Y$ $\bar{X}_2 \xrightarrow{0,018} Y$

Legenda: Coeficientes estimados para o efeito de cada variável construída (X_1 e X_2) sobre Y , quando a variável denotada por uma barra superior é agregada. A pseudo base de dados foi construída de forma a gerar os valores "verdadeiros" em (i). A primeira coluna refere-se a simulações onde a agregação da variável com barra é feita de forma aleatória. A segunda coluna refere-se a simulações onde a agregação é estratificada de acordo com valores de X_1 .

A idéia deste tipo de análise é considerar simultaneamente o impacto das características dos alunos e das escolas no desempenho do aluno, levando em conta também que alunos que pertencem à mesma escola não são independentes uns dos outros.

No restante deste artigo, construímos um modelo hierárquico linear “de baixo para cima”, ou seja, do modelo mais simples possível até chegar ao modelo completo, onde são identificadas as características escolares capazes de tornar a escola brasileira mais eficaz e equitativa. Na seção 3, apresentamos brevemente o modelo incondicional e um modelo com todas as características individuais disponíveis para cada aluno. Nesta seção buscamos respostas para três perguntas:

1. Que proporção da variância dos rendimentos nos testes do SAEB dá-se entre escolas, e que parte disto deve-se à composição da clientela, versus a características da própria escola?
2. Qual o efeito do nível sócio-econômico do aluno sobre o seu desempenho?
3. O grau de associação entre o nível sócio-econômico do aluno e seu desempenho é o mesmo em todas as escolas?

Na seção 4, expandimos o modelo para incluir variáveis escolares, e analisamos os resultados referentes às mesmas.

3. Um Modelo Hierárquico do Rendimento Escolar

O modelo hierárquico mais simples é conhecido como modelo incondicional, que representa o desempenho do aluno como a soma do desempenho médio da escola, \mathbf{b}_{0j} ; e de um resíduo, r_{ij} , que depende do aluno e da escola. Pressupõe-se que este resíduo siga uma distribuição normal com média zero e variância constante σ^2 . Como cada aluno foi testado em uma disciplina apenas e é de nosso interesse considerar o maior número possível de alunos por escola, a fim de tornar nossas estimativas mais confiáveis, foram incluídas variáveis dicotômicas ("dummies") na equação do nível 1. Essa equação pode ser assim expressa:

$$(2) \quad Y_{ij} = \mathbf{b}_{0j} + \mathbf{b}_{1j} * dcie_{ij} + \mathbf{b}_{2j} * dgeo_{ij} + \mathbf{b}_{3j} * dhit_{ij} + \mathbf{b}_{4j} * dport_{ij} + r_{ij}$$

No nível da escola, representa-se o desempenho médio da escola em Matemática, \mathbf{b}_{0j} , como função da média geral de todas as escolas, (\mathbf{g}_{00}), mais um componente aleatório, u_{0j} . Em um primeiro momento, considera-se que o desempenho médio das escolas nas demais disciplinas não varia de escola para escola, e, portanto, estas não apresentam um termo de erro aleatório¹¹. A equação de nível 2 pode ser observada a seguir:

¹¹ O desempenho médio das escolas nas demais disciplinas é considerado fixo, porque não dispomos de um grande número de alunos dentro de cada escola.

$$\begin{aligned}
 \mathbf{b}_{0j} &= \mathbf{g}_{00} + u_{0j} \\
 \mathbf{b}_{1j} &= \mathbf{g}_{10} \\
 \mathbf{b}_{2j} &= \mathbf{g}_{20} \\
 \mathbf{b}_{3j} &= \mathbf{g}_{30} \\
 \mathbf{b}_{4j} &= \mathbf{g}_{40}
 \end{aligned}
 \tag{3}$$

Novamente, supõe-se que u_{0j} se comporta como um erro aleatório, distribuído normalmente com média zero e variância \mathbf{t}_{00} . Os valores desses coeficientes encontram-se na tabela 2.

Tabela 2: Modelo Incondicional¹²

Efeitos Fixos		Coefficiente	Erro Padrão	Teste t	Valor-p
Média das escolas em Matemática,	\mathbf{g}_{00}	253.1	0.65	385.9	0.00
Média das escolas em Ciências,	\mathbf{g}_{10}	-2.5	0.44	-5.7	0.00
Média das escolas em Geografia,	\mathbf{g}_{20}	3.5	0.44	8.1	0.00
Média das escolas em História,	\mathbf{g}_{30}	3.3	0.44	7.6	0.00
Média das escolas em Português,	\mathbf{g}_{40}	-14.2	0.44	-32.1	0.00

Efeitos Aleatórios		Variância	Graus de Liberdade	\mathbf{c}^2	Valor-p
Média da escola,	u_{0j}	733.9	2578	44727.6	0.00
Efeito do aluno,	r_{ij}	1846.7			

A tabela 2 contém dois resultados de interesse. Primeiro, a variância de \mathbf{b}_{0j} ao redor de $\tilde{\alpha}_{00}$ é alta ($\hat{\mathbf{t}}_{00}=733.9$) e estatisticamente significativa. Desse modo, a hipótese nula de nenhuma diferença entre os coeficientes de intercepto das escolas é descartada, ou seja, a eficácia das escolas varia significativamente em torno da média nacional. Em segundo lugar, a correlação

¹² A terminologia habitual na literatura de HLM refere-se aos coeficientes estimados como "efeitos fixos", e às estimativas de variância como "efeitos aleatórios". Os termos são utilizados aqui com esta interpretação, e não aquela associada à econometria de dados em painel.

intra-classe é dada por $\hat{r} = \hat{t}_{00} / (\hat{t}_{00} + \hat{s}^2) = 733.9 / (733.9 + 1846.7) = 0.28$.¹³ Ou seja, 28% da variância total do desempenho estudantil ocorre entre as escolas, o que torna recomendável a utilização dos Modelos Hierárquicos Lineares na estimação da função de produção educacional brasileira.¹⁴

Os próximos passos a serem tomados buscam entender quais as características dos alunos e das escolas tornam as últimas mais eficazes e equitativas. Inicialmente são incluídas seis variáveis de controle no nível dos alunos: três variáveis dicotômicas ligadas a repetência, cada uma representando o número total de anos que o aluno repetiu; duas variáveis dicotômicas relacionadas a sua raça; e, o sexo dos alunos. A inclusão da variável de repetência dos alunos é motivada pelo fato da variável dependente - o desempenho do aluno, Y_{ij} - ser uma medida única em todo o seu percurso escolar. Assim, a nota obtida no teste padrão reflete todo o seu aprendizado durante anos de permanência na escola até o momento do teste¹⁵. Quando as variáveis de repetência são incluídas, elas espelham o percurso escolar dos indivíduos que fizeram o teste. A inclusão das variáveis de sexo e de raça no nível dos alunos tem por objetivo controlar possíveis efeitos do gênero e da etnia sobre o seu desempenho.

