

# 29

ISSN 1414-0640

Série Documental  
TEXTOS PARA DISCUSSÃO



# O Impacto da Infra-Estrutura Escolar na Taxa de Distorção Idade-Série das Escolas Brasileiras de Ensino Fundamental – 1998 a 2005

Sergei Soares  
Natália Sátyro

Ministério da Educação

**INEP**

Instituto Nacional de Estudos  
e Pesquisas Educacionais  
Anísio Teixeira

**MEC**

Ministério da Educação

**INEP**

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

**O Impacto da Infra-Estrutura Escolar na Taxa de Distorção  
Idade-Série das Escolas Brasileiras de Ensino Fundamental  
– 1998 a 2005**

Sergei Soares\*  
Natália Sátyro\*\*

\* Da Diretoria de Estudos Sociais do Instituto de Pesquisa Econômica Aplicada (Ipea).

\*\* Do Departamento do Cadastro Único, Secretaria Nacional de Renda de Cidadania (Senarc)  
Ministério do Desenvolvimento Social e Combate à Fome(MDS).

© Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep)  
É permitida a reprodução total ou parcial desta publicação, desde que citada a fonte.

ASSESSORA DE EDITORAÇÃO E PUBLICAÇÕES

Lia Scholze

PROGRAMAÇÃO VISUAL

Márcia Terezinha dos Reis

EDITOR EXECUTIVO

Jair Santana Moraes

REVISÃO

Antonio Bezerra Filho  
Rosa dos Anjos Oliveira

NORMALIZAÇÃO BIBLIOGRÁFICA

Regina Helena Azevedo de Mello

DIAGRAMAÇÃO E ARTE-FINAL

Raphael Caron Freitas

TIRAGEM

1.000 exemplares

EDITORIA

Inep/MEC – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira  
Esplanada dos Ministérios, Bloco L, Anexo II, 4º Andar, Sala 414  
CEP 70047-900 – Brasília-DF – Brasil  
Fones: (61) 2104-8438, (61) 2104-8042  
Fax: (61) 2104-9812  
editoria@inep.gov.br

DISTRIBUIÇÃO

Inep/MEC – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira  
Esplanada dos Ministérios, Bloco L, Anexo II, 4º Andar, Sala 404  
CEP 70047-900 – Brasília-DF – Brasil  
Fone: (61) 2104-9509  
publicacoes@inep.gov.br  
<http://www.publicacoes.inep.gov.br>

ESTA PUBLICAÇÃO NÃO PODE SER VENDIDA. DISTRIBUIÇÃO GRATUITA.  
A exatidão das informações e os conceitos e opiniões emitidos são de exclusiva  
responsabilidade do autor.

Dados Internacionais de Catalogação na Publicação (CIP)  
Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

---

Soares, Sergei.

O impacto da infra-estrutura escolar na taxa de distorção idade-série das escolas brasileiras de ensino fundamental – 1998 a 2005 / Sergei Soares ; Natália Sátyro. – Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2008.

22 p. : il. – (Série Documental. Textos para Discussão, ISSN 1414-0640 ; 29)

1. Distorção idade-série. 2. Infra-estrutura escolar. 3. Ensino de 1º grau. I. Sátyro, Natália. II. Título. III. Série.

CDU 371.212.72:37.046.12

---

## Sumário

### O Impacto da Infra-Estrutura Escolar na Taxa de Distorção Idade-Série das Escolas Brasileiras de Ensino Fundamental – 1998 a 2005

RESUMO .....	5
INTRODUÇÃO .....	5
1 REFERENCIAL TEÓRICO .....	6
2 MODELO DE REGRESSÃO LINEAR DE PAINEL COM EFEITO FIXO .....	6
2.1 Dados .....	6
2.2 Medidas .....	8
2.2.1 Variáveis resposta: taxa de distorção idade/série escolar .....	8
2.2.2 Variáveis explicativas .....	9
2.3 Modelo .....	10
3 RESULTADOS (NÃO PARAMÉTRICOS) .....	11
4 CONCLUSÕES .....	16
REFERÊNCIAS BIBLIOGRÁFICAS .....	19
ANEXO .....	21



# O Impacto da Infra-Estrutura Escolar na Taxa de Distorção Idade-Série das Escolas Brasileiras de Ensino Fundamental – 1998 a 2005

## RESUMO

O efeito dos insumos escolares sobre o desempenho educacional tem sido freqüentemente estudado na literatura, mas não há consenso sobre seu impacto. Este trabalho apresenta resultados que apontam para a existência de um impacto significativo destes insumos, nas escolas brasileiras de ensino fundamental, sobre a taxa de distorção idade/série escolar entre 1998 e 2005. A análise empírica é feita por meio de mais de um método de estimação, mas os principais resultados são advindos do modelo de regressão linear para dados de painel com efeito fixo das escolas. O estudo é baseado em dados do Censo Escolar de 1998 a 2005 sobre uma coorte de 132.603 escolas, e, para os municípios, foram utilizados dados do Censo Demográfico de 2000. O principal resultado é de que insumos escolares importam sim para o desempenho educacional, principalmente no caso de escolas com maior precariedade das condições infra-estruturais.

Palavras-chave: insumos escolares, função de produção educacional, infra-estrutura escolar.

## INTRODUÇÃO

Em discussão anterior, Sátyro e Soares mostraram que a relação entre a evolução das condições materiais do ensino fundamental e os resultados educacionais não é clara. Por um lado, há um grupo de escolas, as rurais, que sofrem com condições extremamente precárias e, também, com resultados educacionais inaceitáveis; por outro lado, em várias dimensões esta relação não se verifica. Em primeiro lugar, houve uma melhoria considerável nas condições materiais médias das escolas brasileiras durante a última década sem que os resultados educacionais mostrassem melhoria significativa no período

(Sátyro; Soares, 2007). O desempenho escolar, medido pelo Sistema Nacional de Avaliação da Educação Básica (Saeb), manteve-se constante ou em queda em todo o País, e as taxas de repetência caíram até o fim da década de noventa, mantendo-se constantes depois. Em segundo lugar, há pouca diferença entre as condições materiais das escolas das redes estadual e privada, mas há grande diferença entre as duas redes no que tange aos resultados educacionais.

É claro que comparações entre resultados agregados são muito limitadas, uma vez que pode haver causalidade tanto entre resultados educacionais e as condições materiais como o reverso. Em outras palavras, é necessária uma abordagem mais cuidadosa para desvendar se há ou não causalidade entre resultados educacionais e as condições materiais. O objetivo deste trabalho é a construção de um modelo explicativo da taxa de distorção idade/série das escolas brasileiras que oferecem ensino fundamental. Mais especificamente analisamos aqui o impacto da infra-estrutura escolar – compreendida *lato sensu*, incluindo formação docente, existência de biblioteca ou sala de leitura, material pedagógico, bem como características municipais – na taxa de distorção idade/série das escolas. Os resultados mostram que, no Brasil, a infra-estrutura tem um impacto forte no desempenho escolar. Pode-se assim inferir que, caso não tivesse havido o investimento que houve na última década em infra-estrutura nas escolas, os resultados educacionais seriam ainda piores.

O texto organiza-se da seguinte forma: na seção 1 apresenta-se brevemente um resumo da literatura, com a exposição das hipóteses testadas; na seção 2 apresenta-se a especificação formal do modelo e a descrição breve do Censo Escolar e do Censo Demográfico; depois, na seção 3, são apresentados os resultados; e, por fim, as conclusões.

