

# TRANSMISSÃO INTERGERACIONAL DE DESIGUALDADE E QUALIDADE ESCOLAR: uma análise multinível para as escolas paranaenses\*

Marco Túlio Aniceto França \*\*  
Fernanda Marie Yonamini \*\*\*  
Flavio de Oliveira Gonçalves \*\*\*\*

## RESUMO

Neste artigo verificam-se os determinantes da qualidade escolar no Paraná considerando-se as características familiares, escolares e do município. Estimamos a complementaridade entre escolas e famílias e verificamos que no ensino fundamental não existem características escolares que reduzam essa complementaridade. Por outro lado, no ensino médio as escolas com nível socioeconômico médio mais alto têm maior complementaridade. O volume de recursos dos municípios destinado ao ensino fundamental é importante para a constituição de um sistema educacional de melhor qualidade. Mostramos também que diferenças de desempenho entre escolas públicas e privadas conduzem os estudantes a uma desigualdade de condições, gerando a perpetuação da desigualdade.

**Palavras-chave:** modelos multinível; educação; desigualdade.

## ABSTRACT

In this paper we estimate the determinants of education quality in the State of Paraná by controlling family, school, and geographic characteristics. We estimate complementarities between families and schools and test educational characteristics which promotes equality and efficiency. Based on data from SAEB (Evaluation System of Basic Education) for the year 2003 a hierarchical linear model is applied. We show that municipal resources dedicated to education are an important factor in explaining a higher quality of regional educational system. We also show that differences in performance of educational system lead students to face unequal opportunities, which perpetuate intergenerational inequalities.

**Key words:** multilevel models; education; inequality.

\*Este trabalho foi desenvolvido no âmbito do NAPPE - UFPR (Núcleo de Avaliação de Políticas Públicas Educacionais). Os autores agradecem à CAPES pelo financiamento à pesquisa.

\*\*Economista, doutorando do Programa de Pós-Graduação em Desenvolvimento Econômico da UFPR. E-mail: tulio\_franca@yahoo.com

\*\*\*Economista, doutorando do Programa de Pós-Graduação em Desenvolvimento Econômico da UFPR. E-mail: yonamini@yahoo.com.br

\*\*\*\*Economista, doutor em Teoria Econômica pela Universidade de Brasília (UnB). Professor adjunto do Programa de Pós-Graduação em Desenvolvimento Econômico da UFPR. E-mail: flaviogoncalves@hotmail.com

Artigo recebido para publicação em novembro/2006. Aceito para publicação em agosto/2007.

## INTRODUÇÃO

A educação é um ativo importante, uma vez que, aliada a outros mecanismos, contribui para o crescimento econômico e o combate à desigualdade. As diferenças regionais de renda são explicadas, em grande parte, pelo nível de escolaridade médio da mão-de-obra (BARROS et al., 2001). Além da quantidade média de anos de estudo de uma região, a qualidade na formação do capital humano<sup>1</sup> é determinante da capacidade de aprendizado dos trabalhadores.

No ambiente familiar, os pais são os responsáveis, na maioria das vezes, pela tomada de decisão em relação a educar as gerações futuras. E a quantidade de educação a ser adquirida mostra-se correlacionada com o nível de educação dos pais. Os pais que detêm um elevado nível de capital humano terão filhos com nível educacional semelhante ou até mesmo superior. Nesse arcabouço, as ações desenvolvidas pelos pais servem de referência para as ações dos seus filhos, bem como as ações oriundas de outros grupos, diferentes da família, e que também exercem influência sobre o comportamento da criança: o grupo de amigos, a igreja, o time de futebol etc. Ademais, existem outros canais pelos quais essas interações ocorrem, isto é, via disponibilidade de renda/riqueza para financiar a educação dos filhos. Configuram-se assim, parcialmente, as complementaridades existentes entre o capital humano da família e a educação (COLEMAN, 1988). Contudo, em segmentos pobres esses mecanismos são frágeis para a geração de indivíduos mais educados. O resultado é a perpetuação da desigualdade.

Em países como o Brasil, com alto nível de desigualdade, a educação não tem se mostrado como instrumento de ruptura desse círculo vicioso, embora o sistema educacional tenha passado por muitas reformas que beneficiaram de maneira marginal os pobres. Além disso, o País exibe características duais de um sistema parcialmente privado de qualidade média superior à do sistema público, que são mantidas desde a sua constituição.

A partir do final dos anos 1980, verificaram-se esforços para a ampliação do número de matrículas nos ensinos fundamental e médio, junto com o combate ao analfabetismo, como metas de política governamental. No entanto, o aumento no número de matrículas não veio acompanhado da manutenção ou aumento na qualidade educacional.

Este quadro foi confirmado após a realização sistemática de testes educacionais (PISA<sup>2</sup>/OCDE<sup>3</sup> e SAEB<sup>4</sup>) que revelaram uma fraca formação<sup>5</sup> dos estudantes de ensino

---

<sup>1</sup> Capital humano é um ativo intangível que representa um conjunto de conhecimentos, habilidades e cuidados com a saúde que são incorporados de forma produtiva no trabalho pelo ser humano e cujos reflexos são observados nos salários. Embora o conceito de capital humano inclua, atualmente, variáveis de educação, saúde e fecundidade, o foco deste trabalho foi apenas sobre o primeiro.