O novo modelo de nível 1 incorpora também o efeito do nível sócio-econômico no desempenho dos alunos. A variável "nível sócio-econômico" (NSE) foi construída a partir das respostas dos alunos quanto à educação dos seus pais, como também a partir das suas respostas sobre os itens de infra-estrutura e bens de consumo aos quais eles têm acesso, presentes no questionário dos alunos, conforme explicitado no Apêndice 2. O novo modelo de nível 1 é representado pela equação (4) abaixo:

¹³ A correlação intraclassa representa a proporção da variância total que corresponde à variância entre as escolas. Os valores tipicamente encontrados para a correlação intraclassa na literatura dos Modelos Hierárquicos estão entre 0,20 e 0,25.

¹⁴ A partir do modelo incondicional calculou-se também o índice de fidedignidade das médias escolares, que ficou em 0,808. Este resultado indica que a informação utilizada pelo modelo multinível para calcular a estimativa da média escolar provém, em média, 80% da própria escola e 20% de outras escolas.

¹⁵ Como comentado na seção 2, o mais correto seria utilizar uma estimativa de valor agregado para a função de produção educacional. No entanto, ainda não se encontra disponível para pesquisa uma base de dados brasileira que conte com duas medidas de desempenho de um mesmo aluno em dois pontos no tempo.

$$\begin{aligned}
Y_{ij} = & \mathbf{b}_{0j} + \mathbf{b}_{1j} * (\mathit{masc})_{ij} + \mathbf{b}_{2j} * (\mathit{dcie})_{ij} + \mathbf{b}_{3j} * (\mathit{dgeo})_{ij} + \mathbf{b}_{4j} * (\mathit{dhit})_{ij} + \mathbf{b}_{5j} * (\mathit{dport})_{ij} + \mathbf{b}_{6j} * (\mathit{nse})_{ij} + \\
& \mathbf{b}_{7j} * (\mathit{repete_1})_{ij} + \mathbf{b}_{8j} * (\mathit{repete_2})_{ij} + \mathbf{b}_{9j} * (\mathit{repete_3})_{ij} + \mathbf{b}_{10j} * (\mathit{repete_4})_{ij} + \mathbf{b}_{11j} * (\mathit{pardo})_{ij} + \\
& \mathbf{b}_{12j} * (\mathit{negro})_{ij} + r_{ij}
\end{aligned}
\tag{4}$$

O modelo de nível 2, por sua vez, incorpora agora o efeito do nível sócio-econômico médio da clientela da escola, $(\mathit{nsemédio})_j$, sobre o desempenho médio da escola \mathbf{b}_{0j} e sobre o nível sócio-econômico dos alunos, \mathbf{b}_{6j} ¹⁶. A equação de nível 2 pode ser assim expressa:

$$\begin{aligned}
\mathbf{b}_{0j} &= \mathbf{g}_{00} + \mathbf{g}_{01} * (\mathit{nsemédio})_j + u_{0j} \\
\mathbf{b}_{1j} &= \mathbf{g}_{10} \\
\mathbf{b}_{2j} &= \mathbf{g}_{20} \\
\mathbf{b}_{3j} &= \mathbf{g}_{30} \\
\mathbf{b}_{4j} &= \mathbf{g}_{40} \\
\mathbf{b}_{5j} &= \mathbf{g}_{50} \\
\mathbf{b}_{6j} &= \mathbf{g}_{60} + \mathbf{g}_{61} * (\mathit{nsemédio})_j + u_{6j} \\
\mathbf{b}_{7j} &= \mathbf{g}_{70} \\
\mathbf{b}_{8j} &= \mathbf{g}_{80} \\
\mathbf{b}_{9j} &= \mathbf{g}_{90} \\
\mathbf{b}_{10j} &= \mathbf{g}_{100} \\
\mathbf{b}_{11j} &= \mathbf{g}_{110} \\
\mathbf{b}_{12j} &= \mathbf{g}_{120}
\end{aligned}
\tag{5}$$

Os efeitos fixos e aleatórios para o modelo de verificação da eficácia e equidade entre as escolas encontram-se na Tabela 3. Os dois resultados centrais desta tabela referem-se ao efeito do nível sócio-econômico sobre o desempenho do aluno. O valor médio desta variável em cada escola está forte e significativamente relacionado com o desempenho médio do aluno, ($\mathbf{g}_{01}=25.5$, $t= 64.20$). Mesmo controlando por este efeito de grupo, o nível sócio-econômico ainda influencia também o desempenho individual de cada aluno, dentro da escola. Ao nível do aluno,

¹⁶ O nível sócio-econômico médio da escola é igual a média aritmética do nível sócio-econômico dos alunos que estudam nas escolas.

seu próprio nível sócio-econômico contribui de forma estatisticamente significativa, mesmo que em menor grau, para o seu desempenho, $g_{60} = 1.2$, $t = 4.35^{17}$.

Tabela 3: Modelo de Verificação da Eficácia e Equidade entre as escolas

Efeitos Fixos		Coefficiente	Erro Padrão	Teste t	Valor-p
Média da escola, b_{0j}					
Intercepto,	g_{00}	260.7	0.47	552.91	0.00
Nsemédio,	g_{01}	25.5	0.40	64.20	0.00
Sexo Masculino,	g_{10}	3.5	0.28	12.35	0.00
Média das escolas em Ciências,	g_{20}	-2.4	0.43	-5.50	0.00
Média das escolas em Geografia,	g_{30}	3.5	0.43	8.10	0.00
Média das escolas em História,	g_{40}	3.6	0.43	8.20	0.00
Média das escolas em Português,	g_{50}	-13.9	0.43	-32.17	0.00
Inclinação desempenho-nse, b_{6j}					
Intercepto,	g_{60}	1.2	0.27	4.35	0.00
Nsemédio,	g_{61}	-1.9	0.33	-5.74	0.00
Aluno com 1 ano de repetência,	g_{70}	-16.0	0.35	-45.80	0.00
Aluno com 2 anos de repetência,	g_{80}	-20.0	0.45	-43.34	0.00
Aluno com 3 anos de repetência,	g_{90}	-21.0	0.67	-31.25	0.00
Aluno com 4 anos de repetência,	g_{100}	-22.8	1.01	-22.53	0.00
Aluno Pardo/Mulato,	g_{110}	-2.5	0.32	-8.01	0.00
Aluno Negro,	g_{120}	-7.5	0.58	-12.79	0.00
<hr/>					
Efeitos Aleatórios	Variância	Graus de Liberdade	c^2	Valor-p	
Média da escola, u_{0j}	143.5	2575	12614.0	0.00	
Inclinação do nse, u_{6j}	34.3	2575	2839.2	0.00	
Efeito do aluno, r_{ij}	1744.2				

Controlando pelo nível sócio-econômico dos alunos, a Tabela 3 revela ainda a existência de três outros fatores individuais com efeitos estatisticamente significativos sobre o desempenho estudantil: histórico escolar, gênero e raça. No que diz respeito ao histórico escolar, enquanto os

¹⁷ O nível sócio econômico do aluno varia de -2.84 até 9.19, enquanto que o nível sócio econômico médio da clientela varia de -2.11 até 2.98. Observa-se, ainda, que quando se inclui a etnia do aluno no modelo, há uma pequena queda no efeito do nível sócio econômico do aluno sobre o seu desempenho.

alunos com um ano de repetência apresentam, em média, desempenho inferior em dezesseis pontos, $g_{70} = -16.0$, os alunos com dois anos de repetência apresentam um desempenho médio ainda inferior, $g_{80} = -20.0$, assim como os alunos que repetiram três e quatro vezes, $g_{90} = -21.0$ e $g_{100} = -22.8$. Este forte impacto negativo da repetência sobre o desempenho dos alunos confirma a importância de tentarmos controlar para o estoque de capital humano com o qual o aluno chega ao teste, de forma a não atribuirmos erroneamente ao ensino recente, problemas advindos de trajetórias passadas.

