ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Soares, Sergei; Sátyro, Natália

Working Paper

O impacto de infra-estrutura escolar na taxa de distorção idade-série das escolas brasileiras de ensino fundamental: 1998 a 2005

Texto para Discussão, No. 1338

Provided in Cooperation with:
Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: Soares, Sergei; Sátyro, Natália (2008) : O impacto de infra-estrutura escolar na taxa de distorção idade-série das escolas brasileiras de ensino fundamental: 1998 a 2005, Texto para Discussão, No. 1338, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília

This Version is available at:
https://hdl.handle.net/10419/91254

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

TEXTO PARA DISCUSSÃO N° 1338

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

Rio de Janeiro, maio de 2008

TEXTO PARA DISCUSSÃO N° 1338

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**

Rio de Janeiro, maio de 2008

* Pesquisador da Diretoria de Estudos Sociais do Ipea.

** Consultora do Ministério do Desenvolvimento Social.

Governo Federal

Ministro de Estado Extraordinário de Assuntos Estratégicos – Roberto Mangabeira Unger

Núcleo de Assuntos Estratégicos da Presidência da República

ipea **Instituto de Pesquisa Econômica Aplicada**

Fundação pública vinculada ao Núcleo de Assuntos Estratégicos da Presidência da República, o Ipea fornece suporte técnico e institucional às ações governamentais, possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro, e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente
Marcio Pochmann

Diretor de Administração e Finanças
Fernando Ferreira

Diretor de Estudos Macroeconômicos
João Sicsú

Diretor de Estudos Sociais
Jorge Abrahão de Castro

Diretora de Estudos Regionais e Urbanos
Liana Maria da Frota Carleial

Diretor de Estudos Setoriais
Márcio Wohlers de Almeida

Diretor de Cooperação e Desenvolvimento
Mário Lisboa Theodoro

Chefe de Gabinete
Persio Marco Antonio Davison

Assessor-Chefe de Comunicação
Estanislau Maria de Freitas Júnior

URL: http:/www.ipea.gov.br

Ouvidoria: http:/www.ipea.gov.br/ouvidoria

ISSN 1415-4765

JEL: 121

TEXTO PARA DISCUSSÃO

Publicação cujo objetivo é divulgar resultados de estudos desenvolvidos pelo Ipea, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Núcleo de Assuntos Estratégicos da Presidência da República.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

SINOPSE

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 desses insumos, nas escolas brasileiras de ensino fundamental, sobre a taxa de distorção idade-série escolar entre 1998 e 2005. A análise empírica é feita através de mais de um método de estimação, mas os principais resultados advêm do modelo de regressão linear para dados de painel com efeito fixo das escolas. O estudo se baseia em dados do Censo Escolar de 1998 a 2005 sobre uma coorte de 132.603 escolas, e, para os municípios, utilizaram-se dados do Censo Demográfico de 2000. O principal resultado é que os insumos escolares têm importância, sim, no desempenho educacional, principalmente no caso de escolas com maior precariedade das condições infra-estruturais.

ABSTRACT

The effect of school inputs upon educational outcomes has been extensively studies in the literature but there are still no conclusions as to its impacts. This paper presents results that suggest a strong impact of inputs upon age-grade lag for Brazilian primary schools between 1998 and 2005. We use more than one estimation method, but the results we believe are those from a fixed effects model based upon a panel of schools. The data come from the annual School Census from 1998 to 2005 which provides information on a cohort of 132.603 schools; municipal data come from the 2000 Demographic Census. Our most important result is that school inputs have strong positive effects upon educational outcomes, particularly in the case of those schools whose inputs are especially low.

SUMÁRIO

1 INTRODUÇÃO 7

2 REFERENCIAL TEÓRICO 7

3 MODELO DE REGRESSÃO LINEAR DE PAINEL COM EFEITO FIXO 9

4 RESULTADOS 14

REFERÊNCIAS 22

ANEXO 22

1 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 pelas condições extremamente precárias e também com os resultados educacionais inaceitáveis. Por outro lado, em várias dimensões essa 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 de Avaliação do Ensino Básico, manteve-se constante ou em queda em todo o país e as taxas de repetência caíram até o fim da década de 1990, permanecendo 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 tanto pode haver causalidade entre os resultados educacionais e as condições materiais, como pode não haver. Em outras palavras, é necessária uma abordagem mais cuidadosa para desvendar se há ou não causalidade entre os 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; e as 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 observado na última década em infra-estrutura nas escolas, os resultados educacionais seriam ainda piores.

