

ESTRATIFICAÇÃO EDUCACIONAL E PROGRESSÃO ESCOLAR POR SÉRIE NO BRASIL *

Eduardo L. G. Rios-Neto

Professor do Departamento de Demografia e do Cedeplar/UFMG

Cibele Comini César

Professora do Departamento de Estatística e do Cedeplar/UFMG

Juliana de Lucena Ruas Riani

Aluna de doutorado do programa de demografia do Cedeplar/UFMG

Este trabalho trata dos determinantes do desempenho escolar por meio da progressão em duas séries do ensino fundamental (progressões na 1ª e na 5ª série, uma vez que a 4ª série foi concluída). O trabalho faz uma decomposição do aumento dos anos médios de estudo para as várias coortes que entraram na escola entre 1945 e 1985. Essa decomposição foi feita utilizando o conceito de probabilidade de progressão escolar por série e mostra que as progressões na 1ª e na 5ª série explicam 79% dos ganhos nos anos de estudo no período.

Essas duas progressões são, por sua vez, as variáveis dependentes na estimação de um modelo de regressão logística hierárquica. Para tanto, foram utilizados os microdados de 12 Pesquisas Nacionais por Amostra de Domicílios (PNAD) das décadas de 1980 e 1990. No primeiro nível de hierarquia foram incluídas as características socioeconômicas das famílias. No segundo, são incluídas variáveis agregadas do local onde a família mora, como a escolaridade média dos professores. O resultado mais importante é que a escolaridade da professora afeta positivamente o intercepto da progressão para a 1ª série e, negativamente, o coeficiente positivo da escolaridade da mãe. Em outras palavras, a escolaridade das professoras é substituto para a escolaridade das mães, mostrando que melhores professores aumentariam o nível e reduziriam a desigualdade na distribuição da educação no Brasil. No caso da progressão na 5ª série o impacto é positivo no intercepto, mas o impacto no coeficiente da escolaridade da mãe não é estatisticamente significativo.

1 INTRODUÇÃO

A maioria dos estudos sobre desempenho educacional no Brasil concentra sua análise na escolaridade média, ou seja, considera como variável-resposta o número de anos completos de estudo de cada indivíduo. Este artigo trabalha com um conceito alternativo, associado ao desempenho escolar, denominado *probabilidade de progressão escolar por série* (PPS), que permite identificar onde, no processo escolar, ocorreram melhorias. Além de descrever a quais séries se deve atribuir a melhoria na escolaridade no Brasil, este estudo tenta estimar os determinantes do desempenho escolar, tanto individuais como sistêmicos. Isso é feito mediante o uso de modelos hierárquicos, em que se busca uma interação de variáveis em

* Os autores agradecem o apoio do CNPq, além do apoio institucional do projeto Análise da Progressão de Matrículas e População, financiado pelo MEC/Inep e executado pelo Cedeplar/UFMG. Agradecem também aos pareceristas anônimos pelas valiosas contribuições.

nível individual com variáveis em níveis mais agregados. A principal pergunta a ser respondida é em que medida a educação materna pode ser substituída pela escolaridade dos professores.

O restante deste trabalho se encontra organizado da seguinte forma: a Seção 2 contém uma breve discussão teórica e metodológica, centrada na escolha da variável dependente mais adequada para o estudo de desempenho escolar. A Seção 3 mostra que o conceito de PPS permite identificar a 1ª e a 5ª série como sendo responsáveis por quase toda a melhoria no perfil educacional da população brasileira. Em seguida, efetua-se uma análise dos determinantes dessas duas promoções, com base no agrupamento dos microdados das PNADs das décadas de 1980 e 1990. A aplicação dos modelos hierárquicos nesta análise permite a interação do nível familiar com o nível das políticas educacionais. O trabalho conclui com reflexões sobre a substituição da escolaridade materna pela escolaridade dos professores no desempenho escolar.

2 METODOLOGIA: A PROGRESSÃO ESCOLAR¹

O desenvolvimento que se segue visa clarear o conceito de progressão por série e sua relação com anos médios de estudos. Para uma determinada coorte define-se e_k como a probabilidade de transição para a série $k+1$, condicionada à conclusão da série k . Quer dizer:

$$e_k = \frac{P_{k+1}}{P_k} \quad (1)$$

onde:

P_{k+1} = número de pessoas na coorte que concluíram pelo menos a série $k+1$; e

P_k = número de pessoas na coorte que concluíram pelo menos a série k .

Em outras palavras, se e_0 representa a probabilidade de progressão da condição de nenhum ano de estudo para possuir um ano completo de estudo, então e_{k-1} representa a probabilidade de progressão da condição de $k-1$ anos completos de estudo para k anos completos de estudo, entre os que completaram $k-1$ anos de estudo.

Um segundo conjunto de fórmulas define a proporção da coorte de nascimento que possui pelo menos k anos de estudo. Tal proporção, que alcança pelo

1. Esta parte do trabalho, referente à analogia com a fecundidade, teve como base um texto preliminar, de autoria de Eduardo L. G. Rios-Neto, apresentado no Seminário Demografia da Educação, promovido pela Associação Brasileira de Estudos Populacionais (Abep), com apoio da Fundação Ford, realizado em Salvador em junho de 2001.

menos k anos de estudo, é o produto das PPSs até $k-1$ anos de estudo, que denominaremos aqui $e_{0,k}$.

$$\begin{aligned} e_{0,0} &= 1 \text{ toda a coorte possui pelo menos zero ano de estudo;} \\ e_{0,1} &= e_0 \text{ possui pelo menos um ano de estudo;} \\ &\dots \\ e_{0,k} &= \prod_{j=0}^{k-1} e_j \text{ possui pelo menos } k \text{ anos de estudo.} \end{aligned} \tag{2}$$

Esse segundo conjunto de fórmulas derivadas das PPSs, que gera a proporção com pelo menos k anos de estudo de uma população, permite também o cálculo da proporção da população com exatos k anos de estudo ou qualquer outro intervalo de escolaridade. Por exemplo, a população com exatos dois anos de estudo seria dada pela subtração ($e_{0,2} - e_{0,1}$).