Em termos de gênero, nota-se um desempenho superior dos alunos de sexo masculino em Matemática, Geografia e Ciências. As meninas tem desempenho superior em Português, e não há diferença estatística nos resultados para História. Estes resultados - que ainda indicam um desempenho melhor para meninos do que para meninas, em média - talvez deveriam servir de cautela para aqueles que, com base no avanço e predomínio feminino em termos de anos de escolaridade completos, imaginavam não haver mais necessidade de cuidados especiais para incentivar o aprendizado das meninas nas escolas fundamentais brasileiras. Permanece, aliás, um interessante problema: pareceria que os meninos precisam de mais incentivo do que as meninas para permanecer na escola, mas currículos e práticas pedagógicas ainda parecem favorecer os meninos sobre as meninas.¹⁸

Por último, a tabela indica que os alunos pardos ou mulatos têm desempenho inferior àquele observado para os alunos brancos¹⁹ em dois pontos e meio. Essa diferença de desempenho é ainda maior no caso dos alunos negros, $g_{120} = -7.5$. Este efeito negativo da cor sobre o rendimento escolar, mesmo após o controle pelo nível sócio-econômico, constitui um resultado altamente preocupante para aqueles interessados em reduzir a desigualdade de oportunidades no Brasil. O negro brasileiro parece não só ter menos chance de estar na escola mas, além disso, os que chegam a escola e aí logram permanecer, parecem ter um desempenho pior do que seus colegas brancos, mesmo controlando pelo nível sócio-econômico.

¹⁸ É possível que as diferenças de rendimento devam-se tão somente a um efeito de seleção: como os meninos saem da escola mais cedo, e imagina-se que os que saem sejam os menos habilidosos em termos escolares, é possível que a amostra masculina tenha uma habilidade média superior. Mas esta possibilidade é meramente teórica, e caberia verificá-la ou falsificá-la empiricamente.

¹⁹ Nesta categoria também se enquadram os alunos amarelos, que têm desempenho médio semelhante ao dos alunos brancos, bem como os alunos indígenas. Estes últimos representam menos de 3% da amostra.

O tema da equidade é abordado quando avaliamos a maneira pela qual o nível sócio-econômico do aluno afeta o seu desempenho. Repare que este conceito está intimamente ligado a eficácia da escola, b_{0j} , pois só faz sentido falar sobre equidade quando existe alguma variável que contribui, em média, para um melhor desempenho da escola. Ainda de acordo com a tabela 3, o incremento de uma unidade no nível sócio econômico médio da clientela da escola reduz o impacto do nível sócio econômico do aluno no seu desempenho em 1.9 pontos.

A tabela 3 também apresenta informações sobre os efeitos aleatórios. A variância residual entre as escolas, $\hat{t}_{00}=143.5$, caiu para 19,6% em relação ao modelo incondicional, onde $\hat{t}_{00}=733.9$. O que implica que 80.4% da variância observada entre as escolas no desempenho médio de seus alunos deriva da composição social interna e da seletividade da escola. Ainda assim, a variância que restou entre as escolas é estatisticamente significativa. Ou seja: há diferenças entre as escolas - tanto no que tange à eficácia quanto no que tange à equidade - que não são explicadas por diferenças relativas a características da clientela, tais como composição sócio-econômica, racial ou de gênero. A próxima seção utiliza dados sobre as escolas propriamente ditas, para investigar se parte desta variância não explicada deve-se a características da escola em si, ou de seus professores.

4. O papel das características e práticas escolares

Como vimos na seção anterior, cerca de 28% da variância total de resultados no teste do SAEB de 1999 era entre médias escolares, o que amplamente respaldava a visão de que as escolas brasileiras não são estatisticamente idênticas em termos de preparo para o teste. Vimos também que, desta porção da variância existente entre escolas, cerca de oitenta por cento é explicada por diferenças no nível sócio-econômico médio da clientela da escola. Ocorre, claramente, um fenômeno de estratificação ("*sorting*") dos alunos por critérios econômicos, que reforça o impacto positivo que esta variável tem a nível individual.²⁰

Não obstante a importância quantitativa deste efeito, verificou-se ainda que a variância restante após controlar-se por ele permanecia estatística e economicamente significativa. Nesta

²⁰ Um modelo econômico de estratificação por critério duplo - de mérito acadêmico e nível sócio-econômico - é o de Fernandez (1997).

seção, incluímos no modelo anterior as variáveis escolares disponíveis na base de dados, para testar o poder explicativo destas para o desempenho dos estudantes. As variáveis escolares estão classificadas em três grupos. O primeiro refere-se a itens de infra-estrutura da escola. O segundo analisa o impacto da disponibilidade de recursos financeiros na escola; enquanto o terceiro está relacionado às práticas escolares. Uma descrição mais detalhada de cada uma das variáveis encontra-se disponível no Apêndice 1.

No intuito de verificar o impacto de cada variável escolar na função de produção educacional, foi introduzida na modelagem do intercepto e da inclinação uma variável de cada vez. Embora muitas variáveis se mostrassem importantes na modelagem do intercepto, somente duas se mostraram (alternativamente) relevantes na modelagem da inclinação. São elas: o nível de escolaridade e o salário do professor. Os próximos modelos reportam como estas variáveis influenciam o desempenho do aluno.

A fim de prosseguir com a análise, observa-se que o modelo de nível 1 continua a ser representado pela equação (4), enquanto o modelo de nível 2 passa a ser representado pela equação (6)²¹:

$$\mathbf{b}_{0j} = \mathbf{g}_{00} + \mathbf{g}_{01} * (\textit{particular})_{.j} + \mathbf{g}_{02} * (\textit{saa})_{.j} + \mathbf{g}_{03} * (\textit{rp})_{.j} + \mathbf{g}_{04} * (\textit{irf})_{.j} + \mathbf{g}_{05} * (\textit{faltaalu})_{.j} + \mathbf{g}_{06} * (\textit{nescpr})_{.j} + \mathbf{g}_{07} * (\textit{nsemédio})_{.j} + \mathbf{g}_{08} * (\textit{prop_rep})_{.j} + u_{0j}$$

$$\begin{aligned} \mathbf{b}_{1j} &= \mathbf{g}_{10} \\ \mathbf{b}_{2j} &= \mathbf{g}_{20} \\ \mathbf{b}_{3j} &= \mathbf{g}_{30} \\ \mathbf{b}_{4j} &= \mathbf{g}_{40} \\ \mathbf{b}_{5j} &= \mathbf{g}_{50} \\ \mathbf{b}_{6j} &= \mathbf{g}_{60} + \mathbf{g}_{61} * (\textit{nescpr})_{.j} + \mathbf{g}_{62} * (\textit{nsemédio})_{.j} + \mathbf{g}_{63} * (\textit{prop_rep})_{.j} + u_{6j} \\ \mathbf{b}_{7j} &= \mathbf{g}_{70} \\ \mathbf{b}_{8j} &= \mathbf{g}_{80} \\ \mathbf{b}_{9j} &= \mathbf{g}_{90} \\ \mathbf{b}_{10j} &= \mathbf{g}_{100} \\ \mathbf{b}_{11j} &= \mathbf{g}_{110} \\ \mathbf{b}_{12j} &= \mathbf{g}_{120} \end{aligned}$$