O artigo organiza-se da seguinte forma: na seção 2 apresenta-se, brevemente, um resumo da literatura com a exposição das hipóteses testadas; na seção 3 apresentam-se a especificação formal do modelo e a descrição breve do censo escolar e do censo demográfico; na seção 4 os resultados são apresentados; as conclusões são expostas na seção 5 final.

2 REFERENCIAL TEÓRICO

Dois fatos contribuíram fortemente para o interesse acadêmico da relação insumos escolares e escola de qualidade: o primeiro é que, na última metade do século XX, observou-se, nos Estados Unidos e em outros países desenvolvidos, um aumento tanto do valor absoluto como do percentual do Produto Interno Bruto (PIB) gasto com educação, e uma concomitante queda nos escores das avaliações educacionais (HANUSHEK, 2003). Esse fato, entre outros, levou ao famoso Relatório Coleman, de 1966, em que se concluiu, a partir de um estudo com cerca de 600 mil alunos, que a escola exercia apenas efeitos marginais sobre os resultados escolares – e que os grandes fatores causais do desempenho seriam a família e o capital social trazido dela e do meio em que vivemos. A partir do Relatório Coleman, 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.

A análise de custos e benefícios educacionais e a análise de eficiência na alocação de recursos visando a 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 Hanushek que, em 1986 e 1997, apresentou metaanálises 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 infra-estrutura desta. Nos estudos compilados podemos citar 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 e resultados de professores em testes de avaliação.

Talvez a exposição feita anteriormente, 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 dessa tese. Autores como Lant Pritchett e Deon Filmer defendem em artigo de 1999 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 apenas 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 e onde as escolas são muito mais heterogêneas no que diz respeito à infra-estrutura, temos de 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 de condições infra-estruturais muito precárias, a melhoria dessas condições 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.

3 MODELO DE REGRESSÃO LINEAR DE PAINEL COM EFEITO FIXO

3.1 DADOS

Neste estudo, insumos escolares serão entendidos como: infra-estrutura de todo tipo, seja o número médio de alunos por turma; o número de horas-aula; os docentes com formação superior; a construção e a melhoria das dependências da escola; a existência de biblioteca ou sala de leitura e outros aspectos positivos. Infra-estrutura, nesse caso, é tudo aquilo que o dinheiro pode comprar.

O censo escolar contém dados de todas as escolas, de todos os níveis e todas as modalidades de ensino básico no Brasil. Ou seja, ele 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. 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 para a educação básica; de 32.296 para a educação infantil (creche); de 105.616 para a educação pré-escolar; de 162.727 para o ensino fundamental; de 23.561 para o ensino médio; de 7.053 para a educação especial; de 44.838 para a educação de jovens e adultos; e de 3.230 para a 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 em 2005. 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 através do Censo Demográfico de 2000 é possível analisar as características gerais da população; das pessoas responsáveis pelos domicílios; dos próprios domicílios e de seus moradores; e das condições habitacionais, educacionais e de renda.

Não serão considerados, a não ser como variáveis de controle, fatores extra-escolares, como a seleção de alunos e o *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 e %	Número de escolas em áreas urbanas e % (não-metropolitanas)	Número de escolas em áreas metropolitanas e %
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 e Soares (2007).

3.2 MEDIDAS

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 de idade, 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 Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP, 2006) 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 a 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.

Ou seja, o conceito de defasagem idade-série usado é a porcentagem de alunos defasados dois anos ou mais. É importante ressaltar que, quanto maior a distorção, pior o desempenho escolar. Desse 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 a relação não seja linear, ela 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 esse argumento, uma escola de boa qualidade seria um 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. O argumento acima é, lamentavelmente, a base de uma crença amplamente difundida entre pais, professores e o público leigo no nosso país. Embora esse 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 do ensino de baixa qualidade.[1] Convidamos os leitores que vêem a afirmação acima com ceticismo a consultarem Ribeiro (1991), 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. Como usamos aquela como *proxy* para esta, nas áreas rurais haverá um ruído indesejável exercendo influência nos resultados.