Sabe-se que, se uma variável X assume apenas valores inteiros positivos, a sua média pode ser obtida por:

$$\mu(X) = \sum_{i=0}^{\infty} \Pr(X > i) = \sum_{i=1}^{\infty} \Pr(X \geq i) \tag{3}$$

ou seja, a escolaridade média é obtida por meio do somatório do conjunto da população que tenha pelo menos k anos de estudo, com k variando de um ao maior número possível de anos de estudo completos. Considerando $k = 16$, temos:

$$\bar{e} = \sum_{i=1}^{\infty} \Pr(X \geq i) = \sum_{i=1}^{16} e_{0,i} \tag{4}$$

Segundo Mare (1981), a estratificação educacional pode ser decomposta em dois efeitos. O primeiro é o efeito distribuição da escolaridade populacional entre as várias séries, que é afetado no tempo pelo processo de expansão do sistema escolar de uma determinada sociedade. O segundo é o efeito alocação, que é determinado pelos parâmetros de alocação, associados ao processo de estratificação social de acordo com variáveis como raça ou posição social.² O efeito alocação indica o impacto das variáveis de *background* socioeconômico sobre a progressão escolar.³

2. Os demógrafos são familiarizados com essa decomposição por causa dos estudos de fecundidade. Uma descrição detalhada e recente da técnica pode ser encontrada em Preston, Heuveline e Guillot (2001).

3. O trabalho de Silva e Hasenbalg (2001) foi pioneiro em levantar essa questão para o contexto brasileiro; foi esse trabalho que chamou a nossa atenção para o artigo de Mare e sua relação com as decomposições de PPSs, que estavam sendo feitas anteriormente.

Quando a variável dependente se refere aos anos médios de estudos completos, então o modelo de regressão é definido da seguinte forma:

$$\bar{e} = \beta_0 + \sum_{m=1}^M \beta_m X_m \quad (5)$$

onde \bar{e} indica os anos médios de estudos completos, X_m é um vetor de m variáveis socioeconômicas e β_m é o seu efeito sobre a escolaridade média. Normalmente, β_m é interpretado como uma estimativa do efeito alocação, mas é possível que reflita aumentos da escolaridade para toda a população. Além do mais, é impossível saber onde, no processo educacional, ocorrem as mudanças.

Os efeitos marginais de X_m em (5) correspondem ao somatório dos efeitos marginais de X_m das equações de progressão por séries e_{ks} , ponderado pelo peso dessas transições em cada série. Esses efeitos marginais em (5) são afetados tanto pelo efeito de alocação quanto pelo de distribuição. Uma análise de determinantes centrada no modelo de progressão escolar permite separar esses dois efeitos.

A estimativa dos determinantes das PPSs em nível individual deve seguir o modelo logito apresentado na equação abaixo:

$$P(P_i = 1) = F \left(\beta_0 + \sum_{m=1}^M \beta_m X_m + e_i \right) \quad (6)$$

onde F representa a distribuição acumulada logística.

As estimativas de β_m nessa equação determinam o efeito alocação puro das variáveis socioeconômicas, independentemente das não-linearidades decorrentes do efeito distribuição. Embora essas estimativas indiquem o efeito alocação puro no logito, o efeito marginal sobre as probabilidades e_{ks} não será linear no valor das probabilidades. Dado um coeficiente β_m fixo, o efeito marginal dependerá de onde está sendo avaliado.

3 O PAPEL HISTÓRICO DA PROGRESSÃO ESCOLAR NO BRASIL

Utilizando-se a PNAD de 1998, é possível estimar a tendência histórica das PPSs, bastando associar um grupo etário em 1998 com o período em que as pessoas desse grupo teriam sete anos de idade (idade adequada para a entrada no sistema escolar). Assim, é possível mostrar o papel histórico dessas progressões no cálculo dos anos médios de estudo completos de cada coorte. Finalmente, é possível decompor o papel isolado da mudança na progressão de cada série sobre a variação

total nos anos médios de estudo.⁴ A Tabela 1 lista as PPSs por coortes etárias em 1998. Essas coortes refletem períodos diferentes de entrada no ensino fundamental.

TABELA 1
PROBABILIDADE DE PROGRESSÃO POR SÉRIES, SEGUNDO COORTE DE NASCIMENTO, PARA A
POPULAÇÃO BRASILEIRA — 1998

PPSs	1981- 1985	1976- 1980	1971- 1975	1966- 1970	1961- 1965	1956- 1960	1951- 1955	1946- 1950	1941- 1945
e_0	0,909	0,898	0,879	0,870	0,845	0,811	0,759	0,708	0,666
e_1	0,979	0,978	0,972	0,969	0,967	0,958	0,948	0,936	0,928
e_2	0,958	0,956	0,948	0,942	0,936	0,911	0,893	0,872	0,859
e_3	0,939	0,928	0,921	0,914	0,895	0,882	0,846	0,819	0,810
e_4	0,875	0,844	0,827	0,778	0,696	0,625	0,556	0,484	0,460
e_5	0,870	0,859	0,868	0,892	0,928	0,943	0,941	0,953	0,950
e_6	0,896	0,898	0,905	0,919	0,927	0,933	0,943	0,942	0,951
e_7	0,880	0,885	0,897	0,903	0,904	0,925	0,928	0,919	0,940
e_8	0,789	0,763	0,743	0,755	0,751	0,758	0,745	0,708	0,699
e_9	0,871	0,913	0,928	0,944	0,955	0,965	0,977	0,977	0,973
e_{10}	0,818	0,884	0,904	0,915	0,926	0,928	0,940	0,951	0,930
e_{11}	0,303	0,313	0,347	0,382	0,413	0,457	0,473	0,424	0,411
e_{12}	0,735	0,891	0,915	0,942	0,950	0,958	0,979	0,978	0,957
e_{13}	0,611	0,833	0,886	0,900	0,913	0,926	0,951	0,962	0,964
e_{14}	0,540	0,786	0,823	0,842	0,856	0,861	0,904	0,899	0,911
e_{15}	0,250	0,325	0,306	0,317	0,312	0,316	0,325	0,326	0,356
Anos médios = $\sum_i \Pi_i e_i$	7,0	6,8	6,6	6,5	6,0	5,4	4,6	3,9	3,5

Fonte: PNAD de 1998.

4. Esta parte do trabalho teve como base um texto preliminar, de autoria de Eduardo L. G. Rios-Neto, apresentado no Seminário Demografia da Educação, promovido pela Associação Brasileira de Estudos Populacionais (Abep), com apoio da Fundação Ford, realizado em Salvador em junho de 2001.

De acordo com a Tabela 2, a coorte com 20 a 24 anos em 1998 (entrou no ensino fundamental entre 1981 e 1985) adquire, em média, quase sete anos completos de estudo. Já a coorte com 60 a 64 anos em 1998 (entrou no ensino fundamental entre 1941 e 1945) tinha em média três anos e meio de estudo. O número médio de anos completos de estudo dobrou entre as coortes extremas de análise.