## 1 REFERENCIAL TEÓRICO

Dois fatos contribuíram fortemente para o interesse acadêmico acerca da relação insumos escolares e escola de qualidade. O primeiro é que, na última metade do século 20, observou-se, nos EUA e em outros países desenvolvidos, um aumento tanto do valor absoluto como do percentual do PIB gasto com educação e uma concomitante queda nos escores das avaliações educacionais (Hanushek, 2003). Este fato, entre outros, levou ao famoso Relatório Coleman de 1966, que concluiu, a partir de um estudo com cerca de 600 mil alunos, que a escola tinha apenas efeitos marginais sobre os resultados escolares – o grande fator causal seria a família e o capital social trazido dela e do meio em que ela vive. A partir desse relatório, muitos trabalhos foram desenvolvidos nessa área; no entanto, a discussão da significância e mesmo da direção do impacto dos insumos escolares ainda é controversa: não há consenso sobre a relação direta entre o aumento de insumos e o aumento dos resultados escolares.

As análises de custos e benefícios educacionais e de eficiência na alocação de recursos com o intuito de melhorar os resultados escolares têm sido cada vez mais exploradas. Um dos mais ardorosos defensores da não correlação entre insumos educativos e desempenho escolar é Eric A. Hanushek, que em 1986 e em 1997 apresentou metanálises da produção acadêmica controversa em seus resultados. Nesses estudos, Hanushek conclui, semelhantemente a Coleman, que são a origem socioeconômica e o capital social dos alunos os fatores que, de fato, importam para o desempenho escolar, em detrimento do papel da escola e da sua infra-estrutura. Nos estudos compilados podemos notar o uso de diversos fatores enquanto indicadores de insumos escolares: número de alunos por turma, gastos por aluno, escolaridade dos professores, experiência dos professores, salários dos professores, qualidade da infra-estrutura da escola, resultados de professores em testes de avaliação.

Talvez a exposição dos resultados de Hanushek seja um tanto caricatural, uma vez que o objetivo principal dele era mais enfatizar a importância de fatores de processo e não afirmar

peremptoriamente que insumos escolares jamais possam trazer melhorias na qualidade do ensino – tampouco há, na literatura internacional, unanimidade em torno desta tese. Lant Pritchett e Deon Filmer, em artigo de 1999, defendem que investimento em equipamentos tem retornos positivos e significativos; Card e Krueger defendem o mesmo em artigo de 1996. O relevante é que a literatura internacional aponta majoritariamente, embora não consensualmente, para a ineficácia de políticas de insumo para melhorar a qualidade educacional.

Entretanto, os resultados desses estudos geralmente consideram somente os países desenvolvidos e reportam apenas marginalmente ou em notas de rodapé que suas considerações não são generalizáveis. Ou seja, para os países em desenvolvimento, que apresentam realidades muito distintas em relação aos equipamentos escolares, onde encontramos uma precariedade maior no que se refere ao sistema educacional público, em que as escolas são muito mais heterogêneas no que diz respeito à infra-estrutura, temos que reconsiderar a hipótese muito peremptória de que o aumento nos insumos não produz resultados diretos no desempenho escolar. Uma das razões para isso é que os estudos feitos nos países em desenvolvimento quase sempre usam um único levantamento de dados transversal, o que leva a problemas grandes de variáveis endógenas e causalidade reversa. O argumento principal deste trabalho é que em casos onde as condições infra-estruturais são muito precárias sua melhoria produz impacto positivo nos resultados escolares. Para tanto usaremos sempre um painel de escolas que permitirá um controle muito melhor, embora certamente não total ou perfeito, da endogeneidade presente nas estimações.

## 2 MODELO DE REGRESSÃO LINEAR DE PAINEL COM EFEITO FIXO

### 2.1 Dados

Neste estudo, insumos escolares serão entendidos como infra-estrutura de todo tipo: número médio de alunos por turma, número de horas-aula, docentes com formação superior, construção e melhoria das dependências da

escola, existência de biblioteca ou sala de leitura – ou seja, tudo aquilo que o dinheiro pode comprar.

O Censo Escolar (CE) contém dados de todas as escolas, de todos os níveis e modalidades de educação básica no Brasil. Ou seja, o CE inclui informações de todo o universo escolar da educação básica brasileira: escolas de todas as redes administrativas de ensino e de todas as localizações: rurais, urbanas e metropolitanas. O CE inclui também informações sobre o grau de instrução dos docentes por escola, e este é, de acordo com a literatura especializada em avaliação educacional, um dos melhores preditores do desempenho escolar. O formulário é enviado para a escola, anualmente, no mês de março. Em 2005 tínhamos um universo de 207.234 estabelecimentos de educação básica, com a seguinte distribuição: 32.296 de escolas oferecendo educação infantil-creche, 105.616 para pré-escola, 162.727 para o ensino fundamental, 23.561 oferecendo ensino médio, 7.053 para educação especial, 44.838 estabelecimentos com educação de jovens e adultos e 3.230

para educação profissional. O recorte deste trabalho se deu nas escolas que oferecem ensino fundamental.

O número total de escolas que oferecem ensino fundamental passou de 187.493, em 1998, para 162.727. Destas, trabalhamos apenas com o universo das que permaneceram em todos os anos analisados, tendo assim uma série temporal de 132.603 escolas: 55,49% rurais, 33,33% urbanas e 11,18% metropolitanas.

Também por meio do Censo Demográfico de 2000 é possível analisar as características gerais da população, das pessoas que são as responsáveis pelos domicílios e, também, dos domicílios e de seus moradores, das condições habitacionais, educacionais e de renda.

Não serão considerados, a não ser como variáveis de controle, fatores extra-escolares como seleção de alunos e *status* socioeconômico, seja de alunos, seja da comunidade.

**Tabela 1 – Número de Estabelecimentos Escolares de Ensino Fundamental segundo localização, de acordo com o Censo Escolar – 1998-2005**

Ano	Número Total de Escolas	Número de Escolas em Áreas Rurais	Número de Escolas em Áreas Urbanas (não metropolitanas)	Número de Escolas em Áreas Metropolitanas
1998	187.493	119.163 (64%)	49.771 (27%)	18.559 (10%)
1999	183.448	113.236 (62%)	51.159 (28%)	19.053 (10%)
2000	181.504	110.333 (61%)	51.965 (29%)	19.206 (11%)
2001	177.780	105.823 (60%)	52.449 (30%)	19.508 (11%)
2002	172.508	99.806 (58%)	53.102 (31%)	19.600 (11%)
2003	169.075	95.573 (57%)	53.657 (32%)	19.845 (12%)
2004	166.484	92.739 (56%)	54.020 (32%)	19.725 (12%)
2005	162.727	88.989 (55%)	54.183 (33%)	19.555 (12%)

Fonte: Sátyro; Soares, 2007, p.8.

## 2.2 Medidas

### 2.2.1 Variável resposta: taxa de distorção idade/série escolar

Num sistema educacional seriado, como a maior parte do sistema educacional brasileiro, há, teoricamente, uma idade adequada para cada série escolar. No caso brasileiro, a idade ideal para iniciar o ensino fundamental é aos sete anos, o que implica uma duração de oito ou nove anos, dependendo da duração do ensino fundamental oferecido por uma determinada escola. A taxa de distorção idade/série é o cálculo de quantas crianças estão acima da idade ideal em uma determinada série. O Inep divulga a seguinte fórmula para fins de cálculo com base no Censo Escolar:

[...] considerando o Censo Escolar do ano  $t$  e a série  $k$  do ensino fundamental, cuja idade adequada é de  $i$  anos, então o indicador será expresso pelo quociente entre o número de alunos que, no ano  $t$ , completam  $i + 2$  anos ou mais (nascimento antes de  $t - [i + 1]$ ) e a matrícula total na série  $k$ . A justificativa deste critério é que os alunos que nasceram em  $t - [i + 1]$  completam  $i + 1$  anos no ano  $t$  e, portanto, em algum momento deste ano (de 1º de janeiro a 31 de dezembro) ainda permaneciam com  $i$  anos, e, por isso, o critério aqui adotado considera estes alunos como tendo idade adequada para esta série. Os que nasceram depois de  $t - [i + 1]$  completam, no ano  $t$ ,  $i$  anos ou menos.<sup>1</sup>

Ou seja, o conceito de distorção idade/série usado é a porcentagem de alunos defasados dois anos ou mais. É importante ressaltar que quanto maior a distorção pior é o desempenho escolar. Deste modo, coeficientes negativos nas estimações a seguir denotam um efeito positivo sobre a qualidade do ensino.