Outra limitação é que, da 5ª série em diante, é possível haver crianças que cursaram da 1ª à 4ª série em outra escola. Isto é, provavelmente a defasagem é 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 à 4ª série.

Para medir a infra-estrutura das escolas, utilizamos os seguintes indicadores:[2] 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, classificadas em 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 controlarmos 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*; a proporção de não-brancos; a proporção de população em áreas urbanas; e o logaritmo da população. Por fim, como variável de controle de autocorrelação e de legado utilizou-se a *lag* da taxa de distorção idade-série escolar, uma variável dependente no seu tempo $t-1$.

1. É possível que esse 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.

2. Mais detalhes no anexo.

3.3 VARIÁVEIS EXPLICATIVAS

As variáveis explicativas usadas encontram-se na tabela 2.

TABELA 2
Variáveis independentes

Variável	Tipo
Todos os modelos	
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
Apenas modelos Mínimos Quadrados	
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 de 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

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 a 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 essas informações simplesmente não existem no Brasil. Ou seja, mesmo reconhecendo-se a importância da origem socioeconômica dos estudantes para o seu desempenho escolar, não há essa informação para o formato analítico aqui proposto. Tampouco existem no censo escolar ou no censo demográfico informações sobre a gestão da escola, sobre o processo de seleção de alunos e o processo de formação de turmas. 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 nessa literatura. Por exemplo, no censo escolar há informações sobre o combustível do forno da escola, mas não há na literatura nenhuma menção a fornos a lenha ou a gás. Portanto, não usamos essa informação.

Finalmente, usamos nossa própria descrição 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 com 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.

3.4 MODELO

Meio século de avanço na econometria legou-nos grande quantidade 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, de variáveis omitidas e demais 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 o 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-mensuráveis 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 entre a defasagem idade-série mediante um modelo de Mínimos Quadrados Ordinários (MQO) para três anos da nossa amostra. Não duvidamos da forte endogeneidade relacionando o nível de insumos e os 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 modelo 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 –, ele não identifica as escolas, de modo que essa variação fica perdida junto com a variação no espaço. Acreditamos que 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 dessa mesma escola. Ou seja, estamos usando apenas a variação no tempo para identificar os efeitos de insumos escolares. *A priori*, é este o modelo no qual 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 dessas relações nos ajudarão na definição da ordem do polinômio que usaremos para cada variável.

4 RESULTADOS

4.1 NÃO-PARAMÉTRICOS

Os quatro painéis do gráfico 1 mostram duas estimações não-paramétricas cada. As linhas em preto pontilhadas denotam a densidade da distribuição de cada variável. Por exemplo, a porcentagem de docentes com formação superior tem uma distribuição fortemente bimodal – há um grupo grande de escolas onde nenhum docente tem formação superior, há outro grupo um pouco menor onde todos docentes a têm, e há o grupo intermediário de escolas, onde alguns professores contam com esse nível de formação e outros não, sendo que estas últimas se encontram distribuídas mais ou menos homogeneamente. 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.

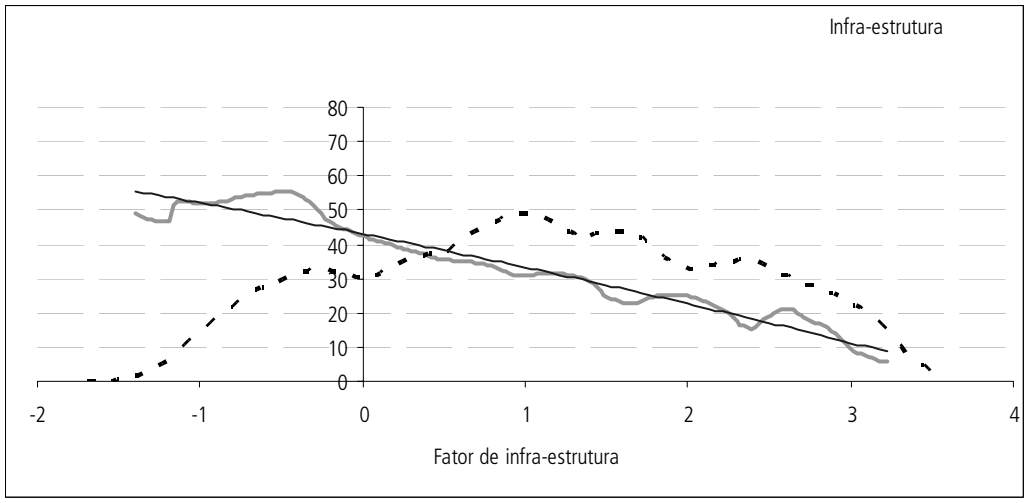
As linhas sólidas em cinza com tendência denotam a relação univariada, não-paramétrica, entre a taxa de defasagem idade-série e a variável em questão. Não há, nessa estimação, a presença de nenhuma outra variável nem qualquer tipo de controle para endogeneidade. O objetivo é fazer uma análise exploratória para indicar o grau do polinômio a ser usado nas outras estimações. 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.