TABELA 2

BRASIL: DECOMPOSIÇÃO DA VARIÇÃO TOTAL NOS ANOS MÉDIOS DE ESTUDO ENTRE A COORTE COM ENTRADA NO ENSINO FUNDAMENTAL DE 1941-1945 E 1981-1985, SEGUNDO A MUDANÇA DE CADA e_x

Coorte/simulação	Número médio de anos de estudo	Varição com relação à coorte 1981-1985	Porcentagem da variação atribuível a e_{ij}
1981-1985	6.973	-	100
1941-1945	3.491	3.482	0
Até $e_{0(41-45)}$	5.107	1.867	53,6
Até $e_{1(41-45)}$	4.878	2.096	6,6
Até $e_{2(41-45)}$	4.505	2.468	10,7
Até $e_{3(41-45)}$	4.136	2.838	10,6
Até $e_{4(41-45)}$	3.239	3.734	25,7
Até $e_{5(41-45)}$	3.312	3.661	-2,1
Até $e_{6(41-45)}$	3.354	3.619	-1,2
Até $e_{7(41-45)}$	3.391	3.582	-1,1
Até $e_{8(41-45)}$	3.344	3.629	1,4
Até $e_{9(41-45)}$	3.373	3.600	-0,8
Até $e_{10(41-45)}$	3.396	3.578	-0,6
Até $e_{11(41-45)}$	3.424	3.549	-0,8
Até $e_{12(41-45)}$	3.444	3.530	-0,6
Até $e_{13(41-45)}$	3.469	3.505	-0,7
Até $e_{14(41-45)}$	3.487	3.486	-0,5
Até $e_{15(41-45)}$	3.491	3.482	-0,1
Total			100,0

Fonte: PNAD de 1998.

É possível efetuar uma decomposição desse aumento de três anos de estudo entre as coortes extremas da Tabela 1. Para tanto, basta lembrar que o número médio de anos de estudo da coorte de 1941-1945 pode ser obtido segundo:

$$\bar{e}(41-45) = \sum_{j=1}^{16} e_{0,j}(41-45) = \sum_{j=1}^{16} \prod_{j=1}^{16} e_j(41-45)$$

Igualmente, $\bar{e}(81-85) = \sum_{j=1}^{16} e_{0,j}(81-85) = \sum_{j=1}^{16} \prod_{j=1}^{16} e_j(81-85)$ representa a média de anos de estudo da coorte de 1981-1985.

Ao substituir progressivamente as PPSs da coorte de 1981-1985 no cálculo da coorte de 1941-1945, podemos decompor a diferença entre as médias das duas coortes em diferenças nas PPSs de cada ano, de acordo com:

$$\bar{e}_{41-45,81-85,J} = \sum_{j=1}^J e_{0,j} + \sum_{i=J}^{16} e_{0,i} = \sum_{j=1}^J \prod_{j=1}^J e_j + \sum_{i=J}^{16} \prod_{i=J}^{16} e_i \quad (7)$$

A Tabela 2 apresenta esse exercício, apontando as PPSs mais importantes para o aumento observado na escolaridade brasileira entre a coorte de 1941-1945 e a de 1981-1985. O somatório da variação de todas as PPSs gera o aumento de 3,482 anos de estudo, que corresponde à diferença total na escolaridade média das duas coortes. Entretanto, é visível que 79% desse aumento se devem a aumentos em duas PPSs: e_0 e e_4 , que representam 54% e 26% da variação, respectivamente. Os Gráficos 1 e 2 mostram a evolução histórica das PPSs: associadas ao ensino fundamental e é visível que e_0 e e_4 apresentam as variações mais importantes no período.

Esse exercício de decomposição deixa evidenciada a importância dos estudos da progressão escolar para se avaliar o desempenho escolar. A experiência histórica sugere que as progressões e_0 (conclusão da 1ª série do ensino fundamental) e e_4 (conclusão da 5ª série do ensino fundamental, para aquelas pessoas que concluíram a 4ª série) são essenciais para o processo. Essas duas PPSs serão as variáveis dependentes do modelo de desempenho escolar a ser estimado a seguir.

GRÁFICO 1
BRASIL: TAXA DE PROMOÇÃO NAS QUATRO PRIMEIRAS SÉRIES — 1998

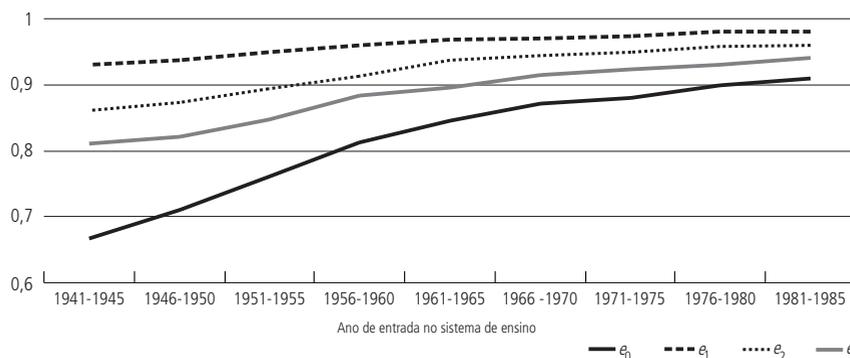
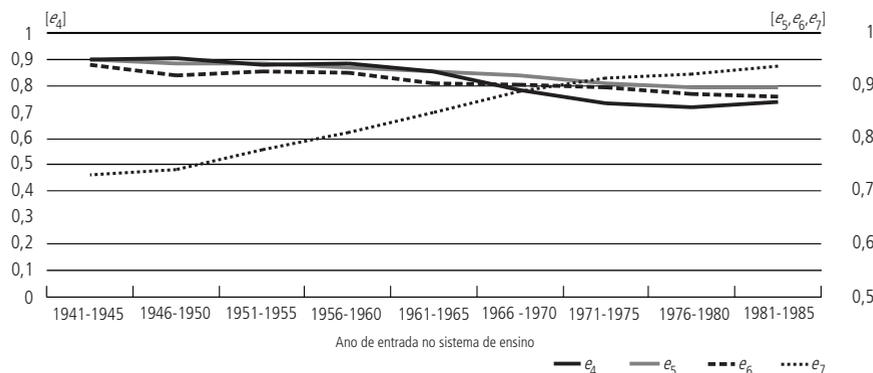


GRÁFICO 2
BRASIL: TAXA DE PROMOÇÃO DA 5ª À 8ª SÉRIE — 1998



4 UM MODELO HIERÁRQUICO DE DETERMINANTES DO DESEMPENHO ESCOLAR

A maioria dos estudos sobre os determinantes do desempenho escolar enfatiza a interação das esferas da família, escola e comunidade. As variáveis na esfera da família mostram a importância do estoque educacional dos pais, suas condições econômicas e sua estrutura ocupacional, entre outras variáveis socioeconômicas, na determinação do desempenho educacional dos filhos. Já as variáveis no nível das escolas e da comunidade indicam em que medida o contexto pode alterar o papel da condição socioeconômica dos pais sobre o desempenho escolar dos filhos. Este estudo pretende adotar as estimativas de modelo hierárquico para separar o nível familiar dos demais níveis. O papel da estrutura socioeconômica familiar será avaliado no nível individual, enquanto a oferta educacional será medida por variáveis de nível agregado.