<sup>2</sup> *Programme for International of Student Assessment (PISA)*, adotado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE), é um programa de avaliação que busca mensurar a qualidade na formação de estudantes com 15 anos de idade, ou seja, no término do ensino fundamental.

<sup>3</sup> Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Mais informações no site: <[www.oecd.org](http://www.oecd.org)>

<sup>4</sup> Sistema Nacional de Avaliação da Educação Básica (SAEB): é um programa de avaliação executado pelo governo federal em parceria com o MEC/INEP desde 1990. Procura avaliar as crianças das 4.ª e 8.ª séries do ensino fundamental, e jovens em fase de conclusão do ensino médio.

<sup>5</sup> De acordo com o informe do MEC a respeito do PISA para o ano de 2000, os alunos brasileiros pouco entendem aquilo que lêem. Além disso, o País ocupou o último lugar no ranking de classificação de um total de 32 países avaliados. Para o ano de 2003, o País teve alguns avanços, porém ocupa a penúltima posição de um total de 54 países. Mais informações no sítio <[www.inep.gov.br](http://www.inep.gov.br)>

fundamental e médio. Além disso, estes testes mostraram diferenças profundas de qualidade entre os sistemas público e privado de educação, assim como intra-escola. Em geral, as escolas da rede pública apresentaram os níveis mais baixos de desempenho, com algumas exceções. O agravante é que a grande maioria das crianças e jovens que estudam nessas unidades escolares é oriunda de famílias pobres.

Trata-se de um quadro bastante perverso, dado que as complementaridades provenientes da família são precárias, resultando, juntamente com a menor qualidade da escola, na baixa qualidade da formação estudantil. A criança compete em desigualdade de condições, já que a escola pública não complementa a defasagem do capital humano da família. Portanto, a reversão desse quadro passa por investimentos que visem a aumentar a qualidade dessa escola. Ao reduzir as influências relacionadas às complementaridades provenientes da família, o círculo vicioso que alimenta a desigualdade poderia ser rompido. Nosso estudo busca verificar a hipótese de redução da complementaridade entre desempenho escolar e nível socioeconômico das famílias no desempenho dos alunos. Definimos como uma boa escola aquela que é igualitária, onde o desempenho do aluno torna-se independente da condição socioeconômica de seus pais.

Para a realização deste estudo utilizou-se o banco de dados do SAEB referente ao Estado do Paraná de 2003. O Estado apresenta um desempenho superior à média brasileira, ficando atrás apenas do Rio Grande do Sul. Em relação à desigualdade de renda, mensurada pelo índice de Gini<sup>6</sup>, o Paraná possui um dos menores indicadores, de 0,539, para o ano de 2004, segundo informações do IBGE (Instituto Brasileiro de Geografia e Estatística) em comparação com o índice referente à desigualdade brasileira, de 0,569. Entretanto, é um índice ainda bastante elevado para padrões mundiais, configurando, assim, o problema da desigualdade de renda no Estado. Todavia, ressaltamos a afirmação de Dias (2000) de que este tipo de desigualdade não é totalmente explicado pela escolaridade, apesar da existência de uma relação estreita entre ambas. Contudo, as diferenças na qualidade de ensino entre as escolas podem corroborar na manutenção de um quadro desigual em razão da desigualdade de condições.

Assim, empregamos na estimação a metodologia hierárquica ou multinível de regressão. A técnica foi utilizada em razão de a metodologia para a construção da amostra levar em conta pesos e conglomerados, além da semelhança nas observações oriundas de escolas pertencentes à mesma região dentro da mesma dependência administrativa. É importante ressaltar que são grandes as semelhanças dos estudantes pertencentes a uma mesma escola. Nesse sentido, a metodologia permite o exame do desempenho do aluno isolando aspectos referentes às instituições de ensino e à família.

Procura-se, desse modo, investigar as características escolares que seriam promotoras de bom desempenho no SAEB e examinar se as escolas do Paraná têm características que contribuem para a maior equidade, ou seja, que reduzem a complementaridade da família no desempenho do estudante. Caso esta hipótese seja

---

<sup>6</sup> Índice de Gini: é uma medida de desigualdade utilizada, em geral, para mensurar a desigualdade de renda. O índice consiste em um número cujo intervalo vai de 0 (igualdade perfeita) a 1 (desigualdade perfeita). O Brasil possui um índice de Gini de 0,569, ocupando a 117.ª posição de um total de 124 países, ficando à frente de nações como Namíbia, Lesoto e Serra Leoa. Mais informações no site: <[www.ibge.gov.br](http://www.ibge.gov.br)>

confirmada, a escola paranaense estaria provendo uma igualdade de condições aos seus estudantes e, dessa forma, agindo na redução da desigualdade através das gerações.

O artigo está dividido em cinco seções, além desta introdução. Na seção seguinte, descreve-se o arcabouço teórico que fundamenta este trabalho. A terceira parte é dedicada à metodologia multinível ou HLM (*Hierarchical Linear Model*), em que se expõem os motivos que conduzem a esta escolha *vis-à-vis* as abordagens paramétrica tradicional e não-paramétrica. A seção quatro destina-se à apresentação da base de dados utilizada, o SAEB, e a sua forma de elaboração, que justificam a adoção da metodologia multinível. Em seguida discutem-se as hipóteses do trabalho e apresentam-se os resultados. E, finalmente, têm-se as considerações finais.