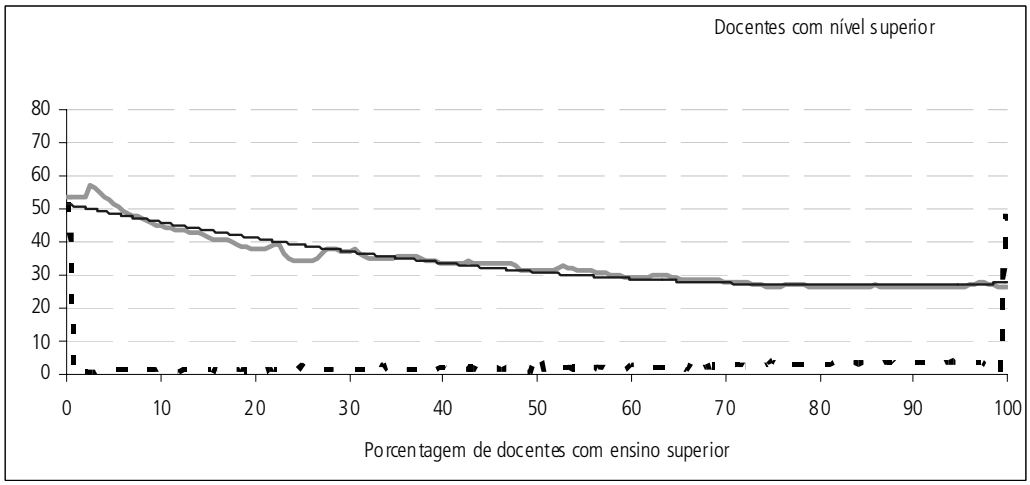
GRÁFICO 1

Densidades e relações com taxa de defasagem idade-série

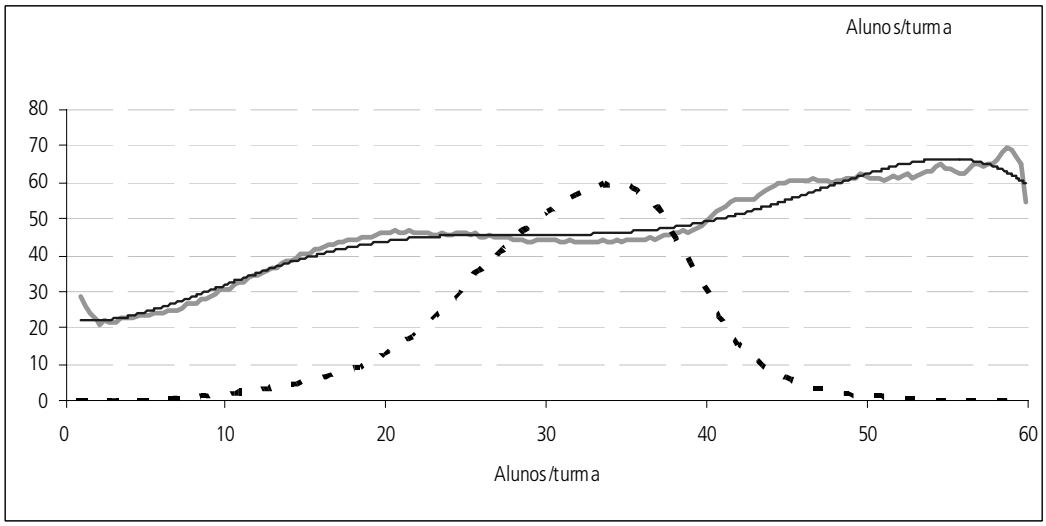
Painel 1
Infra-estrutura



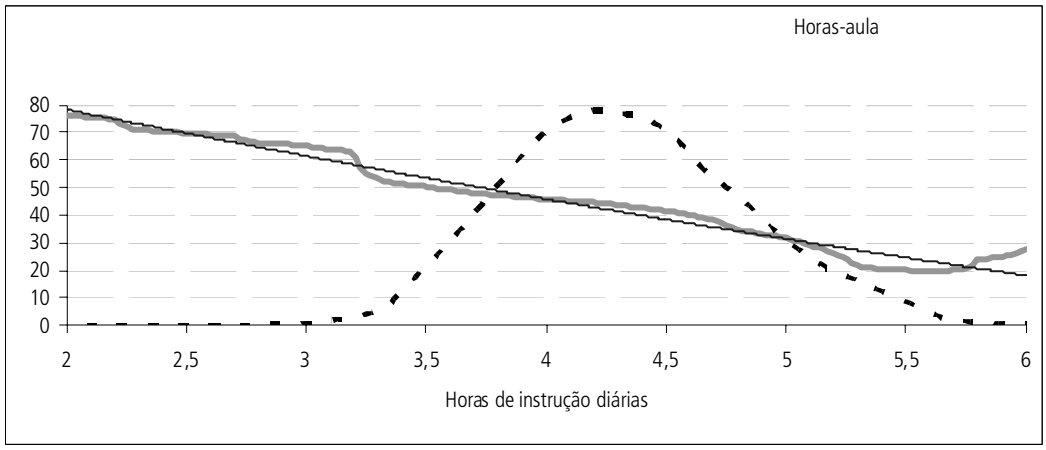
Painel 2
Porcentagem de docentes com nível superior



Painel 3
Alunos/turma



Painel 4
Horas-aula



Fonte: Censo escolar.

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

TABELA 3
Coeficientes e valores-p das estimações

Variável	(1) Empilhado		(2) 1999		(3) 2002		(4) 2004		(5) Modelo de 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 de infra-estrutura									0,00	
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		
Nível educacional dos adultos (%)										
Analfabetos	Base		Base		Base		Base			
Até a 8ª série	–0,09	0,00	–0,14	0,00	–0,12	0,00	–0,13	0,00		
Mais que a 8ª	–0,09	0,00	–0,13	0,00	–0,09	0,00	–0,02	0,07		
Renda domiciliar *per capita*										
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	

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 (p.p.) no modelo de efeitos fixos e tem como único efeito significativo uma redução de 0,17 p.p. 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. Nesse caso, somente a renda média tem comportamento diferente do esperado.

Em suma, as estimações dos modelos (1) a (4) mostram o padrão usual em estimações 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 estimações mostram efeitos fortes e significativos dessa variável. Talvez por que no Brasil o número de horas-aula seja excepcionalmente baixo, esse resultado seja tão forte.

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, exceto 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.

Os quatro painéis do gráfico 2 mostram o impacto do fator 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, apenas que estas são 5,6 p.p. maiores que a média das escolas com cinco horas de aula por dia.