Há uma íntima relação entre a taxa de distorção idade/série e a taxa de repetência,

sendo que esta é causa daquela. Embora não seja linear, essa relação é monótona crescente: quanto maior for a repetência nos anos anteriores ao ano em questão, maior será a defasagem. Por que não optamos por usar a taxa de repetência, já que é esta a variável que nos interessa? A razão é que a repetência, sendo um fluxo, é sujeita a um ruído de medida no Censo Escolar muito maior que a defasagem, que é um estoque. Como o Censo ainda não acompanha alunos de um ano para o próximo, a repetência não é medida diretamente, devendo ser estimada, o que gera um ruído demasiadamente grande.

Há um argumento contrário ao uso da defasagem (ou repetência) como medida de qualidade. Segundo ele, uma escola de boa qualidade seria uma escola exigente com seus alunos e com critérios elevados para passar de ano. Isto naturalmente levaria a uma relação positiva entre a repetência e a qualidade do ensino. Este argumento é, lamentavelmente, a base de uma crença amplamente difundida entre pais, professores e o público leigo no nosso País. Embora este efeito exista, em nossa opinião ele deve ser pequeno, uma vez que a quase totalidade da literatura, tanto qualitativa como quantitativa, aponta a repetência alta como uma característica marcante de um ensino de baixa qualidade.<sup>2</sup> Convidamos os leitores que vêm a afirmação acima com ceticismo a consultarem a *Pedagogia da repetência*, de Sérgio Costa Ribeiro, entre outros numerosos estudos sobre o tema.

Há, no entanto, que explicitar os limites do uso da defasagem idade/série como indicador de qualidade. O primeiro, cuja relevância quantitativa se limita à área rural, é que algumas crianças ingressam no ensino fundamental após os sete anos; conseqüentemente, a relação entre defasagem e repetência torna-se menos nítida. Se usarmos aquela como *proxy* para esta nas áreas rurais, haverá um ruído indesejável exercendo influência nos resultados.

<sup>1</sup> *Glossário do Edudatabrasil* – Sistema de Estatísticas Educacionais: <<http://www.edudatabrasil.inep.gov.br/>>; acessado em 30 de outubro de 2006.

<sup>2</sup> É possível que este efeito adquira importância em uma análise dinâmica. Ou seja, enquanto não há dúvida de que a alta repetência seja uma característica de um ensino de baixa qualidade, é possível que melhorias na qualidade do ensino, a curto prazo, não sejam acompanhadas por reduções na repetência. Isto ocorreria, por exemplo, se uma melhoria organizacional na escola levasse simultaneamente a maiores expectativas e melhor ensino.

Outra limitação é que, da 5ª série em diante, é possível que haja crianças que cursaram da 1ª à 4ª em outra escola. Ou seja, é possível que a defasagem seja produto de repetência, sim, mas em outra escola – novamente temos uma fonte de ruído indesejável. Um modo de testar isso seria repetir a análise apenas para as escolas de 1ª a 4ª série.

Para medir infra-estrutura das escolas, utilizamos os seguintes indicadores:<sup>3</sup> proporção de docentes com grau de formação de nível superior completo, média de alunos por turma, média de horas-aula diária, fator infra-estrutura linear e quadrático, existência de biblioteca, localização das escolas – rurais e não-rurais (urbanas e metropolitanas) – e proporção da população educada dividida em três categorias: adultos analfabetos,

adultos com até 8ª série completa e adultos com ensino médio ou mais anos de escolaridade. Para controlar as características dos municípios nos quais as escolas estão inseridas, utilizamos o logaritmo da mediana da renda domiciliar *per capita*, o logaritmo do desvio padrão da renda domiciliar *per capita*, proporção de não brancos, proporção de população em áreas urbanas, logaritmo da população. Por fim, como variável de controle de autocorrelação e de legado utilizados, a *lag* da taxa de distorção idade/série escolar – trata-se da variável dependente no seu tempo *t-1*.

### 2.2.2 Variáveis explicativas

As variáveis explicativas usadas encontram-se na Tabela 2.

**Tabela 2 – Variáveis Independentes**

Variável	Tipo
<b>Todos os Modelos</b>	
Porcentagem de docentes com ensino superior	Porcentagem entre 0 e 100
Alunos por turma	Razão variando entre 1 e ?
Horas diárias de instrução	Número de horas
Índice de infra-estrutura escolar	Índice entre -1,5 e 4
Existência de biblioteca	Variável binária
<b>Apenas Modelos Mínimos Quadrados</b>	
Localização (rural, urbana, metropolitana)	Variável triária
Mediana do log da renda do município	Variável contínua positiva
Desvio padrão do log da renda do município	Variável contínua positiva
Proporção não brancos na população municipal	Porcentagem entre 0 e 100
Taxa de urbanização do município	Porcentagem entre 0 e 100
Logaritmo da população do município	Variável contínua positiva
Taxa de distorção no ano anterior	Porcentagem entre 0 e 100
Variáveis indicatrizes de ano e Estado	Variáveis binárias

<sup>3</sup> Mais detalhes no Anexo.

Usamos quatro critérios na definição das variáveis. O primeiro é que estamos interessados primordialmente em insumos escolares. A pergunta que motiva este texto é: Em que medida os insumos escolares contribuem para a qualidade escolar? Fatores extra-escolares, como *status* socioeconômico, ou intra-escolares, mas não relacionados com insumos, são de interesse apenas como controle.

O segundo, é claro, é que estamos limitados à informação disponível. Por exemplo, gostaríamos de contar com um painel de informações sobre crianças e suas famílias, mas isto simplesmente não existe no Brasil. Ou seja, mesmo reconhecendo a importância da origem socioeconômica dos estudantes para o seu desempenho escolar, não há essa informação para o formato analítico aqui proposto. Igualmente informação sobre a gestão da escola, o processo de seleção de alunos e o processo de formação de turmas não se encontram nem no Censo Escolar nem no Censo Demográfico. Ou seja, só podemos usar aquilo a que temos acesso.

O terceiro critério é a literatura nacional e internacional sobre qualidade escolar. Usamos apenas variáveis que aparecem nesta literatura. Por exemplo, no Censo Escolar há informações sobre o combustível do forno da escola, mas não há na literatura nenhuma menção a fornos à lenha ou a gás. Portanto, não usamos esta informação.

Finalmente, usamos nossa própria discricção na definição de diversas variáveis. Por exemplo, para modelar a relação entre a renda do município e os resultados escolares, optamos por usar o logaritmo da mediana da renda, quando o log da renda média, a média dos log da renda, ou até a renda média ou a mediana (sem log) teriam sido defensáveis. Para tanto, além de muita discussão, rodamos vários modelos para ver qual o melhor ajuste. Não cabe aqui discutir em maior detalhe o processo de escolha das variáveis usadas em suas formas funcionais. A única exceção é a forma funcional das principais medidas de insumos: o índice de infra-estrutura, horas-aula, alunos por turma e formação docente.