## 1 REFERENCIAL TEÓRICO

A literatura reconhece que há uma forte relação entre características familiares e qualidade da formação estudantil. Se considerarmos uma função de produção educacional, duas variáveis estão sob o controle da família: o volume de investimentos e o tempo destinado à educação. Segundo Glomm e Ravikumar (1992), estas variáveis têm como *proxy* o nível de escolaridade dos pais, já que este é um importante determinante para a renda, além de indicador dos anos de escolaridade formal a serem atingidos pela criança. Nesse cenário, pode-se observar uma relação circular, já que filhos de pais com menor nível educacional estão propensos a possuir uma quantidade de capital humano semelhante à dos seus pais, mantendo uma distribuição desigual de riqueza na forma de capital humano ao longo das gerações. O quadro pode ser contornado se houver acesso ao ensino público de qualidade ou redução nas restrições de crédito por parte das famílias com menores níveis de capital humano.

Contudo, do ponto de vista institucionalista, Engerman e Sokoloff (2002) afirmam que o acesso ao ensino por parte dos segmentos menos abastados ocorreu em velocidades distintas entre os países. Isto porque o grau de desigualdade na distribuição inicial da riqueza e, conseqüentemente, na velocidade de ampliação dos mecanismos de sufrágio refletiu-se na constituição e falta de acesso ao sistema educacional por parte dos menos abastados.

Segundo esses autores, nos países que apresentaram uma estrutura fundiária concentrada nas mãos de poucos, as instituições de sufrágio desenvolveram-se em um ritmo mais lento.<sup>7</sup> Nesse cenário, as primeiras escolas de ensino fundamental e médio eram financiadas de forma privada. As escolas públicas eram praticamente inexistentes, ou seja, os mais pobres não tinham acesso ao ensino básico. Isto explica, em parte, o porquê de o Brasil permanecer com índices elevados de analfabetismo na população adulta até os anos recentes (segundo dados do IBGE, 20,1% da população adulta permanecia analfabeta no início dos anos 1990).

Por outro lado, nos países cuja estrutura fundiária mostrou-se menos concentrada, as restrições para o exercício do voto foram suspensas mais rapidamente. Logo, adotaram-se formas mais universais de financiamento e acesso ao sistema educacional.

---

<sup>7</sup> As restrições em torno dos direitos de sufrágio, riqueza e/ou escolaridade permaneceram até o ano de 1988 no Brasil (NICOLAU, 2002). De acordo com Engerman e Sokoloff (2002), em países como Estados Unidos e Canadá, estas restrições foram retiradas a partir do final do século XIX, ou seja, aproximadamente 100 anos antes.

Dentro desse mesmo arcabouço, Glomm e Ravikumar (1992) mostram que os determinantes para a educação das gerações futuras podem ser decompostos em dois: as características familiares e a qualidade da escola. Os autores ressaltam que a qualidade da educação é o principal legado que os pais deixam para os seus sucessores. A combinação tem reflexo no crescimento econômico e na forma pela qual a desigualdade evolui no decorrer dos anos.

Para isso, eles utilizam um modelo de gerações sobrepostas cujas comunidades são constituídas por agentes heterogêneos. Esta característica influencia na estrutura de financiamento adotada (pública ou privada), por ser uma decisão da maioria e, dessa forma, ser um indicador para o grau de desigualdade existente. O impacto de tal decisão recai sobre a qualidade da escola, uma vez que no sistema público esta seria decorrente do volume de impostos arrecadados da população pelo Estado. No âmbito privado, por sua vez, fica estritamente a cargo da família o volume de recursos destinados para a formação dos mais jovens ou para consumo próprio. As escolas não apresentam diferenças de qualidade entre si. Por último, a heterogeneidade no capital humano dos pais torna a sua transmissão específica a cada família.

Glomm e Ravikumar (1992) destacam que, sob um regime privado de financiamento, os pais com elevado capital humano e renda destinam um volume maior de recursos para a educação dos filhos. O maior investimento em educação se reflete em um alto nível de renda no futuro por parte dessas crianças e, em última instância, em acréscimos de qualidade na escola.<sup>8</sup> Os resultados mostram que a economia cresce mais rapidamente sob esta situação, se a desigualdade não for excessiva. Por outro lado, o regime público mostra-se mais igualitário, uma vez que a influência dos pais recai apenas sobre a qualidade da escola. Além disso, a adoção desse sistema em uma economia com graus elevados de desigualdade possibilita um maior crescimento, quando comparado ao sistema privado.

Para Benabou (1996), o sistema público de financiamento educacional favorece, em geral, as camadas mais baixas da população. Contudo, sua adoção é mais comum em sociedades cujos níveis de desigualdade não são elevados, uma vez que os segmentos mais abastados arcam com os custos educacionais de seus descendentes. Desta forma, os abastados podem mostrar-se pouco interessados em financiar uma educação para todos. Logo, um sistema privado de financiamento é escolhido, e um equilíbrio com graus elevados de desigualdade emergiria.

Para Galor e Zeira (1993), as diferenças no volume de recursos são um dos condicionantes para que algumas famílias fiquem impedidas de realizar investimentos educacionais. Isto é demonstrado utilizando-se instrumento semelhante ao de Glomm e Ravikumar (1992), com a existência de altruísmo entre as gerações. Porém, a distribuição inicial da riqueza é que determina os investimentos futuros em capital humano.

---

<sup>8</sup> Os autores ressaltam que a qualidade da educação ofertada pelas escolas está diretamente ligada ao volume de recursos destinados pela família. O regime público de financiamento à qualidade da escola é financiado pelos impostos arrecadados da população. No regime privado está diretamente ligado aos recursos ofertados pela família. Por fim, é importante destacar que não há a coexistência dos regimes.