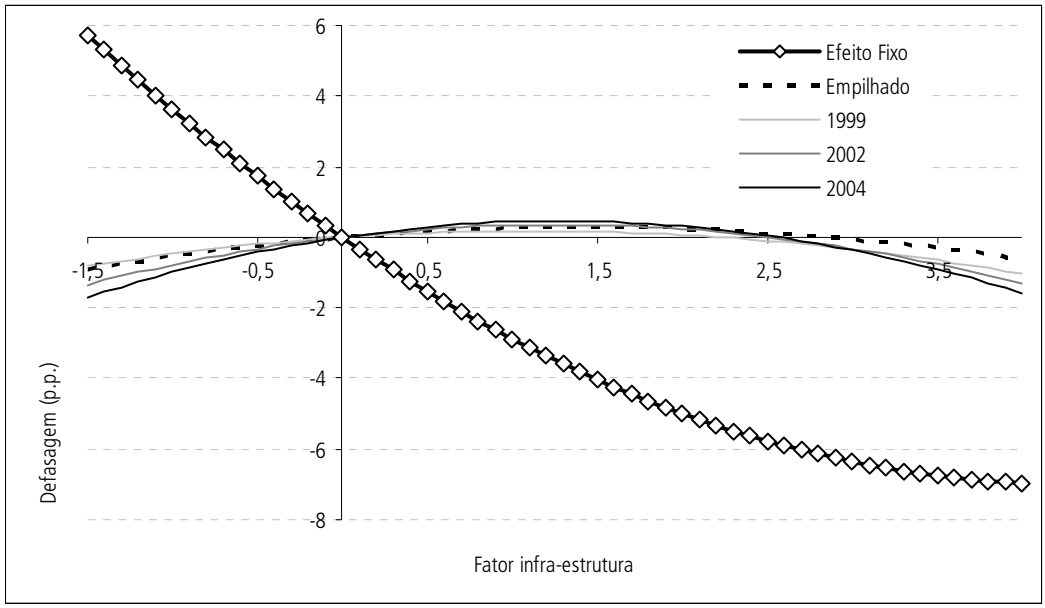
Em cada painel, o efeito estimado no modelo de efeitos fixos é mostrado por losangos; o efeito do modelo empilhado, por uma linha pontilhada; e os efeitos dos modelos de 1999, 2002 e 2004, por linhas cinza claro, cinza escuro e preta, respectivamente.

Conforme já mencionado, para todas as variáveis, o modelo de efeitos fixos estima o maior impacto. 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 p.p., segundo o modelo de efeitos fixos, contra apenas 2,2 p.p., segundo o modelo empilhado.

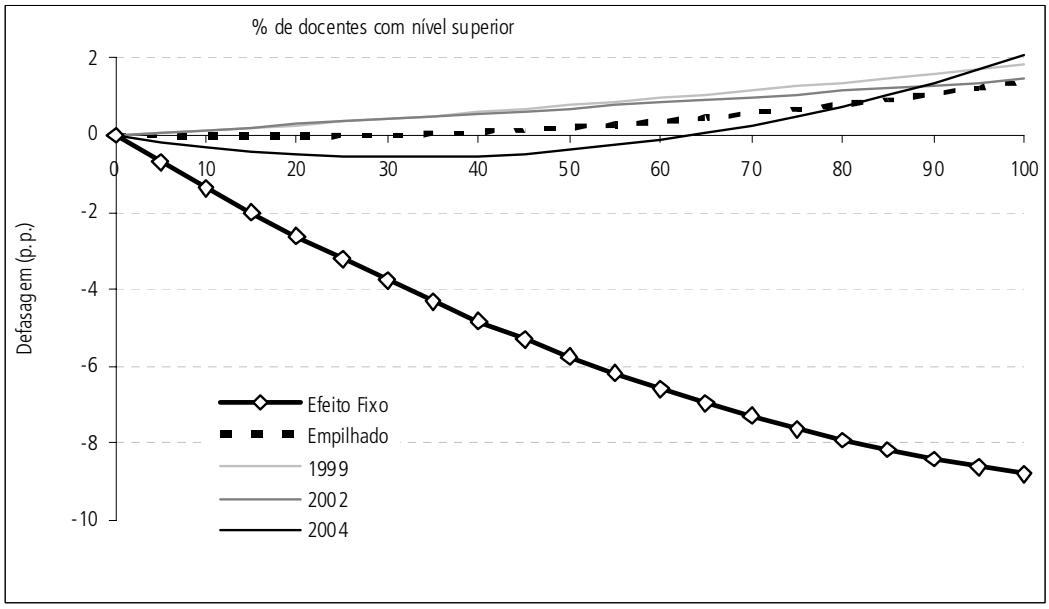
No caso das outras variáveis, a forma da curva difere bastante de modelo de efeitos fixos, por um lado, e das estimações ano a ano e o modelo empilhado, por outro. Por exemplo, no caso de infra-estrutura, o modelo de efeitos fixos estima resultados fortes: uma melhoria de infra-estrutura de –1,5 para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