No caso brasileiro, os estudos econômicos de desempenho escolar enfatizam a variável dependente anos médios de estudo concluídos. Barros e Lam (1996) representam um bom exemplo de aplicação nessa tradição. As duas variáveis independentes centrais para o estudo são a educação dos pais e a renda domiciliar, ambas com efeito esperado positivo. Já Barros *et alii* (2001) incorporam os seguintes determinantes do desempenho escolar: qualidade e disponibilidade dos serviços educacionais, custo de oportunidade do tempo, ambiente familiar, ambiente comunitário e características pessoais/geográficas. O resultado mostra a importância da escolaridade da mãe na determinação do desempenho escolar, mas também sugere que um ano a mais de escolaridade dos pais é equivalente a três anos a mais de escolaridade dos professores. Esse tipo de resultado será empreendido, aqui, mediante uma análise em dois níveis dos determinantes da progressão escolar.⁵

A decomposição da variável anos de estudo nas várias PPSs demonstrou a importância das transições para a 1ª e a 5ª série, definindo assim as duas variáveis dependentes para a estimativa do modelo hierárquico. Já os estudos revisados mostram o papel da educação materna, que apresenta impacto positivo sobre o desempenho educacional. A ocupação paterna é uma boa *proxy* para a classe social e renda domiciliar, tendo um sinal esperado positivo sobre o desempenho escolar.

O impacto de oferta será avaliado por meio da mensuração do efeito de três variáveis educacionais de nível dois: o salário dos professores, a educação dos professores e a razão professores do ensino fundamental/população de 7 a 14 anos de idade. A principal questão é ver em que medida essas variáveis reforçam as tendências de determinação no nível familiar ou os dois mecanismos operam como substitutos entre si.

4.1 A estrutura dos dados

Para poder gerar uma base de dados adequada ao modelo em dois níveis, optou-se por concatenar os dados retirados da PNAD para as décadas de 1980 e 1990, selecionando os anos de acordo com a PPS analisada. Tratou-se de concatenar os microdados de crianças em uma dada idade⁶ e sua respectiva família com os dados de nível dois referentes a um determinado ano e localização.

5. A sociologia tem contribuído muito para a discussão sobre o desempenho escolar e o uso da progressão escolar. A esse respeito ver Mare (1981) e Grusky e DiPrete (1990) no contexto internacional. Ainda na tradição sociológica e para o caso brasileiro, o trabalho de Silva e Hasenbalg (2001) introduz a discussão sobre o modelo logístico para a análise da progressão escolar.

6. As PPSs podem se referir a período, idade ou coorte. Nesse sentido, uma PPS pode apresentar um valor transitório se estiver captando uma coorte que ainda não tenha completado a trajetória escolar. Para o propósito de estudo de determinantes isto não é um maior impedimento, uma vez que aqueles que ainda não tenham concluído uma série estarão nessa situação como resultado de uma dinâmica entre os dois níveis de determinação. A escolha de crianças com 10 anos de idade para a análise da progressão na 1ª série e de 14 anos para a progressão na 5ª série reflete essa constatação. Em princípio, qualquer idade poderia ser escolhida como corte; o corte efetuado reflete um interesse em captar a moda da matrícula na série ao mesmo tempo em que se capta a criança ainda vivendo no domicílio.

No caso da probabilidade de conclusão da 1ª série, e_0 , foram selecionados os filhos de 10 anos de idade que moravam em domicílios com casais intactos (pai e mãe presentes). Dessa forma, os anos selecionados para o nível individual e familiar foram 1983 a 1985, 1987 a 1990, 1992, 1995 e 1997 a 1999, o que resultou em uma amostra de 80.700 casos. Para a aprovação na 1ª série, considerou-se que uma defasagem de dois anos seria adequada, uma vez que os alunos demoram pouco mais de dois anos para serem aprovados na 1ª série no Brasil. No caso da avaliação dos efeitos de segundo nível sobre e_0 , optou-se por admitir que as variáveis de nível 2 têm um efeito defasado sobre o aluno, uma vez que a progressão pode não ter sido efetuada no ano corrente da pesquisa.

Para a análise da probabilidade de conclusão da 5ª série, e_4 , o filtro utilizado foi o seguinte: filhos de domicílios intactos na idade de 14 anos. Nesse caso, a defasagem usada foi de quatro anos, pois a média de tempo gasto para se progredir na 5ª série é de cerca de dois anos, somando-se mais dois anos para se chegar às pessoas de 14 anos de idade. Portanto, os anos selecionados para o nível micro foram 1985 a 1990, 1992, 1993, 1996, 1997 e 1999, resultando em uma amostra de 36.911 casos.

Para o nível 2, a unidade de estudo é elaborada utilizando o ano, a unidade da federação (UF), a situação de domicílio (rural ou urbano) e o tipo de área (metropolitana ou não-metropolitana). Para cada ano existem 63 unidades, pois se trabalhou com 26 UFs (Tocantins foi agregado a Goiás) e nove regiões metropolitanas, todas divididas em rural e urbana. Destaca-se que seis estados e uma região metropolitana da região Norte não possuem a situação rural. Como se considerou que as variáveis de nível 2 possuem um efeito defasado sobre o nível individual e que tal defasagem é diferente, dependendo da série, a amostra do nível 2 para a 1ª série consiste de 819 casos e para a 5ª série, de 693 casos.⁷

Com relação às variáveis explicativas, no nível 1 (individual ou familiar) foram utilizadas a educação da mãe, oito variáveis indicadoras para o grupo de ocupação do pai, quatro indicadoras para a posição na ocupação do pai e uma indicadora para os que estão fora da ocupação. No nível 2 foram utilizados o salário médio do professor do ensino fundamental, a educação média do professor do ensino fundamental e a razão entre o professor e a população de 7 a 14 anos. Essas variáveis estão descritas detalhadamente no Apêndice.

7. Um dos pareceristas argumentou sobre o risco de erro amostral decorrente da pequena área rural das regiões metropolitanas, podendo causar forte viés nas estimativas do impacto das variáveis macro. Seguindo sua sugestão para utilizar o corte metropolitano, urbano não-metropolitano e rural não-metropolitano, obteve-se resultados similares às estimativas originais. Esses resultados poderão ser disponibilizados pelos autores.