Os resultados da estimação do modelo (4) - (6) encontram-se na Tabela 4. Na literatura das funções de produção da educação, a variável escolar que se mostrou um maior número de vezes significativa para explicar a variância do desempenho dos alunos foi a experiência do professor. No contexto educacional brasileiro, essa variável não parece ser relevante, pois em nenhum momento mostrou-se significativa: nem na modelagem do intercepto, nem na modelagem da inclinação. Em contrapartida, a variável "particular", indicando se a escola pertence à rede pública ou particular, mostrou-se altamente significativa na modelagem da eficácia. Controlando-se por todas as variáveis escolares e individuais, bem como pelo nível sócio-econômico médio dos alunos, a escola particular ainda agrega 6,1 pontos à média dos resultados de seus alunos.

A variável de rede ("particular") não foi significativa na modelagem do parâmetro de equidade ($\hat{\alpha}_6$) e foi, portanto, omitida da especificação apresentada.²²

A partir desta tabela, observa-se que o desempenho médio das escolas depende de uma série de fatores escolares, depois de controlado o efeito do nível sócio econômico médio de seus alunos, da rede de ensino a que ela pertence e da proporção de repetentes. Em comparação com modelos intermediários (ver nota 22), a introdução desses novos fatores cancelou parte dos efeitos antes atribuídos a rede de ensino e ao nível sócio econômico médio de seus alunos, cujos coeficientes neste momento passam a ser respectivamente: $g_{01}=6.1$, $t=6.00$ e $g_{07}=18.8$, $t=26.34$.

Um resultado que chama a atenção é o impacto grande e significativo da proporção de repetentes sobre o desempenho médio das escolas, $g_{08} = -12.5$, $t = -2.54$. Como este impacto é significativo mesmo após o controle do nível sócio econômico médio dos alunos da escola e da rede de ensino, isto significa que nem sempre os alunos mais pobres que estudam nas escolas públicas são os alunos repetentes.

²¹ Foram testadas as interações entre as variáveis e a grande maioria revelou-se não significativa. Deste modo, optou-se por não reportá-las.

²² De acordo com a metodologia "de baixo para cima" da modelagem hierárquico-linear, estimou-se anteriormente uma especificação que incluía todas as variáveis referentes ao aluno (da equação 4) e a variável de rede ("particular"), excluindo outras variáveis escolares. O efeito da rede sobre eficácia e equidade foram qualitativamente os mesmos que os apresentados na Tabela 4.

Tabela 4 Modelo educacional com nível de escolaridade do professor

Efeitos Fixos		Coefficiente	Erro Padrão	Teste t	Valor-p
Média da escola, b_{0j}					
Intercepto,	g_{00}	266.2	1.58	168.68	0.00
Particular,	g_{01}	6.1	1.09	6.00	0.00
Sala de aula arejada,	g_{02}	3.5	0.91	3.91	0.00
Ruído prejudica,	g_{03}	-1.2	0.69	-1.79	0.07
Insuficiência de rec. financeiros,	g_{04}	-5.1	1.05	-4.86	0.00
Falta aluno,	g_{05}	-4.3	1.19	-3.60	0.00
Nível escolaridade professores,	g_{06}	2.9	1.17	2.50	0.01
NSE Médio,	g_{07}	18.8	0.69	26.34	0.00
Proporção repetentes	g_{08}	-12.5	1.95	-6.43	0.00
Sexo masculino,	g_{10}	3.5	0.29	12.40	0.00
Média das escolas em Ciências*,	g_{20}	-2.6	0.44	-5.84	0.00
Média das escolas em Geografia*,	g_{30}	3.4	0.44	7.78	0.00
Média das escolas em História*,	g_{40}	3.5	0.44	7.97	0.00
Média das escolas em Português*,	g_{50}	-14.0	0.44	-31.95	0.00
Inclinação Desempenho -NSE, b_{6j}					
Intercepto,	g_{60}	2.8	0.75	3.75	0.00
Nível escolaridade professores,	g_{61}	2.1	1.06	2.03	0.04
NSE Médio,	g_{62}	-2.8	0.47	-6.09	0.00
Proporção repetentes,	g_{63}	-4.0	1.56	-2.54	0.01
Aluno com 1 ano de repetência,	g_{70}	-15.5	0.36	-43.48	0.00
Aluno com 2 anos de repetência,	g_{80}	-19.3	0.47	-41.33	0.00
Aluno com 3 anos de repetência,	g_{90}	-20.3	0.68	-29.66	0.00
Aluno com 4 anos de repetência,	g_{100}	-21.9	1.04	-21.06	0.00
Aluno Pardo/Mulato,	g_{110}	-2.4	0.32	-7.52	0.00
Aluno Negro,	g_{120}	-7.4	0.59	-12.42	0.00
<hr/>					
Efeitos Aleatórios	Variância	Graus de Liberdade	c^2	Valor-p	
Média da escola, u_{0j}	126.9	2475	11238.8	0.00	
Inclinação do NSE, u_{6j}	31.9	2480	2715.2	0.00	
Efeito do aluno, r_{ij}	1745.5				

O primeiro grupo de variáveis escolares, referente às condições de infraestrutura física, está representado pelas variáveis: sala de aula arejada e nível de ruído das salas de aula. Espera-se que quanto melhor a infra-estrutura escolar, melhor o ambiente de estudo para os alunos, o que vêm a favorecer o seu desempenho. Isso pode ser confirmado na tabela, pois quando a escola possui salas de aula arejadas, o desempenho médio dos alunos da escola aumenta, $g_{02} = 3.5$ pontos, e se o nível de ruído da escola prejudica os alunos e professores, o desempenho cai, em média, 1.2 pontos.

A variável que está relacionada ao segundo grupo mede a insuficiência de recursos financeiros. Quando a escola apresenta problemas de caráter financeiro, o desempenho do aluno cai, em média, 5.1 pontos. O terceiro e último grupo está relacionado às práticas escolares. A incidência de falta às aulas, por parte dos alunos, e o nível de escolaridade do professor²³ tratam dessas questões. A tabela 6 descreve o importante impacto do nível de escolaridade do professor sobre o desempenho médio dos alunos, $g_{06} = 2.9$.²⁴ Como era de se esperar, quanto maior a incidência de falta às aulas, por parte dos alunos, menor é o desempenho médio dos mesmos, $g_{05} = -4.3$. Na realidade, esta variável também acaba por captar, mesmo que parcialmente, o grau de interesse nas aulas, por parte dos alunos. Assim, nada mais razoável de se observar do que uma diminuição no desempenho médio dos alunos se os mesmos não se mostram interessados nas aulas.