GRÁFICO 2
Impactos dos insumos escolares sobre a taxa de defasagem idade-série

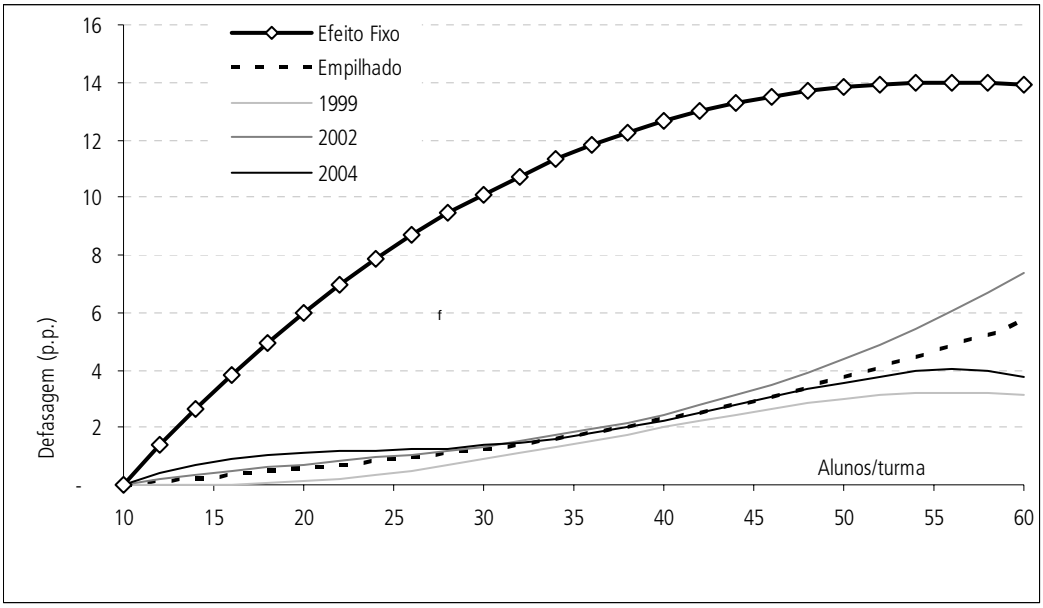
Painel 1
Infra-estrutura



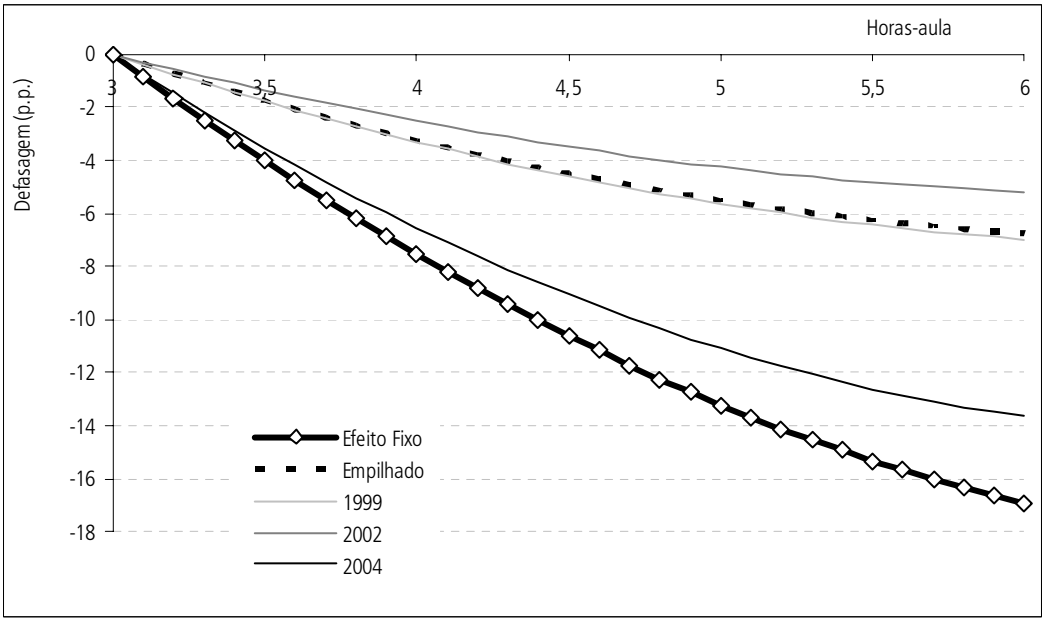
Painel 2
Porcentagem de docentes com nível superior



Painel 3
Alunos/turma



Painel 4
Horas-aula



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 de 1 p.p. O modelo de efeitos fixos discorda e afirma que seriam 2,5 p.p. A forma das curvas também é nitidamente diferente. Dada a ampla literatura indicando que o tamanho da turma seria uma variável com relação fraca com resultados educacionais, encaramos esse 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 tamanho de turma.