4.2 O modelo hierárquico

O modelo estimado é um modelo de regressão logística multinível (ou hierárquico). O modelo logístico hierárquico mais simples seria um modelo de efeito aleatório no intercepto. Nesse caso, o modelo de nível 1 seria especificado pela equação (8) e o modelo de nível 2 incorporaria um termo aleatório.

Nível 1

$$P(e_{ij}) = F\left(\beta_{0j} + \sum_k \beta_{kj} X_{kij} + \varepsilon_{ij}\right) \quad (8)$$

Nível 2

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (9)$$

$$\beta_{kj} = \gamma_{k0} \quad k \neq 0 \quad (10)$$

onde i indexa os indivíduos (unidades de nível 1) e j indexa as unidades de nível 2; F representa a densidade de probabilidade acumulada logística.

O coeficiente β_0 , que corresponde à constante, contém um termo aleatório e é chamado de variável de efeito aleatório, mas os outros coeficientes não contêm esse termo e são, portanto, denominados variáveis de efeito fixo.

O modelo pode ser expandido, considerando que um subconjunto das variáveis independentes tem efeito aleatório. Dividindo-se o vetor X_k de variáveis independentes em dois vetores — (X_{k1}) contendo as variáveis de efeito fixo e (X_{k2}) contendo as variáveis de efeito aleatório —, obtém-se um segundo modelo logístico hierárquico, com efeito aleatório no intercepto e nos coeficientes das variáveis X_{k2} . O modelo é escrito da seguinte forma:

Nível 1

$$P(e_{ij}) = F\left(\beta_{0j} + \sum_k \beta_{kj} X_{kij} + \varepsilon_{ij}\right) \quad (11)$$

Nível 2

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (12)$$

$$\beta_{k2j} = \gamma_{k20} + u_{k2j} \quad (13)$$

$$\beta_{k1j} = \gamma_{k1j} \quad (14)$$

Os dados do modelo 1 estimados nas Tabelas 3 e 4 referem-se a uma estimativa do modelo básico coerente com as equações (11) a (14). Os efeitos do intercepto e da variável educação da mãe são aleatórios, enquanto as demais variáveis — indicadoras de ocupação e de posição na ocupação do pai — têm efeito fixo.

A próxima etapa na especificação dos modelos hierárquicos surge quando, além de se admitir que o efeito do intercepto ou de algumas variáveis de nível 1 é aleatório, admite-se também que variáveis de nível 2, avaliadas nas unidades macro j , Z_j , explicam os coeficientes aleatórios. Os modelos 2 e 3 estimados nas Tabelas 3 e 4 são dessa natureza. A equação de nível 1 para o modelo final é dada pela equação (15), e as equações de nível 2 são dadas por (16), (17) e (18):

Nível 1

$$P(e_{ij}) = F\left(\beta_{0j} + \beta_{1j}X_{1j} + \sum_{k>1} \beta_{kj}X_{kij} + \varepsilon_{ij}\right) \quad (15)$$

Nível 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \quad (16)$$

$$\beta_{1j} = \gamma_{k10} + \gamma_{11}Z_j + u_{1j} \quad (17)$$

$$\beta_{kj} = \gamma_{kj}, k > 1 \quad (18)$$

As Tabelas 3 e 4 apresentam estimativas das equações (8) a (10) no modelo 1 e das equações (15) a (18) nos modelos 2 e 3. A Tabela 3 se refere à probabilidade de promoção na 1ª série (e_0). A Tabela 4 se refere à probabilidade de promoção na 5ª série (e_4), uma vez que a pessoa concluiu a 4ª série. As variáveis de nível 1 com termo aleatório são o intercepto e a educação da mãe. As variáveis Z_j são o salário médio do(a) professor(a) do ensino fundamental na unidade j , a educação média do(a) professor(a) do ensino fundamental na unidade j e a razão entre o número de professores do ensino fundamental e a população em idade escolar adequada para o ensino fundamental (7 a 14 anos de idade) na mesma unidade.

Finalmente, as variáveis independentes do nível 1 estão expressas em termos de desvio em relação à média total da amostra. Esse procedimento faz parte das alternativas do pacote estatístico utilizado para a aplicação do modelo hierárquico, HLM-5.⁸ No caso do modelo logístico hierárquico, a vantagem de se centrar as variáveis independentes na média consiste na estimativa da probabilidade de

8. HLM é um pacote estatístico de Scientific Software International (SSI), sigla que quer dizer Hierarchical Linear and Nonlinear Modeling.

TABELA 3
PROBABILIDADE DE PROMOÇÃO NA 1ª SÉRIE — e_0

Efeito fixo	Modelo 1	Valor-p	Modelo 2	Valor-p	Modelo 3	Valor-p
Intercepto	1,794	0,000	-0,502	0,001	-0,043	0,784
Salário médio do professor					0,044	0,000
Educação média do professor			0,216	0,000	0,143	0,000
Professor/população em idade escolar					-5,112	0,003
Educação da mãe	0,279	0,000	0,401	0,000	0,388	0,000
Salário médio do professor					-0,001	0,357
Educação média do professor			-0,011	0,000	-0,008	0,057
Professor/população em idade escolar					-0,225	0,526
Ocupação do pai						
Técnicas e afins	0,739	0,000	0,693	0,000	0,689	0,000
Administrativas	0,926	0,000	0,886	0,000	0,878	0,000
Indústria de transformação	0,405	0,000	0,345	0,000	0,337	0,000
Comércio	0,547	0,000	0,484	0,000	0,480	0,000
Transporte e comunicação	0,699	0,000	0,641	0,000	0,635	0,000
Prestação de serviços	0,509	0,000	0,459	0,000	0,446	0,000
Outras	0,315	0,000	0,261	0,000	0,255	0,000
Posição na ocupação do pai						
Empregado com carteira	0,359	0,000	0,355	0,000	0,348	0,000
Conta-própria	0,255	0,000	0,258	0,000	0,257	0,000
Empregador	0,197	0,019	0,206	0,015	0,209	0,014
Fora da ocupação	0,596	0,000	0,539	0,000	0,535	0,000
Efeito aleatório						
Intercepto	0,791	0,000	0,508	0,000	0,453	0,000
Educação da mãe	0,006	0,000	0,005	0,000	0,005	0,000
Cov (intercepto, educmãe)	-0,003		0,012		0,012	
Proporção da variância explicada — intercepto			0,357		0,428	
Proporção da variância explicada — educação da mãe			0,128		0,111	