A Tabela 4 retrata também qual o impacto do nível sócio-econômico médio dos alunos, do nível de escolaridade do professor e da proporção de repetentes na escola, sobre a maneira como o nível sócio econômico do aluno afeta a sua proficiência, b_{6j} . Em termos de equidade, observa-se que um maior nível sócio econômico médio dos alunos faz com que o nível sócio econômico familiar do aluno afete menos o seu desempenho²⁵, $g_{62} = -2.8$. Por sua vez, um maior

²³ Como está descrito no Apêndice 1, os valores tomados por esta variável oscilam entre 0 e 3, onde 3 se refere aos professores que possuem formação de ensino superior.

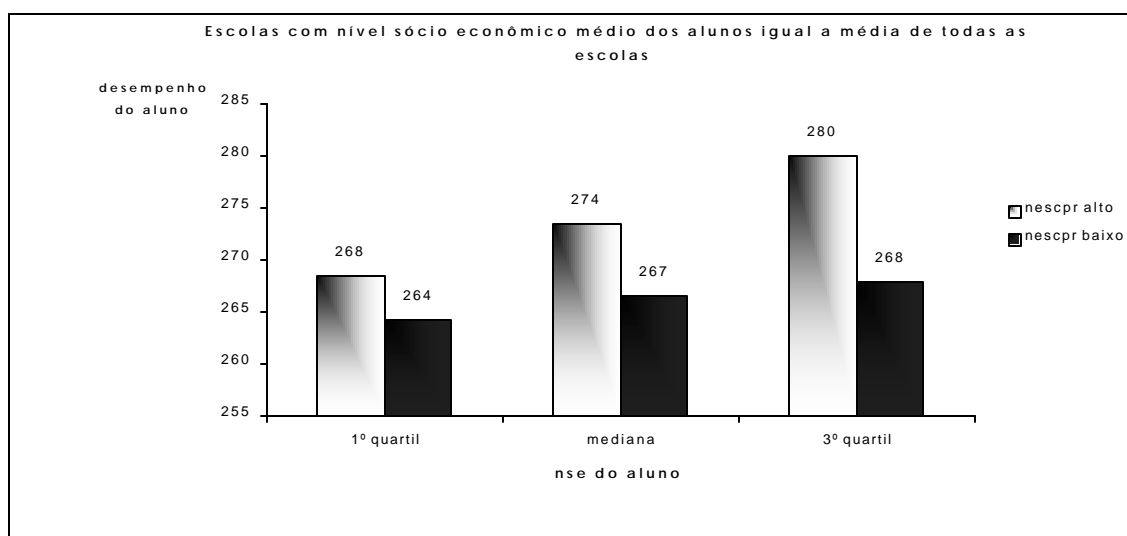
²⁴ Quando se inclui a variável de salário do professor na modelagem de b_{0j} e b_{6j} , o nível educacional do professor torna-se insignificante para explicar tanto o melhor desempenho do aluno, quanto a relação de seu nível sócio econômico com o seu desempenho. Substituiu-se, então, a variável salarial pelo nível educacional do professor. A partir deste novo modelo conclui-se que o salário do professor impacta o desempenho dos alunos de forma semelhante que o nível de escolaridade do professor o faz. Enquanto que um maior salário se traduz num maior desempenho médio da escola, à medida que o salário do professor aumenta, o grau de associação entre o nível sócio econômico do aluno e seu desempenho se torna mais forte.

²⁵ Ou seja, o resultado da Tabela 3 permanece válido neste modelo mais geral.

nível de escolaridade do professor tem o efeito oposto: quanto mais instruído o professor, maior o impacto do nível sócio-econômico do aluno no seu desempenho, $g_{61} = 2.1$.

Ou seja: por um lado, um maior nível de escolaridade do professor faz com que a escola seja mais eficaz - aumenta o desempenho médio dos alunos que freqüentam essa escola. Mas, por outro, esse maior nível de escolaridade torna a escola menos equitativa - aumenta a importância do nível sócio econômico do aluno para o seu desempenho. Uma outra maneira de ver o mesmo efeito encontra-se ilustrada na Figura 2: o aumento na escolaridade dos professores beneficia todos os alunos da escola²⁶. Mas este benefício é crescente com o NSE familiar do aluno, de forma que professores mais qualificados contribuem para o aprendizado de todos, mas de forma mais acentuada para o dos alunos de nível sócio-econômico mais elevado.²⁷

Figura 2: Efeito de um Aumento na Escolaridade do Professor



Um último exercício foi realizado, no sentido de testar a robustez dos parâmetros estimados à omissão da variável "rede" do modelo. Assim, re-estimamos o modelo (4) - (6), excluindo a variável rede, e reexaminando todas as outras variáveis disponíveis, a fim de verificar qualquer mudança de significância ou magnitude nos coeficientes. A única variável dantes não significativa a 5% que passou a sê-lo foi a não-disponibilidade de professores para lecionar certas

²⁶ A Figura 2 apresenta o desempenho de alunos em escolas com NSE médio entre -0,4 e 0,4, de forma a controlar para o efeito do NSE médio da escola.

²⁷ Na Figura 2, a distribuição de escolas por NSE médio é representada por três pontos: o primeiro quartil, a mediana e o terceiro quartil.

matérias (resposta dada pelo diretor da escola). Quando a variável "rede" é omitida, esta variável passa a afetar negativamente o desempenho do aluno, indicando que uma das razões por trás da menor eficácia da escola pública poderia ser um menor sucesso em garantir a disponibilidade efetiva dos professores necessários para a completa cobertura do currículo.

5. Conclusões

Através da aplicação de um método estatístico apropriado - os modelos hierárquicos lineares - a uma base de dados relativamente nova - os resultados dos testes padronizados do SAEB de 1999 - tentamos, neste artigo, contribuir para uma melhor compreensão dos determinantes da qualidade do ensino fundamental no Brasil. Os resultados encontrados agrupam-se, de forma bastante natural, em duas categorias: a primeira refere-se ao papel das características individuais e familiares dos próprios estudantes, e a segunda ao papel das variáveis escolares e dos professores.

Na primeira categoria, o resultado central é a enorme importância do nível sócio-econômico (NSE) da família como determinante do desempenho do aluno, tanto entre escolas como dentro delas. Quanto mais alto o NSE médio da escola, melhor o desempenho médio dos alunos da escola. E mesmo após controlar pelo NSE médio da escola, quanto maior o NSE individual, melhor o rendimento escolar do aluno dentro da escola.²⁸ Mas este efeito positivo do NSE individual não é igual em todas as escolas. Na verdade, quanto mais "rica" a escola (em termos de seu NSE médio), menor o efeito das condições familiares de cada aluno. Ou seja, a escola com alunos mais "ricos" tende a compensar melhor pelas desvantagens educacionais de seus alunos oriundos de famílias mais "pobres".