No primeiro período, o nível inicial de riqueza determina se o indivíduo decide investir em educação ou se ingressa diretamente no mercado de trabalho. No período seguinte, ambos os trabalhadores (educados ou não) consomem e deixam um legado. Este altruísmo determina o volume de educação adquirido pelas gerações seguintes. Assim, Galor e Zeira (1993) afirmam que os segmentos ricos sempre destinam recursos para a educação dos filhos, enquanto as camadas menos abastadas estariam presas em uma armadilha de pobreza. A situação pode ser contornada caso o mercado de crédito impeça esta distorção.

Contudo, as elevadas taxas de juros impostas àqueles que necessitam de financiamento tornam o empréstimo impraticável devido aos problemas de risco moral, e isto seria perpetuado para todas as gerações (GALOR; ZEIRA, 1993). Este quadro reforça a afirmação de Barros, Henriques e Mendonça (2002), uma vez que o crédito não é acessível às camadas de baixa renda para o financiamento de recursos educacionais.

Assim, as disparidades na distribuição da riqueza mantêm-se no longo prazo com impactos sobre as taxas de crescimento da economia. As sociedades com níveis elevados de desigualdade teriam um menor crescimento no longo prazo. O contrário aconteceria com aquelas que tivessem um perfil distributivo mais homogêneo, as quais, além de apresentarem uma renda menos concentrada, mostrariam um maior desempenho da economia no longo prazo.

O que os modelos descritos até aqui têm em comum é o fato de que mecanismos geradores da desigualdade são reproduzidos no sistema educacional com reflexos sob a qualidade. Dessa forma, investigar-se-ão alguns determinantes da qualidade nas escolas paranaenses e se esses determinantes são capazes de romper o círculo vicioso da desigualdade, bem como se as restrições de crédito à educação são um condicionante para o não-acesso a uma educação de qualidade.

Na seção seguinte, descreve-se a metodologia multinível de regressão. Embora as metodologias paramétrica tradicional e não-paramétrica sejam comumente utilizadas nesta abordagem, inicialmente serão destacadas algumas restrições a que essas metodologias estão sujeitas, relacionadas à construção da amostra e à não consideração de uma estrutura hierárquica nas relações existentes entre as características escolares.

## 2 OS MODELOS LINEARES HIERÁRQUICOS

As metodologias paramétrica tradicional e não-paramétrica exigem que a amostra seja construída de maneira aleatória, ou seja, que as observações sejam independentes umas das outras. Contudo, esta hipótese não é observada em amostras construídas através de estratos da população, como o SAEB. Os estudantes que pertencem a uma mesma classe, escola ou área geográfica demonstram mais similaridades entre si do que quando comparados a estudantes que estão em um contexto diverso de outras áreas (HOX; KREFT, 1994, p.284). Com isso, Kish (1987) mostra que a realização de estimativas que não consideram tais peculiaridades pode conduzir a estimativas muito pequenas dos desvios, produzindo resultados "significativos" espúrios, chamados de efeito *cluster*.

Além disso, é importante considerar a estrutura hierárquica existente nas características escolares. De acordo com Hox (1995), a escola apresenta uma estrutura

hierárquica, podendo ser estabelecidos até quatro níveis de agregação para a estimação dos parâmetros. A forma de agrupamento ocorre da seguinte forma: os alunos estão no nível mais inferior, sendo o primeiro nível de análise. Estes estariam organizados em turmas, que seriam o segundo nível. As turmas estão estabelecidas em escolas e, por conseguinte, as escolas encontram-se localizadas em bairros ou comunidades, que seriam, assim, o terceiro e o quarto nível de agregação, respectivamente.

A não-observância da estrutura hierárquica das relações pode conduzir o analista a inferir conclusões que não se referem ao nível estimado, denominado de *ecological fallacy*.<sup>9</sup> A verificação da estrutura hierárquica das observações corrobora a redução de erros de identificação. Esta redução ocorre devido ao fato de a técnica hierárquica buscar estimar os efeitos diretos e as interações existentes segundo os diferentes níveis de hierarquia.

A metodologia hierárquica de regressão difere de outros métodos ao permitir que o intercepto e o coeficiente de inclinação sejam os regredidos em um segundo nível de estimação.

O trabalho de Bryk e Raudenbush (1986) foi um dos primeiros a ressaltar a estrutura hierárquica das características escolares, tendo desenvolvido, então, a metodologia HLM (*Hierarchical Linear Models*) para analisar esta forma de amostra.

Eles estimaram a relação existente entre o nível socioeconômico e a proficiência do aluno. O objetivo era testar, por meio desta modelagem, as conclusões encontradas por Coleman et al. (1982). Enquanto estes autores realizaram inferências mediante equações paramétricas de regressão cujos alunos estavam divididos segundo a rede, Bryk e Raudenbush (1986) estimaram simultaneamente as regressões com o cuidado de verificar a qual nível pertencia cada variável. O resultado de Coleman et al. (1982) é que as escolas católicas seriam mais igualitárias em detrimento das escolas públicas. Ou seja, o nível socioeconômico teria menor impacto no desempenho do estudante. O método de Bryk e Raudenbush (1986) substitui  $n$  estimações do intercepto e as que estão relacionadas ao termo erro, pela estimação de um único intercepto e da variância dos resíduos entre grupos. As conclusões de Bryk e Raudenbush (1986) mostraram que a relação entre proficiência e nível socioeconômico varia de escola para escola.