## 2.3 Modelo

Meio século de avanço na econometria deixou uma abundância de abordagens para modelar a relação causal entre duas ou mais variáveis em um contexto não-experimental. No entanto, não há, até hoje, nenhuma técnica que garanta estimar a verdadeira relação causal entre duas variáveis. Ou seja, os problemas de causalidade reversa, variáveis omitidas e todos os outros problemas que cabem debaixo do guarda-chuva “endogeneidade” não têm solução técnica e continuam a depender das hipóteses identificadoras.

Em educação, os problemas de endogeneidade são especialmente graves. Pais e alunos reagem a mudanças nas condições de ensino; secretarias de educação também reagem a resultados escolares, procurando outorgar mais recursos às escolas com piores resultados ou, então, de modo inverso, premiando as melhores; o processo de formação de turmas depende crucialmente do poder político na escola de certos grupos de pais, que podem ser justamente os que têm maior poder na própria secretaria de educação; enfim, existe um sem-número de relações sociais não men-suráveis que podem potencialmente viesar os parâmetros ligando insumos a resultados.

Neste trabalho, tentaremos três modelos. O primeiro é, sem dúvida, o mais fraco: estimaremos uma relação linear da defasagem entre idade e série mediante um modelo de mínimos quadrados ordinários para três anos da nossa amostra. Não duvidamos da forte endogeneidade relacionando o nível de insumos e resultados escolares. O objetivo deste exercício não é produzir estimativas críveis, mas sim uma linha de base que sirva de comparação com os outros modelos.

O segundo é o modelo empilhado: a estimação da mesma relação linear usando todos os anos para os quais temos dados. Para controlar variações temporais, usamos uma variável indicatriz para cada ano. Embora o modelo empilhado permita o uso de uma importante fonte de variação – a variação no tempo –, mas como não identifica as escolas, esta variação fica perdida junto com a variação no espaço. Acreditamos que ele também sofre de problemas

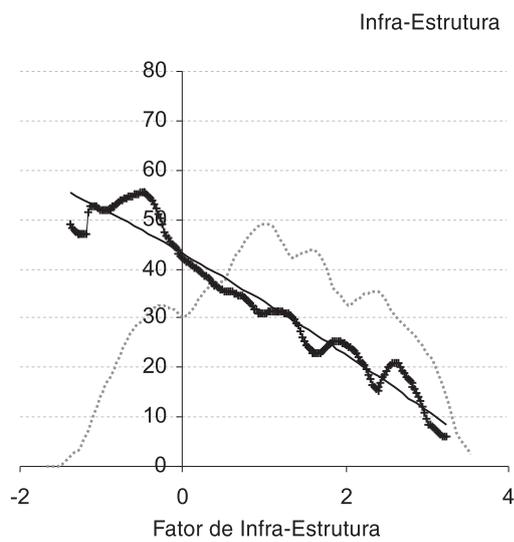
de endogeneidade, embora, *a priori*, acreditemos mais em seus resultados que nos modelos estimados para cada ano.

Por fim, estimaremos um modelo de efeitos fixos que consiste em identificar todas as observações relativas a uma escola e compará-las com a média desta mesma escola. Ou seja, estamos usando apenas a variação no tempo para identificar os efeitos de insumos escolares. *A priori*, é este o modelo em que mais acreditamos.

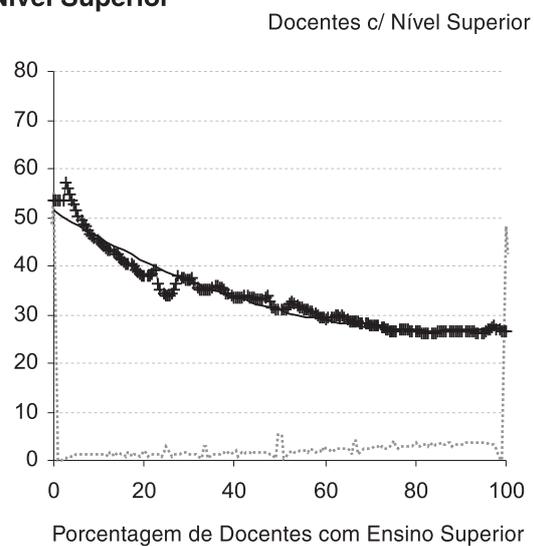
Finalmente, gostaríamos de permitir que os insumos exerçam influência sobre os resultados escolares não apenas de modo linear. Para tanto, estimaremos relações não paramétricas para as seguintes variáveis: o índice de infra-estrutura, horas-aula, alunos por turma e formação docente. Os gráficos destas relações nos ajudarão na definição da ordem do polinômio que usaremos para cada variável.

### 3 RESULTADOS (NÃO PARAMÉTRICOS)

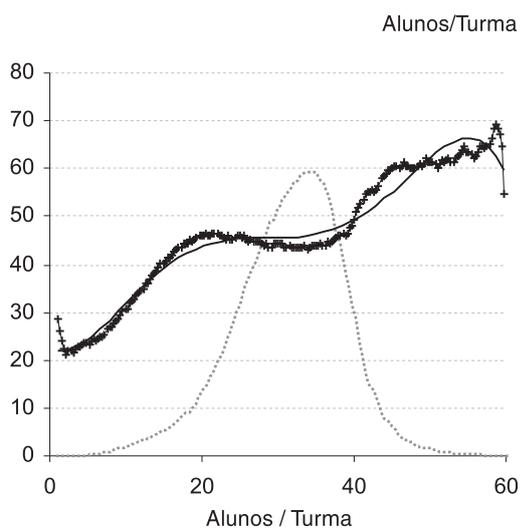
**Painel 1 – Infra-estrutura**



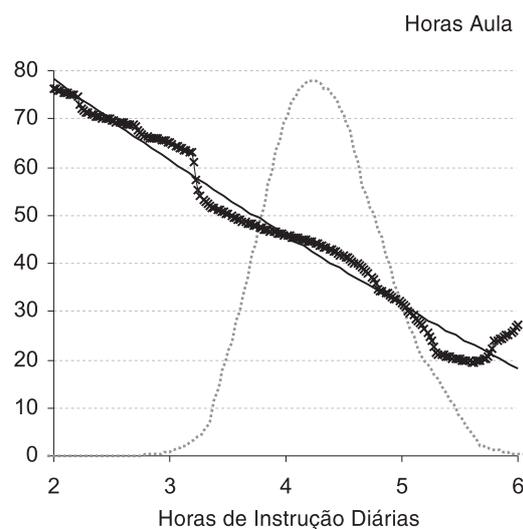
**Painel 2 – Porcentagem de Docentes com Nível Superior**



**Painel 3 – Alunos / Turma**



**Painel 4 – Horas-Aula**



**Gráfico 1 – Densidades e Relações com Taxa de Distorção Idade/Série**

Fonte: Censo Escolar

Os quatro painéis do Gráfico 1 mostram duas estimações não paramétricas cada um. As linhas em cinza pontilhado denotam a densidade da distribuição de cada variável. Por exemplo, a porcentagem de docentes com formação superior tem uma distribuição fortemente bi-modal – há um grupo grande de escolas onde nenhum docente tem formação superior e um outro, um pouco menor, onde todos os docentes a têm; e as escolas no meio, onde alguns contam com este nível de formação e outros não, se encontram distribuídas mais ou menos homoganeamente. Já a distribuição de horas-aula é aproximadamente normal, com quase nenhuma escola oferecendo menos que três ou mais que seis horas de instrução por dia, com uma média em 4,2 horas/dia.