Uma vez controlado o efeito do nível sócio-econômico, duas outras características individuais permaneceram significativamente correlacionadas com o desempenho do aluno: gênero e raça. O melhor desempenho dos meninos na maioria das matérias soa um alerta contra a complacência relativa à questão do desempenho feminino na escola brasileira. Ainda que as mulheres estejam de fato permanecendo mais tempo na escola do que os homens (o que em si só é um problema de gênero, ainda que masculino), seu pior desempenho - especialmente em

²⁸ Este resultado também é encontrado em outros países, tanto na América Latina (ver Mizala e Romaguera, 1998, e Mizala et. al., 1999), como nos Estados Unidos (ver Hanushek, 1989).

matemática e ciência - merece atenção. Parece ainda haver escopo para reformas pedagógicas que visem melhorar o aprendizado das meninas em matérias como estas.

O pior desempenho do aluno negro (preto ou pardo), mesmo após controlarmos pelo NSE, é igualmente preocupante. Supondo que erros de medida na variável NSE sejam ortogonais à variável raça, a significância estatística deste resultado depõe contra a existência de igualdade de oportunidades na escola brasileira. Trabalhos anteriores (como, por exemplo, Henriques, 2001) já apontavam que os negros têm menores taxas de acesso e frequência à escola. O presente resultado sugere que, mesmo dentro da escola e controlando pelo seu nível sócio-econômico, o aluno negro tem um pior desempenho de aprendizado. Num contexto onde políticas afirmativas começam a ser debatidas no país, cremos ser necessária uma investigação cuidadosa e detalhada dos aspectos raciais da prática educacional nas escolas brasileiras. Existe alguma forma de discriminação - ainda que sutil e talvez inconsciente - na alocação do aluno negro à sala de aula? Ou por parte do professor em sua atenção individual? Existe algum problema de relevância cultural do currículo ou do material didático que possam torná-los menos interessante ou apropriado para alunos de minorias raciais? Estas são questões que vão muito além do escopo deste trabalho, mas que os presentes resultados sugerem como temas relevantes para pesquisa futura.

Passando à categoria dos resultados sobre as variáveis escolares, cabe ressaltar que a significância estatística dos efeitos de um bom número delas sobre o desempenho do aluno contrasta com os resultados usuais na literatura econométrica americana e latino-americana. Isto pode dever-se a uma diferença legítima entre o Brasil e os outros países estudados, ou a diferenças metodológicas, principalmente devido a uma redução no viés de agregação que pode ter levado à subestimação dos coeficientes destas variáveis em análises anteriores. Seja como for, tanto a qualidade do professor, medida principalmente por seu nível de escolaridade, como a qualidade da estrutura física da escola (recursos financeiros, salas arejadas e silenciosas), contribuem para um melhor desempenho de seus alunos.²⁹

²⁹ Mesmo controlando por todos estes fatores explicitamente, assim como para a composição étnica e sócio-econômica da clientela, a escola pública teve um desempenho estatisticamente significativamente inferior ao da escola particular no Brasil. Caberia investigar se há aspectos de gestão e/ou incentivos internos aos funcionários da rede pública que podem ser responsáveis por este pior desempenho. Este resultado também se coaduna com modelos da economia política do financiamento à educação no Brasil (ver Ferreira, 2001).

Cabe notar que este resultado não confirma, apenas, que escolas melhores levam a melhores resultados. Isso não chegaria a surpreender. O que ele mostra é que essa variância de qualidades escolares pode ser efetivamente observada na amostra das escolas brasileiras em operação em 1999. Ou seja, há escolas no Brasil onde as crianças estão aprendendo menos do que poderiam, por causa da insuficiência de recursos financeiros; da insuficiência de professores; de sua baixa escolaridade; de salas barulhentas e/ou abafadas. Não é preciso um grande esforço intelectual para discernir implicações deste fato para a política pública, com conseqüências tanto para a eficácia quanto para a igualdade de oportunidades em nosso sistema educacional.

Outro resultado com possíveis implicações para a política educacional é o fato de que incrementos na escolaridade do professor contribuem para um melhor desempenho de todos os alunos, mas de forma crescente com o nível sócio-econômico destes. Claramente, o aumento da escolaridade do professor, por aumentar a eficácia das escolas é, por si só, desejável. Caso almejemos, entretanto, reduzir a desigualdade de oportunidades educacionais entre famílias mais e menos afortunadas, conviria tentarmos desenvolver políticas educacionais e pedagógicas voltadas para identificar e retificar as razões específicas por trás do pior desempenho dos alunos mais pobres. Tais políticas poderiam incluir uma maior disponibilidade do professor fora do horário de aula, para quem deseje procurá-lo; programas de incentivo ao uso da biblioteca; a disponibilidade de acompanhamento psicológico, ou uma série de outras idéias.³⁰ A investigação do mérito de cada uma foge ao escopo do presente estudo, no qual apenas constatamos a necessidade de tais políticas para evitar que aumentos na eficácia escolar sejam acompanhados por aumentos na desigualdade de oportunidades dentro da escola.

³⁰ É óbvio que cada uma destas políticas teria custos. O que se aponta aqui é que elas também teriam benefícios, com os quais os custos devem ser comparados.

Bibliografia:

Arrow, K.J. (1962): "The Economic Implications of Learning by Doing", *Review of Economic Studies*, **29**, pp.155-173.

Barbosa, M. e Fernandes, C. (2001): "A Escola Brasileira Faz Diferença?: Uma Investigação dos Efeitos da Escola na Proficiência em Matemática dos Alunos da Quarta Série.", em C. Franco (ed.) (2001): *Promoção, Ciclos e Avaliação na Educação* (Porto Alegre: ArtMed).

Barros, R., Mendonça, R. e Santos, D., (1999): "Determinantes do Desempenho Educacional no Brasil", mimeo, IPEA.

Barros, R., Henriques, R. e Mendonça, R., (2000): "Pelo fim das décadas perdidas: educação e desenvolvimento sustentado no Brasil", *Desigualdade e Pobreza no Brasil*, IPEA, cap.14, pp.405-423.

Becker, G., Murphy, K., Tamura, R., (1990): "Human capital, fertility, and economic growth", *Journal of Political Economy*, **98**: pp.s12-s37.

Bidwell, E. e Kasarda, D., (1980): "Conceptualizing and Measuring the Effects of School and Schooling", *Journal of Education*: pp.401-430.

Bryk, S. e Raudembush, W., (1986): "A Hierarchical Model for Studying School Effects", *Sociology of Education*, **59**: pp.1-17.

Bryk, S. e Raudembush, W., (1992): *Hierarchical Linear Models: Applications and data analysis method*, Newbury Park, Sage.

Coleman, S., (1966): *Equality of Educational Opportunity*, Office of Education, U.S., Washington D.C.

Fernandez, R., (1997): "To each according to?: Markets, Tournaments, and the Matching Problem with Borrowing Constraints", mimeo, New York University.