O modelo desenvolvido por Bryk e Raudenbush (1986) consiste de  $n$ -equações e  $n$ -estágios. A primeira equação é a regressão para cada escola,  $j$ , do nível de proficiência ( $y_{ij}$ ) com todas as variáveis independentes no nível do aluno. Abaixo elas estão representadas para uma quantidade  $k$  ( $k = 0, \dots, K$ ) de regressores:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{1jk} + \beta_{2j}X_{2jk} + \dots + \beta_{k-1j}X_{ijK} + r_{ij} \quad (1)$$

onde  $X_{jk}$  refere-se à  $K$ -ésima característica dos estudantes e o coeficiente  $b_{jk}$  corresponde às escolas. O subscrito  $i = 1, \dots, n$  refere-se aos alunos. As escolas são representadas por  $j = 1, \dots, J$ .

<sup>9</sup> Uma completa taxonomia dessas falácias pode ser encontrada em Alker (1969). Outra falácia encontrada sobre dados agrupados e não analisada de forma apropriada é o Paradoxo de Simpson, isto é, trabalhar com dados agrupados e estimá-los como se estes viessem de uma população homogênea.

No primeiro estágio, o processo de estimação é semelhante ao método de mínimos quadrados ordinários (MQO). Segundo Bryk e Raudenbush (1986), a diferença em relação ao método de MQO é percebida na segunda etapa da regressão. Avalia-se se o intercepto ( $\beta_{0j}$ ) e a inclinação ( $\beta_{1j}$ ) variam entre as escolas. Estes são considerados parâmetros aleatórios mensurados através da variância amostral e é por intermédio dela que as inter-relações entre os níveis (aluno e escolas) são verificadas (BRYK; RAUDENBUSH, 1992). Assim, na segunda fase da estimação são regredidos os parâmetros estimados em função das características escolares:

$$\beta_{jk} = \gamma_{0k} + \gamma_{1k}Z_{1j} + \gamma_{2k}Z_{2j} + \dots + \gamma_{p-1,k}Z_{p-1,j} + u_{jk} \quad (2)$$

o subscrito  $p = 0, \dots, P$  denota o número de regressores pertencentes ao segundo nível.  $Z_{pj}$  são as variáveis que captam as características da escola  $j$ .  $\gamma_{pk}$  são os parâmetros que captam os efeitos de interação entre as variáveis escolares e o ambiente intra-escolar ( $\beta_{jk}$ ).

Em geral, centram-se os regressores no primeiro nível em torno de sua média. Dessa forma, o intercepto da regressão intra-escolar passa a refletir a proficiência média dos alunos para cada escola.

A modelagem hierárquica tem como hipóteses: a homocedasticidade de variância no nível 1; a independência e normalidade dos resíduos entre os níveis ( $r_{ij}$  e  $u_{jk}$ ).<sup>10</sup>

No segundo nível, as características acima são mantidas e assume-se independência entre os resíduos intraníveis, ou seja,  $cov(u_{jk}, u_{jk'}) = \tau_{jj}$ .

Ao empregar esta modelagem, os coeficientes de regressão que mensuram as características escolares mostram-se estatisticamente significativos.

Para Hwang (2002), a metodologia multinível é preferível às técnicas ordinárias de regressão se o coeficiente de correlação intraclasse ( $\rho$ ) mostrar-se alto. Esta estatística mede o grau de homogeneidade nas observações e as diferenças no desempenho entre as escolas.

Este fato viola o pressuposto de independência nas observações. Além disso, esta estatística é utilizada para mensurar a proporção da variância total ( $\tau_{00} + \sigma_{R0}^2$ ) que é explicada pela variância do erro no segundo nível. Logo, é a proporção da variância total que é explicada única e exclusivamente pelas características escolares, ( $\tau_{00}$ ).

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma_{R0}^2} \quad (3)$$

Hwang (2002) afirma que quanto mais elevado  $\rho$ <sup>11</sup> maior a homogeneidade entre as observações de um mesmo nível. Em outras palavras: os parâmetros fixos não

<sup>10</sup>  $r_{ij} \sim N(0, \sigma_{R0}^2)$  e  $u_{jk} \sim N(0, \tau_{00})$   
 $cov(r_{ij}, u_{jk}) = 0, \forall i, j, k$

<sup>11</sup> Em simulações de Monte Carlo, Maas e Hox (2004) mostram a importância de se utilizar a metodologia multinível a partir de um coeficiente de 0,1 para a estatística  $r$  e consideram acima de 0,3 um valor alto.

podem ser considerados similares entre os agrupamentos (turmas ou escolas). Neste caso, é relevante permitir que as variáveis explicativas do segundo nível (as características escolares) sejam aleatórias. A não consideração deste tipo de agrupamento e, portanto, a realização da estimação com métodos ordinários de regressão conduzem a estimativas viesadas das variâncias.

Ainda segundo Hwang (2002), se o valor de  $\rho$  é pequeno as técnicas multiníveis não seriam necessárias, uma vez que as observações mostrar-se-iam suficientemente independentes, o que viabilizaria a utilização de técnicas de regressão múltipla.