TABELA 4
PROBABILIDADE DE PROMOÇÃO NA 5ª SÉRIE — e_4

Efeito fixo	Modelo 1	Valor-p	Modelo 2	Valor-p	Modelo 3	Valor-p
Intercepto	1,058	0,000	-0,534	0,000	-0,424	0,004
Salário médio do professor					0,013	0,004
Educação média do professor			0,145	0,000	0,106	0,000
Professor/população em idade escolar					3,688	0,007
Educação da mãe	0,184	0,000	0,220	0,000	0,234	0,000
Salário médio do professor					0,001	0,351
Educação média do professor			-0,003	0,224	-0,005	0,188
Professor/população em idade escolar					-0,139	0,661
Ocupação do pai						
Técnicas e afins	0,760	0,000	0,686	0,000	0,673	0,000
Administrativas	0,698	0,000	0,636	0,000	0,622	0,000
Indústria de transformação	0,240	0,000	0,158	0,000	0,145	0,001
Comércio	0,458	0,000	0,371	0,000	0,359	0,000
Transporte e comunicação	0,329	0,000	0,253	0,000	0,241	0,000
Prestação de serviços	0,321	0,003	0,232	0,031	0,223	0,039
Outras	0,181	0,000	0,111	0,022	0,098	0,043
Posição na ocupação do pai						
Empregado com carteira	0,040	0,291	0,048	0,194	0,053	0,155
Conta-própria	0,024	0,519	0,061	0,101	0,065	0,081
Empregador	0,127	0,116	0,157	0,053	0,165	0,041
Fora da ocupação	0,337	0,000	0,267	0,000	0,259	0,000
Efeito aleatório						
Intercepto	0,233	0,000	0,138	0,000	0,131	0,000
Educação da mãe	0,002	0,000	0,002	0,000	0,002	0,000
Cov (intercepto, educmãe)	0,001		0,004		0,003	
Proporção da variância explicada — intercepto			0,408		0,436	
Proporção da variância explicada — educação da mãe			0,033		0,022	

progressão prevista no valor médio das variáveis independentes simplesmente a partir do intercepto. Isso porque, na média, os valores das variáveis independentes são iguais a zero. Tal estratégia facilita também os exercícios de simulação. Já as variáveis do nível 2 são medidas em termos de seus valores métricos.

5 ANÁLISE DOS RESULTADOS

A análise das Tabelas 3 e 4 se inicia pela explicação da variância. Os componentes da variância estimada relativos ao intercepto e ao coeficiente angular da educação da mãe no modelo 1 são apresentados na parte inferior das tabelas. Essa variância é denominada variância não-condicional. À medida que as variáveis do nível 2 (variáveis de nível macro) são incluídas na análise, a variância passa a ser condicional. A proporção da variância do intercepto e do coeficiente de educação da mãe, explicada pelas variáveis de nível 2, é calculada para os modelos 2 e 3 e apresentada no final das tabelas. Essa proporção é dada pela seguinte fórmula [Bryk e Raudenbush (1992)]:

$$\% \text{ da variância explicada} = \frac{\hat{\tau}_{qq(\text{n\~{a}o-condicional})} - \hat{\tau}_{qq(\text{condicional})}}{\hat{\tau}_{qq(\text{n\~{a}o-condicional})}} \quad (19)$$

A análise do efeito aleatório do modelo 1 da Tabela 3 mostra a variância não-condicional do intercepto e do coeficiente da educação da mãe. A análise do valor- p mostra ser aceitável a hipótese de esses dois coeficientes serem aleatórios no nível 2. O modelo 2 testa o efeito isolado da educação média do(a) professor(a). A educação média do(a) professor(a) explica cerca de 36% do componente aleatório do intercepto (modelo 2). A educação do(a) professor(a) explica cerca de 13% da variância do coeficiente aleatório da educação da mãe (modelo 2).

O modelo 3 incorpora as três variáveis de nível 2 simultaneamente. Embora a variância explicada do intercepto suba para cerca de 43%, o modelo não acrescenta variância explicada ao coeficiente estimado para a variável educação materna. Como parece haver uma certa multicolinearidade entre os três indicadores de nível 2, estimados simultaneamente no modelo 3, escolheu-se o modelo 2, para efeito de simulação e análise dos efeitos fixos.

No caso da promoção na 5ª série para aqueles que completaram a 4ª série, uma análise da Tabela 4 mostra uma variância não-condicional mais baixa que a observada no caso da promoção na 1ª série, embora o componente aleatório seja estatisticamente significativo. No modelo 2, a variância explicada dos parâmetros estimados — incluindo exclusivamente a variável macro escolaridade média do(a) professor(a) — é de aproximadamente 41% da variância do intercepto, embora de apenas 3% da variância do coeficiente estimado da educação da mãe. Ressalta-se o fato de a proporção da variância do parâmetro de educação materna, explicada

pela variável macro, ser bem mais baixa no caso da promoção na 5ª série (Tabela 4) do que na 1ª série (Tabela 3).

Ainda no que tange ao efeito aleatório, a co-variância entre o intercepto e a educação materna é positiva na maior parte dos casos das Tabelas 3 e 4, embora os valores sejam baixos. Uma co-variância positiva significa que as unidades de nível 2, que promovem mais na série estudada, são aquelas em que o efeito da educação materna é maior.

Após a análise do efeito aleatório consideram-se os resultados dos efeitos fixos. Estes são analisados graficamente, por intermédio da simulação dos efeitos total, direto e indireto das variáveis de interesse.

Os dados das Tabelas 3 e 4 (1ª e 5ª séries do ensino fundamental) mostram que todas as variáveis independentes consideradas no nível 1 são estatisticamente significativas. As variáveis de ocupação, posição na ocupação e fora da ocupação (desemprego e inatividade econômica) do pai merecem uma referência específica. No caso das ocupações, a categoria omitida refere-se aos pais em ocupações agrícolas. Por isso mesmo o impacto de todas as ocupações na probabilidade de promoção é positivo, embora ocupações de maior *status* socioeconômico apresentem impacto positivo maior (por exemplo, ocupações técnicas e afins, além das administrativas). No caso da posição na ocupação do pai, a categoria omitida é a ocupação dos empregados sem carteira. Sendo assim, não surpreende que os coeficientes estimados sejam positivos. Comparando-se o valor dos coeficientes estimados para a aprovação na 1ª série (Tabela 3) com a aprovação na 5ª série (Tabela 4), observa-se uma redução no impacto das ocupações sobre a promoção. Esses coeficientes das Tabelas 3 e 4 não fornecem uma clara idéia sobre o efeito marginal das mudanças na estrutura ocupacional sobre a taxa de promoção. Para melhor entender esse efeito marginal, uma simulação a partir do modelo 2 é apresentada nos Gráficos 3 e 4.