As linhas sólidas em preto com tendência denotam a relação univariada, não paramétrica, entre a taxa de distorção idade/série e a variável em questão. Não há, nesta estimacão, a presença de nenhuma outra variável nem qualquer tipo de controle para endogeneidade. O objetivo é fazer uma análise exploratória para indicar o grau do polinômio a ser usado nas outras estimacões. Os resultados são claros: as relações entre defasagem e infra-estrutura e entre defasagem e horas-aula são aproximadamente lineares. Ou seja, um termo quadrático é suficiente para modelá-las. A relação entre defasagem e porcentagem de docentes com ensino superior, também monótona decrescente, é bem aproximada por uma relação quadrática.

Ao contrário, a relação entre defasagem e alunos por turma consiste de duas áreas de crescimento com um platô no meio, e é necessário um polinômio de ordem quatro para fazer uma boa aproximação.

A Tabela 3 mostra os resultados dos diferentes modelos. Vale lembrar que nossa variável-resposta é a taxa de distorção idade/série. Portanto, um coeficiente negativo denota um efeito positivo sobre a qualidade do ensino.

O primeiro resultado que salta aos olhos é que os modelos ano a ano e empilhado são

bastante semelhantes entre si, mas diferem bastante do modelo de efeitos fixos. Em particular, todos os efeitos escolares são mais fortes no modelo de efeitos fixos. Um exemplo é a presença da biblioteca, que reduz a defasagem em 1,5 ponto percentual no modelo de efeitos fixos e tem como único efeito significativo uma redução de 0,17 para o ano 1999.

Os controles sociodemográficos nos primeiros quatro modelos apresentam, *grosso modo*, os valores esperados: a presença de não-brancos, adultos pouco educados e um aumento na renda média reduzem a qualidade escolar, enquanto municípios com populações grandes, escolas particulares ou federais, alta desigualdade de renda (medida pelo desvio padrão da renda) a aumentam. O objetivo deste texto não é avaliar o impacto dos fatores extra-escolares, mas se muitos dos coeficientes fossem diferentes do usual, nossos resultados seriam postos em xeque. Neste caso, somente a renda média tem comportamento diferente do esperado.

Em suma, as estimacões dos modelos (1) a (4) mostram o padrão usual em estimacões transversais de funções de produção educacional: efeitos fracos e pouco significativos dos fatores escolares com efeitos mais fortes das variáveis usuais que pouco iluminam nossa busca por melhores políticas, como, por exemplo, que escolas da rede particular obtêm melhores resultados. A notável exceção é uma variável pouco comum na literatura internacional: horas-aula. Todas as estimacões mostram efeitos fortes e significativos desta variável, talvez por ser no Brasil o número de horas-aula excepcionalmente baixo.

O modelo de efeitos mostra resultados muito mais fortes para todos os fatores escolares e, é claro, nada pode dizer sobre fatores extra-escolares, já que a fonte para estes é o censo demográfico, que ocorre uma vez a cada dez anos. Dado que, fora o impacto da existência de uma biblioteca, todos os efeitos escolares foram modelados usando polinômios, a melhor maneira de mostrar os seus efeitos é mediante o uso de gráficos.

**Tabela 3 – Coeficientes e p-valores das estimações**

(continua)

Variável	(1) Empilhado		(2) 1999		(3) 2002		(4) 2004		(5) Modelo Efeitos fixos	
	B	P>t	B	P>t	B	P>t	B	P>t	B	P>t
Docentes nível superior (%)										
Linear	-0,01	0,00	0,01	0,00	0,01	0,00	-0,04	0,00	-0,14	0,00
Quadrático	0,00		0,00		0,00		0,00		0,00	
Média de alunos por turma										
Linear	0,10		-0,08		0,23		0,82		0,94	
Quadrático	0,00	0,00	0,00	0,00	-0,01	0,00	-0,04	0,00	-0,01	0,00
Cúbica	0,00		0,00		0,00		0,00		0,00	
Quártico	0,00		0,00		0,00		0,00		0,00	
Média de horas-aula										
Linear	-6,78	0,00	-6,78	0,00	-5,29	0,00	13,53	0,00	14,23	0,00
Quadrático	0,50		0,50		0,39		1,00		0,95	
Fator infra-estrutura										
Linear	0,40	0,00	0,33	0,00	0,56	0,00	0,71	0,00	-3,25	0,00
Quadrático	-0,14		-0,15		-0,22		-0,28		0,38	
Biblioteca	-0,02	0,47	-0,17	0,04	0,13	0,12	-0,07	0,47	-1,53	0,00
lag[taxa de distorção]	0,79	0,00	0,78	0,00	0,81	0,00	0,78	0,00		
Escola não-rural	-1,45	0,00	-1,59	0,00	-1,25	0,00	-0,96	0,00		
Rede										
Estadual	Base		base		base		base			
Federal	-3,59	0,00	-4,46	0,00	-0,89	0,27	-8,68	0,00		
Municipal	-0,83	0,00	-1,27	0,00	-0,95	0,00	-0,92	0,00		
Particular	-4,10	0,00	-5,34	0,00	-3,43	0,00	-4,71	0,00		

**Tabela 3 – Coeficientes e p-valores das estimações**

Variável	(conclusão)									
	(1) Empilhado		(2) 1999		(3) 2002		(4) 2004		(5) Modelo Efeitos fixos	
	B	P>t	B	P>t	B	P>t	B	P>t	B	P>t
Nível educacional dos adultos (em %)										
Analfabetos	Base		base		base		base			
Até oitava série	-0,09	0,00	-0,14	0,00	-0,12	0,00	-0,13	0,00		
Mais que oitava	-0,09	0,00	-0,13	0,00	-0,09	0,00	-0,02	0,07		
Renda domiciliar <i>per capita</i>										
Mediana	0,14	0,03	-0,29	0,08	0,13	0,41	0,04	0,85		
Desvio padrão	-0,14	0,00	-0,11	0,05	0,03	0,64	-0,29	0,00		
Proporção de não-brancos	0,02	0,00	0,00	0,02	0,01	0,00	0,00	0,87		
Proporção urbana	0,00	0,35	-0,07	0,01	0,00	0,04	0,00	0,37		
ln(população)	-0,06	0,00	3,06	0,00	-0,01	0,81	-0,03	0,36		
N	810.929		130.830		130.639		110.721		953.948	

Os quatro painéis do Gráfico 2 mostram o impacto dos fatores infra-estrutura, porcentagem de docentes com ensino superior, alunos/turma e horas-aula sobre a qualidade medida pela defasagem idade-série. É importante ressaltar que os gráficos são apenas relativos: não se pode afirmar qual é a defasagem média das escolas com quatro horas de aula por dia, mas apenas que estas são 5,6 pontos percentuais maiores que a média das escolas com cinco horas de aula por dia.

Em cada painel, o efeito estimado no modelo de efeitos fixos é mostrado usando uma linha preta com losangos, o efeito do modelo empilhado usando uma linha pontilhada preta, e os efeitos dos modelos de 1999, 2002 e 2004 usando linhas cinza claro, cinza escuro e preta, respectivamente.

Conforme já foi mencionado, o modelo de efeitos fixos estima maior impacto para todas as

variáveis. No caso de horas-aula, a forma das curvas é semelhante, mas os impactos são maiores. Um aumento de quatro para cinco horas de aula por dia reduziria a defasagem em 5,7 pontos segundo o modelo de efeitos fixos, contra apenas 2,2 segundo o modelo empilhado.