No entanto, a utilização da variância revela dificuldades de estimação, já que a mesma é menos confiável em comparação às técnicas que empregam a média amostral. Quando a amostra de segundo nível apresenta um número pequeno de observações, isto se reflete no aumento do erro amostral, provocando, em último caso, a redução da confiabilidade dessas regressões. Vale destacar que a complexidade da estimação é aumentada quando a mesma amostra não possui quantidades semelhantes de observações, isto é, quando o número de alunos avaliados por escolas é diferente. Dessa forma, faz-se necessária a utilização de métodos iterativos de máxima verossimilhança para a obtenção de estimativas eficientes.

A especificação do modelo segue a metodologia *bottom-up*, ou seja, parte do menor número possível de regressores, e a inclusão de novas variáveis no modelo ocorre pela observação da estatística de *deviance*<sup>12</sup>. A interpretação da *deviance* é o contrário daquela do coeficiente de determinação, isto é, um menor valor corresponde a um maior ajustamento dos dados. A metodologia utilizada para especificar o modelo busca, portanto, minimizar a estatística proposta.

A metodologia proposta por Hendry (1995) – comumente empregada em séries temporais – não é utilizada em virtude das dificuldades de definição, por parte do analista, de quais variáveis serão aleatórias no nível 1, pois, quanto maior o número de parâmetros aleatórios definidos neste nível, mais elevada é a quantidade de informações necessárias para que os parâmetros referentes à matriz de variância-covariância sejam estimados.

Na seção seguinte explicaremos a forma como a base de dados utilizada foi elaborada. Além de os dados educacionais apresentarem uma estrutura hierárquica, o banco de dados utilizado, o SAEB (Sistema Nacional de Avaliação da Educação Básica), possui características importantes que justificam a modelagem hierárquica em nosso trabalho. Ademais, serão abordadas as variáveis utilizadas, assim como as hipóteses para o nosso estudo.

### 3 O SAEB E A ABORDAGEM HIERÁRQUICA

Os dados do SAEB utilizados neste trabalho são do ano de 2003. O SAEB é um exame que avalia o desempenho dos alunos em algumas séries consideradas chaves e é realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), agência de pesquisa subordinada ao Ministério da Educação. Apesar de o SAEB não ser satisfatório

<sup>12</sup> Especificamente a estatística de *deviance* refere-se a duas vezes a diferença entre a máxima verossimilhança atingível e a máxima verossimilhança ajustada. É uma medida que generaliza a soma dos quadrados dos resíduos de modelos lineares.

para comparações internacionais, os seus objetivos e *design* estatístico, além dos procedimentos empregados na aplicação do teste, não o tornam muito diferente de outras experiências bem conhecidas de avaliação do desempenho dos alunos entre países, como PISA, TIMSS/PIRLS e LLECE.<sup>13</sup>

O SAEB é um teste aplicado desde 1990 em nível nacional que avalia as habilidades cognitivas dos alunos nas disciplinas de Português e Matemática. Os testes são elaborados segundo a metodologia das matrizes de referência que possibilitam aos alunos responderem a diferentes testes, os quais possuem, porém, itens comuns. O desempenho do estudante é mensurado de acordo com a teoria de resposta ao item, comumente chamada de TRI. Esta variável é o indicador das habilidades e competências dos estudantes, ou seja, o indicador de qualidade na sua formação. São avaliadas a quarta e a oitava séries do ensino fundamental e o terceiro ano do ensino médio de escolas públicas e privadas.

Juntamente com os resultados do teste são incorporadas as características familiares desses alunos e informações relevantes a respeito dos professores, diretores e escolas. O banco de dados consiste de dados transversais em seqüência (porém sem formar um painel) de uma amostra representativa das escolas e estudantes. A característica hierárquica da amostra é inata à forma pela qual o banco de dados é construído. Primeiramente, as escolas que fazem parte do SAEB são aleatoriamente escolhidas. Posteriormente, no interior de cada escola, uma ou duas classes são selecionadas. Todos os estudantes da classe que foi selecionada são submetidos ao exame, mas somente em um dos assuntos.

A proficiência dos alunos corresponde a escalas específicas ao assunto elaboradas pelo *staff* do INEP juntamente com professores, pesquisadores e especialistas em *surveys* nacionais e internacionais. Os resultados variam de 0 a 425 e avaliam as habilidades e conhecimentos dos alunos.

A escala dos resultados no SAEB é contínua e comparativa. Isto significa que o aluno cuja proficiência é 400 no exame de Português incorpora todas as habilidades de escrita e leitura possuídas pelos estudantes cujos resultados no exame foram de 150, 300 ou 380, além de algumas habilidades adicionais. Assim, o estudante seria capaz de entender e interpretar textos mais complexos, enquanto outros com níveis inferiores de proficiência não teriam esta capacidade. Devido à característica de invariância na escala do teste, os resultados podem ser comparados entre anos e séries analisadas. Contudo, não é possível construir uma variável que capte o diferencial de rendimento nos exames de uma série para outra (por exemplo, diferenças na proficiência entre 4.º e 5.º séries).

O SAEB poderia revelar um viés de seleção se não conseguisse reproduzir com exatidão as características da população. Desta forma, é um sistema de avaliação que utiliza estratos na construção da amostra, e depois a pondera. Estas características corroboram o uso de uma metodologia multinível de forma a evitar viés no estimador da variância.

---

<sup>13</sup> TIMSS: *Trends in International Mathematics and Science Study*; PIRLS: *Progress in International Reading Literacy Study*. Mais informações no site <http://timss.bc.edu/>.