Os exercícios de simulação apresentados nos Gráficos 3 e 4 mostram o pequeno efeito marginal da estrutura ocupacional quando se contrastam as diferentes estruturas ocupacionais e de posição na ocupação entre dois estados totalmente díspares com a média nacional. Apesar da significância estatística das variáveis referentes à estrutura social dos pais, o efeito marginal na progressão seriada é mínimo. Além disso, o impacto declina entre a 1ª e a 5ª série.

Analisando o modelo 2 na Tabela 3, observa-se que a educação materna apresenta um efeito altamente positivo na probabilidade de progressão na 1ª série. A educação média do(a) professor(a) no nível macro, por sua vez, causa um efeito positivo e direto na promoção por meio do efeito intercepto, e um efeito negativo e significativo através do coeficiente de educação materna. Em outras palavras, a educação do(a) professor(a) no nível macro, uma variável *proxy* para polí-

GRÁFICO 3
IMPACTO DA OCUPAÇÃO E POSIÇÃO NA OCUPAÇÃO DO PAI NA PROBABILIDADE DE
PROGRESSÃO NA 1ª SÉRIE

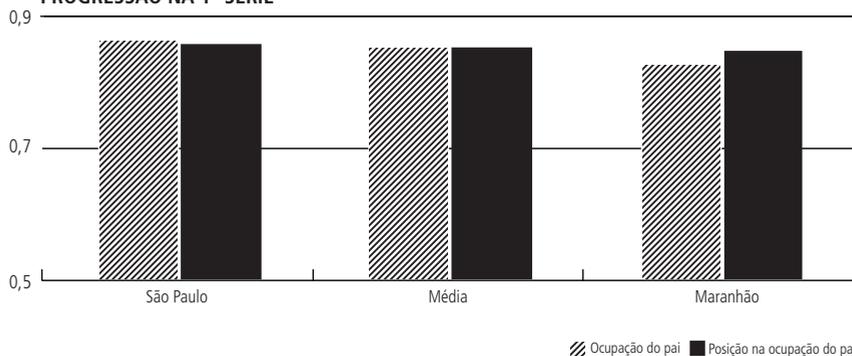
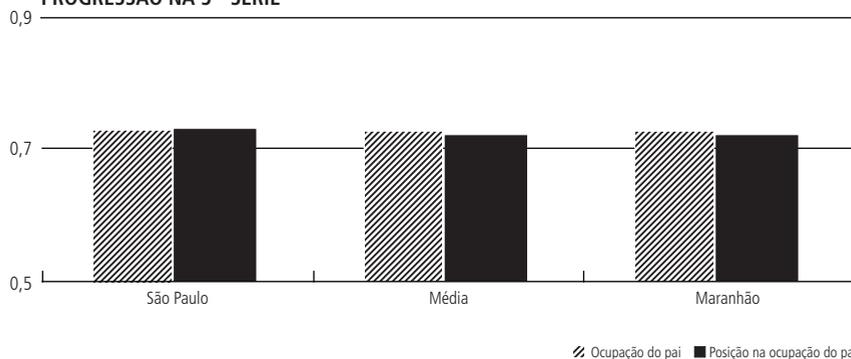


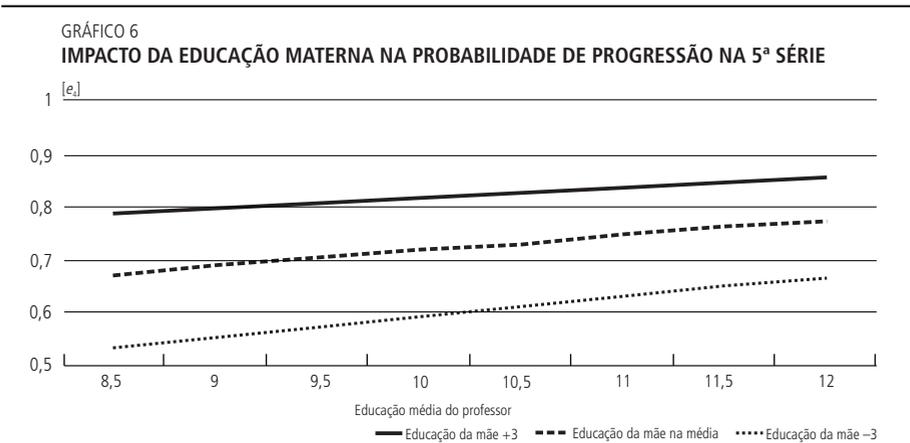
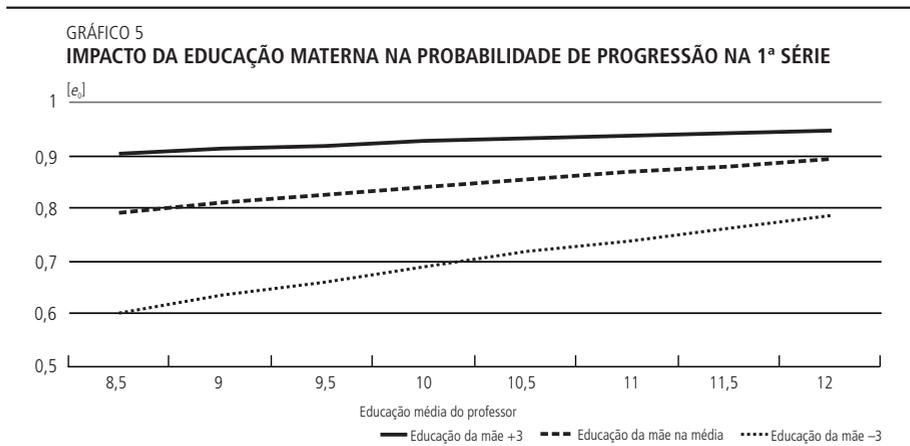
GRÁFICO 4
IMPACTO DA OCUPAÇÃO E POSIÇÃO NA OCUPAÇÃO DO PAI NA PROBABILIDADE DE
PROGRESSÃO NA 5ª SÉRIE



tica de qualidade educacional, é, em alguma medida, substituta para a educação materna na determinação da promoção na 1ª série.

A análise do modelo 2 na Tabela 4 mostra um efeito positivo e significativo da educação materna, embora menor no caso da 5ª série do que na promoção na 1ª série. O efeito direto da educação média do(a) professor(a) sobre o intercepto é positivo e significativo, mas também menor do que no caso da 1ª série. O efeito substituição da educação do(a) professor(a) no nível macro com relação à educação materna é nulo, uma vez que o coeficiente estimado não é estatisticamente diferente de zero.

Os Gráficos 5 e 6 apresentam os exercícios de simulação a partir do modelo 2, tratando simultaneamente do impacto da educação da mãe e da educação do(a) professor(a) do ensino fundamental no nível 2.