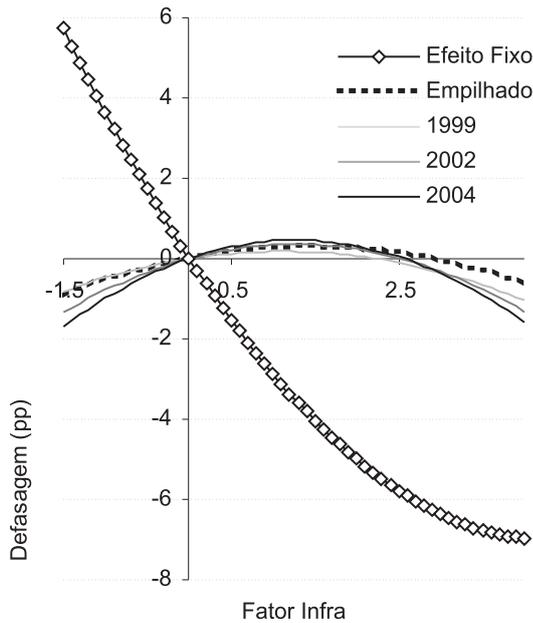
No caso das outras variáveis, a forma da curva difere bastante de modelo de efeitos fixos, por um lado, e das estimações ano a ano e o modelo empilhado, por outro. Por exemplo, no caso de infra-estrutura, modelo de efeitos fixos estima resultados fortes: uma melhoria de infra-estrutura de -1,5 para 0, que equivale a ir de uma escola péssima para uma escola ruim, reduz a defasagem idade-série em 5,7 pontos. Já os outros estimam que esta melhoria resultaria em um uma piora dos resultados educacionais.

Uma variável especialmente polêmica é tamanho de turma. Nosso modelo empilhado nos diz que reduzir o tamanho de turma de uma

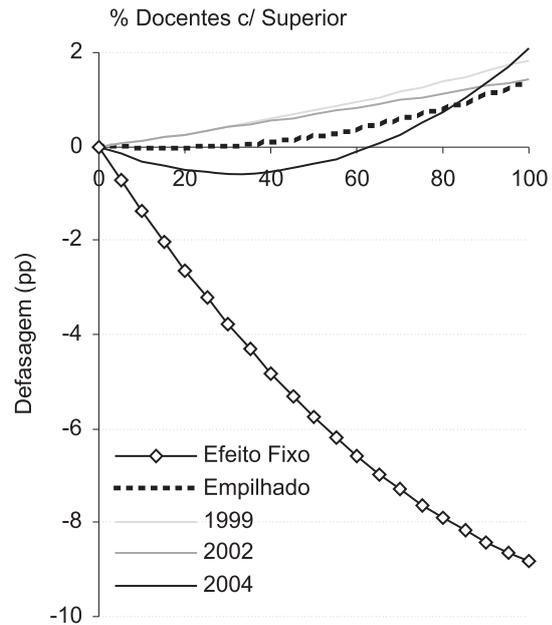
escola de 40 para 30 alunos reduziria a defasagem em pouco mais que um ponto percentual. O modelo de efeitos fixos discorda e afirma que seriam 2,5 pontos. A forma das curvas também é nitidamente diferente. Dada a ampla literatura indicando que tamanho de turma seria uma variável com relação fraca com resultados educacionais, encaramos este resultado com alguma

desconfiança. Além do mais, ao contrário das outras variáveis aqui investigadas, nas quais nossas escolas são muito maldotadas, as escolas brasileiras contam com turmas relativamente pequenas em comparação com outros países. Um dos resultados deste trabalho que merece ser melhor investigado é justamente o tamanho de turma.

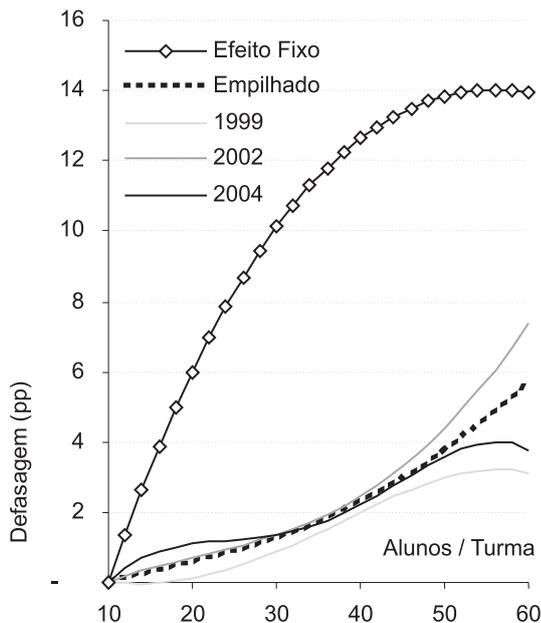
**Painel 1 – Infra-estrutura**



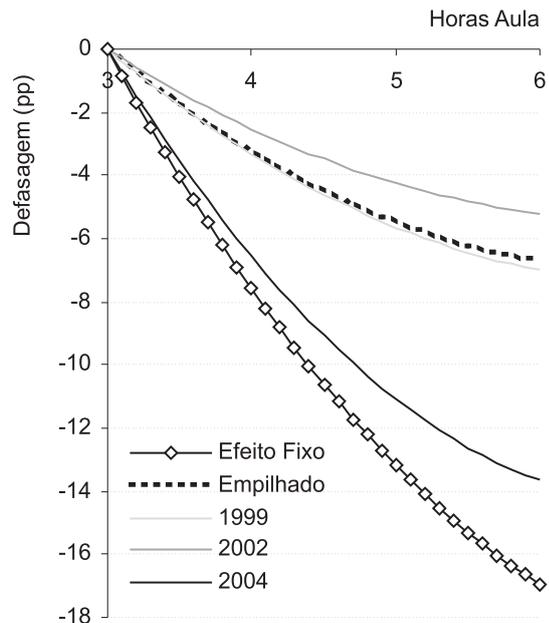
**Painel 2 – Porcentagem de Docentes com Nível Superior**



**Painel 3 – Alunos / Turma**



**Painel 4 – Horas-Aula**



**Gráfico 2 – Impactos dos Insumos Escolares sobre a Taxa de Distorção Idade/Série**

Fonte: Censo Escolar

Finalmente, o impacto da porcentagem de professores com nível superior, variável de considerável interesse para a política educacional no Brasil nos últimos anos, é um dos pontos nos quais há maior discordância entre os modelos. O modelo de efeitos fixos aponta um impacto grande e convexo, enquanto os outros apontam essencialmente a inexistência de qualquer relação ou uma relação negativa.

Em quem devemos acreditar? Nossa opinião é de que, com poucas dúvidas, devemos acreditar no modelo de efeitos fixos. Por quê? Por três razões.

A primeira é que, se queremos iluminar a política educacional, não estamos falando do impacto do *nível* de uma ou outra variável sobre a qualidade, e sim de *mudanças* em cada variável. Como já foi explicado acima, o modelo de efeitos fixos ignora o nível e estima os impactos de variações ao redor da média de cada escola. Ou seja, o modelo de efeitos fixos estima diretamente aquilo que queremos saber.

A segunda razão é que, em um campo minado pela endogeneidade, acreditamos haver menor espaço para relações causalidade reversa e variáveis omitidas no modelo de efeitos fixos. Em modelos transversais, causalidade reversa e variáveis omitidas tanto nos níveis como na variação levam a parâmetros endógenos. Já em modelos nos quais o nível não é importante, como o modelo de efeitos fixos, as relações de endogeneidade teriam que ocorrer na variação para viesar os parâmetros.

É claro que pode haver causalidade reversa e variáveis omitidas também na variação. É possível contar uma história na qual as autoridades educacionais reagem continuamente à deterioração na qualidade de uma escola com aumento de insumos, assim como é possível contar uma história em que pais educógenos com poder político fazem com que as melhorias nos insumos acompanhem seus filhos quando mudam de escola. No entanto, outras histórias seriam filtradas pelo modelo de efeitos fixos.