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 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 se trata de falar do impacto do *nível* de uma ou outra variável sobre a qualidade, e sim de *mudanças* em cada variável. Como já foi explicado, 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 de 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 onde 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 viés 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 Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb) foi criado durante o período estudado, levando a 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 artigo.

Finalmente cabe um comentário sobre a aplicação desses resultados na política educacional. O fato de uma medida ter impactos positivos sobre a qualidade educacional não quer dizer que seja uma boa medida. 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 três para quatro reduziria a defasagem em 5,7 p.p. 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 levaria a uma redução de 3,9 p.p. – significativa mas bem menor que 5,7 p.p.

5 CONCLUSÕES

A principal conclusão deste artigo 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 – e, quando estes estão disponíveis, de se tratar 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 sobrepujá-los. O que, sim, indicamos 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 tivemos acesso aos microdados da Prova Brasil, será possível usá-los para medir o 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, para tanto, metodologias em dois estágios, e 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 aqui examinadas. Métodos sugeridos são o uso de estimadores

que melhor adotam a estrutura de painel para lidar com endogeneidade, como Arellano-Bond, e as 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 aos encontrados na literatura internacional.

REFERÊNCIAS

CARD, D.; KRUEGER, A. B. *Labor market effects of school quality:* theory and evidence. Feb. 1996 (NBER Working Paper, n. 5.450).

HANUSHEK, E. A. The failure of input-based schooling policies. *Economic Journal,* v. 113, p. F64-F98, Feb. 2003.

INEP. *Glossário do Edudatabrasil* – Sistema de Estatísticas Educacionais. Disponível em: <http://www.edudatabrasil.inep.gov.br/>. Acesso em: 30 out. 2006.

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.

RIBEIRO, S. C. 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. S. *Anatomia das escolas brasileiras*: um estudo sobre a infra-estrutura das escolas de ensino fundamental com base no Censo Escolar de 1997 a 2005. Ipea, 2007 (Texto para Discussão). No prelo.

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 através do número médio de alunos por turma. A medida 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".[3]

3. Inep (2006).

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".[4]

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. Esse 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.[5]

Variável independente 5

Fator infra-estrutura^2 [*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, 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.

---

4. Idem.

5. 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 labaratórios/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.

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 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 da 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 da 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*].

O lag da variável dependente é utilizado como controle de autocorrelação, trata-se da variável dependente no seu tempo t–1.

© Instituto de Pesquisa Econômica Aplicada – **Ipea** 2008

EDITORIAL

Coordenação
Iranilde Rego

Revisão
Lucia Duarte Moreira
Alejandro Sainz de Vicuña
Eliezer Moreira
Elisabete de Carvalho Soares
Míriam Nunes da Fonseca

Editoração
Roberto das Chagas Campos
Aeromilson Mesquita
Camila Guimarães Simas
Camila Oliveira de Souza
Carlos Henrique Santos Vianna

Brasília
SBS – Quadra 1 – Bloco J – Ed. BNDES,
9º andar – 70076-900 – Brasília – DF
Fone: (61) 3315-5090
Fax: (61) 3315-5314
Correio eletrônico: editbsb@ipea.gov.br

Rio de Janeiro
Av. Nilo Peçanha, 50/609
20044-900 – Rio de Janeiro – RJ
Fone: (21) 3515-8522 – 3515-8426
Fax (21) 3515-8585
Correio eletrônico: editrj@ipea.gov.br

Tiragem: 135 exemplares

COMITÊ EDITORIAL

Secretário-Executivo

Marco Aurélio Dias Pires
SBS – Quadra 1 – Bloco J – Ed. BNDES,
9º andar – sala 912
70076-900 – Brasília – DF
Fone: (61) 3315-5406
Correio eletrônico: madp@ipea.gov.br