LLECE: Laboratório Latino-Americano de Qualidade da Educação (mais informações no site <http://llece.unesco.cl/index.act>).

Construímos três variáveis: 1) NSE para o nível socioeconômico da família dos estudantes; 2) EQUI\_ESC para a infra-estrutura de equipamentos disponível na escola; 3) SEG\_PRED para captar as condições de segurança do edifício e das redondezas. A primeira (NSE) foi construída por análise fatorial<sup>14</sup>, técnica que procura identificar as variáveis ou fatores subjacentes que explicam os padrões de correlação dentro de um conjunto de variáveis observadas. Variáveis como número de televisores, rádios, videocassetes, carros, quartos para dormir, contratação de diarista, número de livros em casa e grau de escolaridade dos pais foram reduzidas a um único fator por meio da extração do componente principal no intuito de calcular o nível socioeconômico da família.<sup>15</sup>

A variável EQUI\_ESC é um índice calculado a partir das respostas ao questionário do diretor, que identifica as condições de uso de equipamentos, como computadores para uso dos alunos, vídeos e televisores para uso pedagógico, projetores, acesso a jornais e revistas, laboratórios de ciência, salas de músicas, entre outros.

Procedimento semelhante foi adotado para a variável relacionada à infra-estrutura de segurança como SEG\_PRED, com informações extraídas do questionário que se refere às condições da escola. Foram adicionadas, também, duas variáveis que captassem características relacionadas aos professores, tais como escolaridade (medida em anos de estudo) e salário.

A proficiência dos estudantes em matemática foi utilizada como *proxy* para a qualidade educacional. A escolha desta disciplina baseou-se no relatório do PISA 2000, que afirma que esta matéria é predominantemente aprendida na escola, diferentemente do português, que pode ser aprendido primeiramente em casa. Os dados são analisados para a quarta série do ensino fundamental e o terceiro ano do ensino médio para as escolas do Paraná. Nas tabelas A.1 a A.4, em anexo, apresentamos as estatísticas descritivas das variáveis analisadas.

A amostra para as quartas séries continha originalmente 2.737 observações no nível do estudante. Contudo, muitos dados foram eliminados em razão de os estudantes, diretores ou professores não responderem a alguns itens do questionário. Dessa forma, a amostra reduziu-se para 2.626 estudantes, sendo eliminadas informações para 111 estudantes na amostra. O número inicial de escolas avaliadas foi de 175, porém, pelas mesmas razões, eliminaram-se observações para 9 escolas, resultando em 166 escolas avaliadas.

Para as escolas do ensino médio a amostra continha originalmente 1.694 estudantes, perdendo-se 68 observações. Nestas, 69 escolas foram avaliadas e, por motivos semelhantes, 2 foram retiradas da amostra, sendo avaliados 1.622 estudantes no interior de 67 escolas. Em ambos os casos, a exclusão das observações se deu pelo método *listwise* em razão da construção de variáveis como NSE, EQUI\_ESC e SEG\_PRED.<sup>16</sup>

<sup>14</sup> Uma discussão mais extensa sobre a técnica de análise fatorial, assim como os itens dos questionários de nível socioeconômico utilizados na construção das variáveis empregadas neste estudo, encontram-se em anexo.

<sup>15</sup> A amplitude de variação desta e de outras variáveis, bem como os fatores subjacentes a cada uma, são descritos nas tabelas A.1 e/ou A.2, em anexo.

<sup>16</sup> A proposta inicial era a construção de variáveis correspondentes à escola por meio da metodologia de análise fatorial, semelhante àquela empregada na elaboração do nível socioeconômico. Contudo, observou-se que a adoção desta técnica reduziu sobremaneira o número de escolas na amostra.

Assim, na seção seguinte, construiu-se um modelo hierárquico linear para investigar quais características são importantes para a eficácia escolar, ou seja, para a qualidade da escola. Ademais, será observado se boas escolas conseguem reduzir as complementaridades da família sobre o desempenho do estudante, ou seja, se uma boa escola pode atuar como um mecanismo capaz de romper o círculo vicioso da desigualdade entre as gerações. Caso contrário, o sistema educacional agiria na reprodução das disparidades já existentes.

Os resultados encontrados mostram que no ensino fundamental não existem características escolares que reduzam a complementaridade entre família e desempenho do estudante na escola. No ensino médio, as escolas com nível socioeconômico médio mais alto apresentam uma maior complementaridade. O volume de recursos dos municípios destinado ao ensino fundamental é importante para a constituição de um sistema educacional de melhor qualidade. Nesse sentido, acredita-se que um instrumento eficaz de política pública para opor-se a esta realidade ocorra por intermédio da elevação da qualidade da educação pública.

## 4 RESULTADOS

### Séries Iniciais do Ensino Fundamental

Utiliza-se um modelo incondicional cujo regredido é a proficiência do aluno  $profic_{ij}$  e o termo de erro é  $r_{ij}$ . O segundo nível é composto pelo intercepto,  $\beta_{0j}$ , que representa a proficiência média da escola e o termo de erro,  $\varepsilon_{jk}$ , que corresponde às características escolares. A estratégia de regressão é partir de um modelo incondicional para então adicionar novas variáveis à regressão. O modelo é assim estimado:

$$\begin{aligned} PROFIC_{ij} &= \beta_{0j} + r_{ij} \\ \beta_{0j} &= \gamma_{00} + u_{0j} \end{aligned} \tag{4}$$

Os resultados são mostrados na tabela A.5, em anexo.