Uma última razão é que acreditamos que parte da melhoria nos insumos escolares dos

últimos anos foi conseqüência de políticas públicas universais, vistas através do Censo Escolar, que também deixam menos espaço para vies de endogeneidade. Por exemplo, a política de aumento das qualificações de professores via cursinhos de formação superior foi muito criticada por colocar os mesmos professores na mesma sala de aula com um papel a mais debaixo do braço, mas não se pode negar que foi resultado de decisões centrais e não de pressões diferenciadas no ambiente educacional. Na mesma linha, o Fundeb foi criado durante o período estudado, levando as melhorias educacionais exógenas a várias redes escolares no período.

É claro que é possível tratar endogeneidade usando técnicas mais apuradas, como variáveis instrumentais, mas este é assunto para outro trabalho.

Finalmente, cabe um comentário sobre a aplicação destes resultados na política educacional. O fato de uma medida ter impactos positivos sobre a qualidade educacional não quer dizer que ela seja boa. Quando se trata de políticas públicas, *sempre* há que se levar em consideração os custos de cada alternativa. Por exemplo, tanto o número de horas de instrução diárias quanto o número de alunos por sala de aula exercem impactos sobre a defasagem; no entanto, o impacto do primeiro é bem maior no intervalo onde se situa a maior parte das escolas. Aumentar o número de horas-aula de 3 para 4 reduziria a defasagem em 5,7 pontos percentuais a um custo de 33% na folha de pagamento; um aumento equivalente feito mediante redução de número de alunos por sala de aula de 40 para 26 proporcionaria uma redução de 3,9 pontos – significativa, mas bem menor que 5,7 pontos.

## CONCLUSÕES

A principal conclusão deste texto é clara: ao contrário do que diz a maior parte da literatura internacional sobre o tema, os insumos escolares são muito relevantes na definição dos resultados educacionais. Possivelmente por não contarem com dados em painel ou quando disponíveis tratem de mudanças pequenas a valores já elevados, os estudos, na sua maior parte,

não encontram os mesmos resultados que nós. Ao contrário, a inutilidade de dotar as escolas com melhores insumos é tema recorrente na literatura internacional e brasileira.

Outra conclusão igualmente relevante é que, com exceção da variável alunos por turma, o efeito é maior para as escolas que estão em pior situação. Ou seja, como os impactos de melhorias nos insumos escolares são maiores justamente para os que menos insumos têm, uma mesma intervenção aumenta a média e reduz a desigualdade na distribuição dos resultados escolares. É mais importante ir de três para quatro horas de aula por dia que de quatro para cinco; tem maior impacto ir de -1,5 para zero que de zero para 1,5 no índice de infra-estrutura; é melhor aumentar a proporção de professores com ensino superior de 0% para 10% que de 90% para 100%.

É claro que todos os condicionantes e limitantes sociais da educação, tão dolorosamente expressivos na maior parte da literatura até hoje, continuam valendo mesmo à luz dos resultados deste texto. A origem socioeconômica dos alunos, os *peer effects*, os efeitos de uma gestão deficiente, todos continuam sendo válidos, e nada indica que insumos escolares melhores possam a eles sobrepujar. O que indicamos, sim, é que os efeitos dos insumos não são nulos e que políticas para sua melhoria não estão, *a priori*, fadadas ao fracasso.

Quais são as principais limitações do trabalho? Uma limitação se refere ao uso exclusivo da defasagem idade/série como indicador de qualidade; em breve, quando tivermos acesso aos microdados da Prova Brasil, será possível usá-los para medir aprendizado diretamente. Outra é que não se levou em conta a interação dos efeitos de infra-estrutura com variáveis de gestão, ou político-pedagógicas, como autonomia escolar, expectativas, participação dos pais, e assim por diante. É possível que estas sejam potencializadas ou potencializem as variáveis de insumos de que tratamos aqui. Há metodologias em dois estágios para tanto, que poderão ser exploradas em estudos futuros.

As direções para futuros trabalhos também são claras. Devemos usar o banco de dados do Censo Escolar, que começamos a explorar neste trabalho, para melhor investigar as relações acima. Métodos que se sugerem são o uso de estimadores que melhor usam a estrutura de painel para lidar com endogeneidade, como arellano-bond, e o uso de mudanças na política educacional como o Fundeb e os requisitos mínimos de formação docente como instrumentos ou experimentos naturais.

Urge também investigar melhor os resultados relativos ao tamanho de turma, uma vez que não há especificidade brasileira que justifique resultados tão discrepantes com relação ao encontrado na literatura internacional.



## REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, D. F.; SINGER, J. M. Análise de dados longitudinais. In: SIMPÓSIO NACIONAL DE PROBABILIDADES E ESTATÍSTICA, 4. 1996. [Anais...]. 1996.

BARBOSA, M. E. Ferrão; FERNANDES, C. Modelo multinível: uma aplicação a dados de avaliação educacional. *Estudos em Avaliação Educacional*, n. 22, p. 135-153, 2000.

BECK, N.; KATZ, J. What to do (and not to do) with time-series-cross-section data in comparative politics. *American Political Science Review*, n. 89, p. 634-647, 1995.

CARD, David; KRUEGER, Alan B. *Labor market effects of school quality: theory and evidence*. 1996. (NBER Working Paper n. 5450)

CÉSAR, C. C.; SOARES, J. F. Desigualdade acadêmica induzida pelo contexto escolar. *Revista Brasileira de Estudos de População*, v. 18, n. 1/2, jan./dez. 2001.

FERRÃO, M. E.; BELTRÃO, K. I.; SANTOS, D. O impacto da política de não-repetência uma proficiência dos alunos da quarta série? Um estudo sobre o Sudeste Brasileiro. *Revista Brasileira de Estatística*, 2002a.

\_\_\_\_\_. Modelo de regressão multinível: aplicação ao estudo do impacto da política de não-repetência no desempenho escolar dos alunos da 4ª série. *Pesquisa e Planejamento Econômico*, v. 32, n. 3, 2002b.

FLETCHER, P. *À procura do ensino eficaz*. Relatório de pesquisa, Pnud/MEC/Saeb, 1997.

FRANCO, F. C. J.; LEE, V. E.; SÁTYRO, N. G. D. *National educational policies and their consequences for quality and equity*. Documento apresentado no encontro anual da American Sociological Association, Philadelphia, 1997.

GOLDSTEIN, H. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, n. 73, p. 43-56, 1986a.

\_\_\_\_\_. *Multilevel statistical models*. Wiley, 1986b.

GOLDSTEIN, H.; RASBASH, J. Improved approximations for multilevel models with binary response. *Journal of the Royal Statistical Society*, v. A, n. 159, p. 505-513, 1996.

HANUSHEK, E. A. School Resources. In: HANUSHEK Eric A.; WELCH, Finis (Ed.). *Handbook of the Economics of Education*. Amsterdam: Elsevier, 2006.

\_\_\_\_\_. The Failure of Input-based Schooling Policies. *Economic Journal*, 113, p. F64-F98, Feb. 2003.

\_\_\_\_\_. Spending on Schools. In: MOE, Terry (Ed.). *A Primer on American Education*. Hoover Press, 2001. p. 69-88.

KLEIN, Rubem; RIBEIRO, Sérgio Costa. A pedagogia da repetência ao longo das décadas. *Ensaio*, Rio de Janeiro, v. 3, n. 6, p. 55-62, jan./mar. 1995.

KREFT, I.; LEEUW, Jan. *Introducing multilevel modeling*. Sage, 1998.

LEE, Valerie 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, n. 3, p. 5 e 11, 1989.

PLEWIS, Ian. Comment on "Centering" predictors in multilevel analysis: choices and consequences. *Multilevel Modelling Newsletter*, v. 1, n. 3, p. 6 e 11, 1989.