Ferreira, F.H.G. (2001): "Education for the Masses?: The Interaction between Wealth, Educational and Political Inequalities", *Economics of Transition*, **9** (2), pp.533-552.

Franco, C., Mandarino, M. e Ortigão, M.I., (2001): "Projeto Pedagógico de Escola Promove Qualidade e Equidade em Educação?", *Revista UNDIME-RJ*, **VII** (2), pp.30-46.

Fletcher, R., (1997): "À procura do Ensino Eficaz". Relatório de Pesquisa, PNUD/MEC/SAEB.

Goldstein, H., (1995): *Multilevel Statistical Models*, Kendall's Library of Statistics.

Hanushek, E., (1986): "The Economics of Schooling: Production and Efficiency in Public Schools", *Journal of Economic Literature*, **24**, pp.1141-1177.

- Hanushek, E., (1989): "The Impact of Differential Expenditures on School Performance", *Educational Researcher*, pp.45-52.
- Henriques, R. (2001): "Desigualdade Racial no Brasil: evolução das condições de vida na década de 90", TD#807, IPEA, Rio de Janeiro.
- Johnson, R., (1998): *Applied Multivariate Statistical Analysis*, Prentice-Hall, Inc.
- Langoni, C. (1973): *Distribuição de Renda e Desenvolvimento Econômico do Brasil*, Rio de Janeiro: Expresso e Cultura
- Lucas, R., (1988): "On the mechanics of economic development", *Journal of Monetary Economics*, **22**: pp. 3-42.
- Mincer, J. (1958): "Investment in Human Capital and Personal Income Distribution", *Journal of Political Economy*, **66**: pp.281-302.
- Mizala, A. e Romaguera, P., (1998): "Desempeño Escolar y Eleccion de Colegios: La experiencia Chilena", Documento de Trabajo #36, Serie Economía, Universidad de Chile.
- Mizala, A., Romaguera, P. e Reinaga, T., (1999): "Determinants of Student Achievement and School Performance in Bolivia", CD- Rom LACEA 2000.
- Relatório do SAEB 1999, Diretoria de Avaliação da Educação Básica, INEP, Ministério da Educação ([http: www.inep.gov.br](http://www.inep.gov.br))
- Summers, A. e Wolfe, B. (1977): "Do schools Make a Difference?", *The American Economic Review*, **67** (4), pp. 639-652.
- Tinbergen, J. (1975): *Income Differences: Recent research*. Oxford: North Holland Publishing.
- Uzawa, H. (1965): "Optimal technical Change in an Aggregative Model of Economic Growth", *International Economic Review*, **6**: pp.18-31.

Apêndice 1 - Descrição das variáveis escolares

A tabela A1 mostra as características das variáveis escolares incluídas nas funções de produção educacional apresentadas ao longo do capítulo 3.

Tabela A1 Estatística Descritiva para as Variáveis Escolares

Variável	Média	Desvio Padrão	Valor Máximo	Valor Mínimo
$(particular)_j$	0.25	0.43	1.0	0.0
$(nescpr)_j$	2.83	0.29	3.0	1.5
$(salário)_j$	4.14	1.59	9.0	0.0
$(irf)_j$	0.71	0.34	1.0	0.0
$(saa)_j$	0.88	0.32	1.0	0.0
$(rp)_j$	0.30	0.46	1.0	0.0
$(faltaalu)_j$	0.32	0.33	1.0	0.0
$(nprof)_j$	0.34	0.35	1.0	0.0

As variáveis $(particular)_j$, $(irf)_j$ e $(rp)_j$ estão presentes no questionário da escola. Estas questões foram respondidas pelo diretor das escolas. As demais questões foram trazidas do questionário do professor para o questionário da escola³¹. A partir da tabela acima, observa-se que existem duas variáveis que oscilam entre valores maiores que uma unidade. São elas: $(nescpr)_j$ e $(salário)_j$

Os valores tomados pela variável $(nescpr)_j$, expressam que os professores apresentam os seguintes níveis de escolaridade média:

- 0.0- ensino fundamental incompleto
- 1.0- ensino fundamental completo
- 2.0- ensino médio
- 3.0- ensino superior

A variável $(salário)_j$, por sua vez, refere-se ao rendimento mensal médio dos professores. Ela pode assumir valores de 0 a 9, de acordo com a faixa salarial recebida pelos professores:

- 0.0- até R\$136,00
- 1.0- de R\$137,00 a R\$272,00
- 2.0- de R\$273,00 a R\$408,00

³¹ O resultado dessas variáveis na base da escola expressa a média aritmética das respostas de todos os professores que trabalham nesta escola.

- 3.0- de R\$409,00 a R\$544,00
- 4.0- de R\$545,00 a R\$816,00
- 5.0- de R\$817,00 a R\$1088,00
- 6.0- de R\$1089,00 a R\$1360,00
- 7.0- de R\$1361,00 a R\$2040,00
- 8.0- de R\$2041,00 a R\$2720,00
- 9.0- mais de R\$2720,00

Quanto as demais variáveis, cada escola recebe o valor de 1 em cada variável se: (1) pertence a rede particular; (2) apresenta insuficientes recursos financeiros; (3) possui salas de aula arejadas; (4) ruído nas salas de aula prejudica; (5) os alunos faltam as aulas; e, (6) não existem professores para lecionar algumas matérias.

Apêndice 2: Definindo o nível sócio-econômico do aluno

Com base na literatura sobre as funções de produção educacional, observa-se que o nível sócio-econômico afeta de forma expressiva o desempenho escolar dos alunos. No entanto, como o nível sócio-econômico representa um construto, ou conceito, ele não pode ser apreendido num único item de questionário, mas sim num conjunto de itens. Neste contexto se insere a análise fatorial, realizada neste estudo a fim de construir uma variável que represente o nível sócio-econômico dos alunos.

A análise fatorial é uma técnica estatística de resumo de informações muito utilizada na análise multivariada de dados. Seu principal objetivo é viabilizar a análise de associações existentes entre um grande número de variáveis, através da geração de um número reduzido de construtos, chamados “fatores”. Pode-se entender esses fatores como uma dimensão latente que se manifesta de forma redundante em algumas variáveis originais, isto é, as variáveis de uma determinada base de dados são agrupadas em função da correlação existente entre elas.