Os gráficos corroboram a análise dos coeficientes estimados nas Tabelas 3 e 4. No caso do Gráfico 5, a visualização do efeito substituição entre educação materna e educação do(a) professor(a) é clara. A linha média é a base (*baseline*) da simulação. Tudo o mais constante, o aumento da educação materna em três anos acima e abaixo da média da amostra faz crescer a probabilidade de progressão na 1ª série. Esse crescimento pode ser da ordem de 30 ou 10 pontos percentuais dependendo do nível de educação média do(a) professor(a) no nível macro. No caso de pobreza, quando a escolaridade da mãe é três anos abaixo da média e a escolaridade média do(a) professor(a) é a mais baixa possível (8,5 anos de estudo), então um aumento de três anos e meio de estudo entre os professores é equivalente a um aumento de três anos de estudo na educação materna, levando à probabilidade de progressão para a média.

A análise do Gráfico 6 mostra um impacto positivo da educação materna sobre a promoção na 5ª série, embora menor do que aquele observado na 1ª série. Entretanto, na 5ª série o efeito substituição é praticamente inexistente, fazendo com que as três curvas sejam praticamente paralelas.

6 COMENTÁRIOS FINAIS

Este trabalho ressalta a importância da progressão por série no estudo do desempenho escolar. O componente descritivo do trabalho destaca o papel histórico da 1ª e da 5ª série do ensino fundamental no aumento dos anos de estudo da população brasileira.

No caso dos determinantes do desempenho escolar, um primeiro ponto se refere ao fato de que, apesar de as variáveis paterna (ocupação) e materna (educação) serem estatisticamente significativas na determinação da progressão por série (1ª e 5ª), o efeito marginal das mães é muito maior que o dos pais.

Tanto os determinantes maternos quanto os paternos são mais fortes na determinação da progressão na 1ª série do que na 5ª série. Esse declínio da importância do componente socioeconômico nas séries mais elevadas é previsto pela literatura sociológica, uma vez que a seleção começa a ter um papel importante na determinação do universo de alunos em cada série.

O trabalho mostra também como as estimativas de modelos hierárquicos podem ajudar na avaliação do impacto de políticas macro sobre o comportamento individual. Nessa linha, a principal conclusão do trabalho se refere ao efeito substituição entre a escolaridade média dos professores e a escolaridade materna, efeito esse que é importante, particularmente na probabilidade de progressão na 1ª série. Uma possível implicação política do trabalho seria que os melhores professores de escolas públicas deveriam lecionar preferencialmente nas séries mais baixas e nas regiões e áreas das cidades onde os alunos fossem filhos de mães menos escolarizadas; nesse caso, o efeito substituição de equidade estaria sendo potencializado.

APÊNDICE

Glossário das variáveis explicativas

Educação da mãe: número de anos de estudo que varia de 0 a 16 anos.

Indicadoras de ocupação do pai: foram utilizados os grupos de ocupação do IBGE, que divide as ocupações em oito grupos: técnicas e afins, administrativas, agropecuária, indústria de transformação, comércio, transporte e comunicação, prestação de serviços e outras ocupações. Para não perder muita informação, foi considerada uma indicadora para inativo.

Indicadora de posição na ocupação do pai: empregado com carteira, empregado sem carteira, empregador e conta-própria.

Salário médio do(a) professor(a): esta variável considerou apenas os professores do ensino fundamental. O salário total foi deflacionado, tendo como ano-base o de 1998, e dividido pelo total de horas trabalhadas. Portanto, é o salário-hora médio do professor.

Educação média do(a) professor(a): a variável considerou apenas os professores do ensino fundamental; ela varia de 0 a 16 anos de estudo.

Número de professores por população em idade escolar: refere-se ao total de professores do ensino fundamental dividido pela população em idade escolar do ensino fundamental (população de 7 a 14 anos).

ABSTRACT

This paper decomposes the increase in average educational levels between cohorts born in 1945 and 1985 using the concept of school progression probability. The most important result in this decomposition is that two transitions, to first and fifth grades, explain 79% of the increase in schooling attainment between 1945 and 1985.

These two probabilities are, in their turn, used as dependent variables in a hierarchical logistic regression estimated using the PNAD microdata from 12 years during the 1980 and 1990. In level one of the regression were included family and individual variables. In the second level, averages of school inputs in the region where the family lives, such as average schooling of teachers. The most important result is that teacher schooling positively affects the intercept and negatively affects the impact of the mother's educational level. This means that more investment in schooling will improve both the level and inequality in the educational distribution. This effect is stronger in the transition to first grade and weaker in the transition into fifth grade.

BIBLIOGRAFIA

- BARROS, R. P. de, LAM, D. Income and educational inequality and children's school attainment. Chap 12. In: BIRDSALL, N., SABOT, R. H. (eds.). *Opportunity foregone — education in Brazil*. Washington: Inter-American Development Bank, The Johns Hopkins University, 1996.
- BARROS, R. P. de, MENDONÇA, R., SANTOS, D. D., QUINTAES, G. Determinantes do desempenho educacional do Brasil. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, v. 31, n. 1, abr. 2001.
- BARRY, J. R. *Probabilidade: um curso em nível intermediário*. CNPq: Projeto Euclides, 1981.
- BRYK, A. S., RAUDENBUSH, S. W. *Hierarchical linear models: applications and data analysis methods*. California: Newbury Park, Sage Publications, 1992.
- GRUSKY, D. B., DIPRETE, T. A. Recent trends in the process of stratification. *Demography*, v. 27, n. 4, p. 617-637, Nov. 1990.

- LAM, D., LEVISON, D. Declined inequality in schooling in Brazil and its effects on inequality in earnings. *Journal of Development Economics*, v. 37, p. 199-225, 1992.
- MARE, R. D. Social background composition and educational growth. *Demography*, v. 16, n. 1, p. 55-71, Feb. 1979.
- . *Change and stability in educational stratification*. 1981.
- PRESTON, S. H., HEUVELINE, P., GUILLOT, M. *Demography — measuring and modeling population processes*. Blackwell, 2001.
- RAUDENBUSH, S. W. *et al. HLM 5: hierarchical linear and nonlinear modelling*. Lincolnwood: Scientific Software International, 2000.
- RIOS-NETO, E. L. G. *Análise da progressão de matrículas e população — o método probabilidade de progressão por série*. Versão preliminar apresentada no Workshop de Demografia da Educação. Associação Brasileira de Estudos Populacionais (Abep), Salvador (BA), jun. 2001.
- SILVA, N. do V., HASENBALG, C. *Recursos familiares e transições educacionais*. Versão preliminar apresentada no Workshop de Demografia da Educação. Associação Brasileira de Estudos Populacionais (Abep), Salvador (BA), jun. 2001.

(Originais recebidos em julho de 2002. Revistos em dezembro de 2002.)