PRITCHETT, L.; FILMER, D. What education production function *really* show: a positive theory of education expenditures. *Economics of Education Review*, v. 18, n. 2, p. 223-239, 1999.

RAUDENBUSH, S.; BRYK, A. S. *Hierarchical linear models*. 2. ed. Newbury Park: Sage, 2002.

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

RIBEIRO, Sérgio Costa. A pedagogia da repetência. *Estudos em Avaliação Educacional*, São Paulo, n. 4, p. 73-86, jul./dez. 1991. Versão eletrônica disponível em: <<http://www.scielo.br>>.

SÁTYRO, N. G. D.; SOARES, D. Sergei. *A infra-estrutura das escolas brasileiras de ensino fundamental: um estudo com base nos Censos Escolares de 1997 a 2005*. Brasília: Ipea, 2007. (Texto para discussão, 1267).

SILVA, P. L. N. et al. Análise Estatística de dados da Pnad: incorporando a estrutura do plano amostral. *Ciência e Saúde Coletiva*, v. 7 n. 4, p. 659-670, 2002.

SOARES, J. F.; Ribeiro, L. M.; CASTRO, C. de M. Valor agregado de Instituições de Ensino Superior em Minas Gerais para os cursos de Direito, Administração e Engenharia Civil. *Dados, Revista de Ciências Sociais*, Rio de Janeiro, v. 44, n. 2, p. 363-396, 2001.

SOARES, J. F.; CÉSAR, C. C.; MAMBRINI, J. Determinantes de desempenho dos alunos do ensino básico brasileiro: evidências do Saeb de 1997. In: FRANCO, C. *Promoção, ciclos e avaliação educacional*. Porto Alegre: Artmed, 2001. p. 121-153.

\_\_\_\_\_. Determinantes de desempenho dos alunos do ensino básico brasileiro: evidências do Saeb de 1997. In: FRANCO, Creso (Org.). *Avaliação, ciclos e promoção na educação*. Porto Alegre: Artmed, 2001.

SOARES, T. Machado. Modelo de 3 níveis hierárquicos para a proficiência dos alunos de 4ª série avaliados no teste de língua portuguesa do Simave/Proeb-2002. *Revista Brasileira de Educação*, n. 29, p. 73-87, maio/jul. 2004b.

STIMSON, James. Regression in space and time: a statistical essay. *American Journal of Political Science*, v. 29, p. 14-47, 1985.

WALTENBERG, Fábio D. Economic theories of the supply of education: historical evolution, current state, and perspectives. *Educ. Pesqui.*, v. 32, n. 1, p. 117-136, Jan./Apr. 2006. ISSN 1517-9702.

WOOLDRIDGE, J. M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge (USA): MIT Press, 2002.

## ANEXO

**Variável independente 1:** Docentes com ensino superior completo por escola [*docentes*]  
Proporção de docentes, por escola, atuando no ensino fundamental com grau de formação de nível superior completo.

**Variável independente 2:** Média de alunos por turma [*alunoturma*]  
O intuito aqui é medir o tamanho das classes escolares por meio do número médio de alunos por turma. A medida aqui apresentada, de acordo com o glossário do Edudatabrasil, “trata-se do quociente entre a matrícula inicial e o total de turmas informadas na data de referência do censo escolar, por série, grupo de séries e nível/modalidade de ensino”.<sup>1</sup>

**Variável independente 3:** Média de horas-aula diária [*horaula*]  
“Número médio de horas de permanência dos alunos na escola. Trata-se de uma média aritmética ponderada, cujo fator de ponderação é a matrícula na data de referência do Censo Escolar, por série, grupos de séries e nível de ensino”.<sup>2</sup>

**Variável independente 4:** Fator infra-estrutura [*finfra1*]  
Para analisar a infra-estrutura das escolas foi construída uma variável síntese das dependências existentes na escola chamada Índice de Infra-Estrutura. Este índice é o primeiro fator de uma análise de componentes principais que engloba vários itens do Censo Escolar que perguntam se a escola tem ou não tem determinada benfeitoria. São elas: diretoria, secretaria, sala de professores, biblioteca, laboratório de ciências, laboratório de informática, cozinha, depósito de alimentos, refeitório, pátio, quadra, parque infantil, dormitório, berçário, sanitário fora do prédio, sanitário dentro do prédio, sanitário adequado à pré-escola, sanitário adequado a alunos com necessidades especiais/acessibilidade.<sup>3</sup>

**Variável independente 5:** Fator infra-estrutura<sup>2</sup> [*finfra2*]  
Fator infra-estrutura elevado ao quadrado.

**Variável independente 6:** Biblioteca [*biblioteca*]  
Variável binária que tem valor 1 se na escola há biblioteca ou sala de leitura, ou 0 para o caso contrário.

**Variáveis independentes 7 e 8:** Localização da escola [*urbana e metropolitana*]  
As escolas foram classificadas pelo tipo de localização: rural, urbana e metropolitana, sendo as escolas localizadas em área rural aquelas utilizadas como referência.

**Variáveis independentes 9 e 10:** Proporção da população educada [*educa1 e educa2*]  
Através da variável relativa à escolaridade completa do Censo Demográfico foram construídas as proporções da população educada em determinadas faixas de escolaridade. Proporção da população analfabeta (*educa0*), proporção da população com até 8ª série completa (*educa1*) e proporção da população com ensino médio ou mais anos de escolaridade (*educa2*). Neste caso, a proporção da

<sup>1</sup> Edudatabrasil. Disponível em: <<http://www.edudatabrasil.inep.gov.br/>>.

<sup>2</sup> Idem.

<sup>3</sup> Esses são os itens comuns para toda série temporal. Outros itens podem ser encontrados para alguns anos, mas como não são comuns a todos não permitem a construção de fatores. A listagem completa das dependências arroladas no Censo é: diretoria, secretaria, sala de professores, almoxarifado, videoteca, sala para TV/Vídeo, laboratório de informática, laboratório de ciências, outros laboratórios/oficinas, brinquedoteca, auditório, solário, ginásio de esportes, cozinha, cantina, refeitório, lactário, parque infantil, dormitório, berçário, quadra de esportes descoberta, quadra de esportes coberta, piscina, depósito de alimentos, sanitário fora do prédio, sanitário dentro do prédio, sanitário adequado à pré-escola, sanitário adequado a alunos com necessidades especiais/acessibilidade, dependências e vias adequadas a alunos com necessidades especiais, lavanderia, fraldário, sala de recursos para alunos com necessidades educacionais especiais.

população analfabeta é utilizada como categoria de referência, por isso não aparece no modelo. A variável *educa1* é o resultado do agrupamento de duas categorias testadas separadamente que se comportaram de forma muito semelhante: proporção de população com até 4ª série e proporção da população de 5ª a 8ª série completa. A variável *educa2* agrega ensino médio com curso superior em função da pouca expressividade desta última categoria como proporção da população e da não significância dela sozinha em relação à categoria de referência.

**Variável independente 11:** Logaritmo da mediana da renda domiciliar *per capita* [*logrendam*]

**Variável independente 12:** Logaritmo do desvio padrão da renda domiciliar *per capita* [*logrendadp*]

**Variável independente 13:** Proporção de não-brancos por município [*propNB*].

**Variável independente 14:** Proporção de população em áreas urbanas por município [*propURB*]

**Variável independente 15:** Logaritmo da população municipal [*logpop*].

**Variável independente 16:** *Lag* da taxa de distorção idade/série escolar [*lagTDI*]

A *lag* da variável dependente é utilizada como controle de autocorrelação, trata-se da variável dependente no seu tempo  $t-1$ .