A técnica é baseada em um modelo explícito em que as variáveis da matriz de dados são expressas como uma função linear de um número reduzido de fatores latentes. No caso do nível sócio-econômico do aluno, os itens de questionários referentes a este construto, deverão ser descritos, através do modelo (Johnson, p.515):

$$X_{(px1)} - \mathbf{m}_{(px1)} = L_{(pxm)} F_{(mx1)} + \mathbf{e}_{(px1)}$$

Onde: X é um item de questionário, \mathbf{m} é o vetor da média da variável, L é a matriz de carga dos fatores (ou seja, a correlação entre as variáveis e o fator), F é o vetor de fatores comuns ou latentes e \mathbf{e} é um vetor de fatores específicos para cada variável. Considera-se que os fatores latentes formam um conjunto de variáveis padronizadas e descorrelacionadas, isto é:

$$E(F) = 0_{(mx1)}, \quad Cov(F) = E[FF'] = I_{(mxm)}$$

$$E(\mathbf{e}) = \mathbf{0}_{(px1)} \quad \text{Cov}(\mathbf{e}) = \mathbf{Y}_{(p \times p)} = \begin{bmatrix} \mathbf{y}_1 & 0 & \cdot & 0 \\ 0 & \mathbf{y}_2 & \cdot & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdot & \mathbf{y}_p \end{bmatrix}$$

F e \mathbf{e} são independentes, logo: $\text{Cov}(\mathbf{e}, F) = \mathbf{0}_{(pxm)}$. A covariância do modelo é dada por: $\Sigma = \text{Cov}(X) = LL' + \mathbf{Y}$, onde \mathbf{Y} é a matriz diagonal ($p \times p$) dos fatores individuais. Desta forma as variâncias de cada variável podem ser decompostas em comunalidade (h^2), devido aos fatores comuns e variância específica \mathbf{y} , devida aos fatores individuais.

$$s_{ii} = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \mathbf{y}_i$$

$\text{Var}(X_i) =$ comunalidades + variância específica

$$h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2$$

$$\text{Var}(X_i) = h_i^2 + \mathbf{y}_i$$

Como o modelo trabalha com variáveis padronizadas, a fim de evitar que uma possível variável apresente uma variância muito grande, influencie excessivamente na determinação das cargas dos fatores, a aproximação $\Sigma = \text{Cov}(X) = LL' + \mathbf{Y}$, pode ser aplicada igualmente na matriz de correlação, uma vez que para variáveis padronizadas a matriz de covariância corresponde a matriz de correlação das variáveis³².

Neste estudo, nove itens foram considerados relevantes na investigação do nível sócio-econômico do aluno. As questões encontram-se no questionário nº 1 do aluno. São elas:

Onde você mora existe:

- 1) água encanada?
- 2) eletricidade?

Quantos dos seguintes itens há no lugar onde você mora: (respostas: 0 a 6)

- 3) rádios?
- 4) televisões a cores?
- 5) geladeira?
- 6) freezer?
- 7) máquina de lavar roupa?
- 8) telefone residencial?
- 9) Até que série seu pai estudou? Até que série sua mãe estudou?

³² A tabela A.3, que se encontra no final deste apêndice, apresenta a matriz de correlação do modelo de fatores para as variáveis consideradas relevantes, na investigação do nível socioeconômico do aluno.

As respostas dos itens de nº 1 e 2 são sim ou não, codificadas para 0 se não e 1 se sim. As respostas das perguntas de nº 3 a 8 revelam a disponibilidade destes itens na casa dos alunos. As respostas podem variar entre 0 e 6. O último item considera duas perguntas presentes no questionário do aluno³³. A resposta deste item corresponde a maior série cursada pelo seu pai e/ou sua mãe.

No presente estudo, a extração de fatores foi realizada a partir da Análise dos Componentes Principais. O objetivo desta análise é determinar uma transformação linear das variáveis originais, em geral correlacionadas, que resulta em novas variáveis chamadas “componentes principais”. Estas são descorrelacionadas entre si e apresentam a mesma variação total das variáveis originais. Este é o método mais indicado quando a principal preocupação é achar um número mínimo de fatores necessários para substituir as variáveis originais.

Como há poucos alunos por escola na amostra e, mais ainda, o trabalho objetivou modelar o nível sócio-econômico do aluno, somente uma variável representante do nível sócio-econômico teve de ser extraída. Desse modo, foi extraída a “componente principal”, ou seja, a componente de maior variância extraída a partir dos dados. O quadro abaixo apresenta o total de variância original das variáveis explicada por essa primeira componente.

Tabela A2: Resultado da análise fatorial

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.036	33.734	33.734	3.036	33.734	33.734
2	1.218	13.529	47.263			
3	.903	10.036	57.299			
4	.779	8.660	65.960			
5	.736	8.181	74.141			
6	.670	7.442	81.583			
7	.662	7.351	88.935			
8	.548	6.092	95.026			
9	.448	4.974	100.000			

Extraction Method: Principal Component Analysis.

A análise fatorial apresentada foi realizada a partir do software *SPSS*. Selecionou-se a opção de se obter somente um fator nesta análise. A partir da tabela A.1, observa-se que a “componente principal” explica 33.734% da variância total das 9 variáveis. Esta análise também nos fornece a tabela A2, que mostra a carga das 9 variáveis neste fator. Estes valores correspondem às correlações de cada variável com a “componente principal”.

Na interpretação dos fatores, grande importância deve ser dada as cargas fatoriais. O valor da carga ao quadrado representa a quantidade de variância total da variável expressa pelo fator. Assim, uma carga de 0,30 traduz aproximadamente 10% de explicação e uma carga de 0,50 significa que 25% da variância original da variável é explicada pelo fator.

³³ Cada pergunta apresenta 6 possibilidades de respostas. Estas variam de 0: nunca estudou; até 4: frequentou a faculdade; e, 5: o aluno não respondeu.

Tabela A3: Matriz de fatores

Component Matrix ^a

	Component
	1
Água encanada ?'	.209
Eletricidade ?	.167
No. de rádios	.642
No. de TV a cores	.783
No. de geladeira	.575
No. de freezer	.612
No. de máquina de lavar roupa	.582
No. de telefone	.727
MAXEDU	.606

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

De forma que $nse_{ij} = .209$ (água encanada) + $.167$ (eletricidade) + $.642$ (nº rádios) + $.783$ (nº TV a cores) + $.575$ (nº geladeiras) + $.612$ (nº freezer) + $.582$ (máq. lavar roupa) + $.727$ (nº telefone) + $.606$ (maxedu).

Esse apêndice objetivou apresentar a obtenção da variável de nível sócio-econômico utilizada nas funções de produção educacional realizadas neste trabalho.

Tabela A4: Matriz de correlação:

Correlation Matrix

	Água encanada ?'	Eletricidade ?	No. de rádios	No. de TV a cores	No. de geladeira	No. de freezer	No. de máquina de lavar roupa	No. de telefone	MAXEDU	
Correlation	Água encanada ?'	1.000	.244	.068	.107	.081	.032	.099	.092	.109
	Eletricidade ?	.244	1.000	.044	.096	.111	.020	.063	.041	.072
	No. de rádios	.068	.044	1.000	.480	.272	.272	.254	.363	.288
	No. de TV a cores	.107	.096	.480	1.000	.367	.366	.327	.502	.425
	No. de geladeira	.081	.111	.272	.367	1.000	.281	.312	.281	.189
	No. de freezer	.032	.020	.272	.366	.281	1.000	.314	.380	.253
	No. de máquina de lavar roupa	.099	.063	.254	.327	.312	.314	1.000	.324	.212
	No. de telefone	.092	.041	.363	.502	.281	.380	.324	1.000	.406
	MAXEDU	.109	.072	.288	.425	.189	.253	.212	.406	1.000

a. Determinant = .208