A estimativa de  $\gamma_{00}$  é 183,53 com um erro-padrão de 3,39. Este valor refere-se à nota média das escolas de quarta série do ensino fundamental no Paraná. Os efeitos aleatórios decompõem-se em duas partes: aluno e escola. Rejeitamos a hipótese nula de insignificância da variância do desempenho no nível da escola com um teste  $\chi^2$  resultando em um p-valor de 0,000. Dessa forma, temos evidências de que existe significativa variabilidade no desempenho entre as escolas.

As estimativas das variâncias corroboraram nossa afirmação ao possibilitarem a estimação do coeficiente de correlação intra-classe,  $\rho$ . O coeficiente estimado  $\rho = 0,36$  sugere que aproximadamente 36% da variância no desempenho é devido às diferenças entre escolas. Valores acima de 10% desta estatística são um indicativo, segundo Hwang (2002), da necessidade de utilização da modelagem multinível em detrimento dos métodos ordinários de regressão (efeito *cluster*).

O modelo seguinte inclui a variável nível socioeconômico no nível do aluno. O objetivo é verificar se o nível socioeconômico da família influencia positivamente o desempenho do estudante, e se esta relação varia entre as escolas. Assim, estimamos o *random-coefficient model*:

$$\begin{aligned} \text{PROFIC}_{ij} &= \beta_{0j} + \beta_{1j} (\text{NSE}_{ij} - \overline{\text{NSE}_j}) + r_{ij} \\ \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \quad (5)$$

Os resultados são mostrados na tabela A.6, em anexo.

As estimativas mostram que, ao isolar o nível socioeconômico do aluno, o desempenho da escola reduz-se marginalmente, 183,52, porém permanece significativa. O sinal positivo do coeficiente  $\beta_{1j}$  mostra que o nível socioeconômico da família influencia positivamente a proficiência do estudante. O aumento em uma unidade no nível socioeconômico eleva em 0,14 desvio padrão no desempenho do estudante dentro da mesma escola. A diferença faria um estudante na mediana ir para o 56º percentil da distribuição.

No que tange aos efeitos aleatórios, o desempenho continua variando entre as escolas após a inclusão do efeito do nível socioeconômico. Porém, o efeito dessa variável sobre o desempenho não varia entre as escolas do ensino fundamental, isto é, as características da família afetam o desempenho com a mesma intensidade, independentemente das escolas.

Dando seqüência à análise, serão incluídas variáveis no nível da escola e do aluno a fim de explicar as características relacionadas ao seu desempenho, bem como realizar uma investigação mais profunda em relação à equidade. Assim, estima-se o modelo condicional.

$$\begin{aligned} \text{PROFIC}_{ij} &= \beta_{0j} + \beta_{1j} \text{SEXMASC}_{ij} + \beta_{2j} \text{NEGRO}_{ij} + \beta_{3j} \text{REP\_UMA}_{ij} \\ &\quad + \beta_{4j} \text{REP\_DUAS}_{ij} + \beta_{5j} (\text{NSE}_{ij} - \overline{\text{NSE}_j}) + r_{ij} \\ \beta_{0j} &= \gamma_{00} + \gamma_{01} \text{LNDESPED}_j - \overline{\text{LNDESPED}} + \gamma_{02} \text{ESTAD}_j + \gamma_{03} \text{PARTIC}_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} \\ \beta_{3j} &= \gamma_{30} \\ \beta_{4j} &= \gamma_{40} \\ \beta_{5j} &= \gamma_{50} \end{aligned} \quad (6)$$

As estimativas dos efeitos fixos mostrados na tabela A.7, em anexo, indicam que todos os parâmetros são estatisticamente significantes a 6%. O desempenho médio das escolas foi 181,97, com um nível de significância de 0,1%. Os resultados mostram que as escolas que pertencem à rede privada e pública estadual obtiveram um desempenho de 47,67 e 17,38 pontos, respectivamente, superior às municipais. No primeiro caso, a frequência a uma escola privada é capaz de levar um aluno na mediana para 88º percentil e, no caso seguinte, leva para o 66º percentil da distribuição. Contudo, as matrículas nas

escolas privadas estão condicionadas à disponibilidade de financiamento por parte das famílias. E as poucas escolas estaduais voltadas para o ensino fundamental localizam-se no interior do Estado ou em comunidades cuja escolaridade é mais elevada.

A variável LNDESPED foi trabalhada no intuito de estimar o peso dos recursos orçamentários destinados pelos municípios para gastos em educação. O resultado dessa estimativa mostrou uma relação positiva e significativa a um nível de 5,7% para explicar o desempenho das escolas do município. O aumento em 1% nesta variável acarreta uma elevação em 0,045 desvio padrão no desempenho das escolas do município. Ao estudar em um município cujos recursos orçamentários são elevados, o estudante na mediana iria para o 52º percentil.

No tocante à distribuição dos alunos por sexo, as estatísticas descritivas apresentam uma quantidade semelhante no número de meninos e meninas que freqüentam as escolas do ensino fundamental. No entanto, os resultados mostram que as crianças do sexo masculino têm um desempenho 7,39 pontos superior às do sexo feminino. Isto pode ser decorrência, segundo estudo da OECD (2001), da maior facilidade dos meninos em disciplinas relacionadas às áreas de exatas, enquanto as meninas possuem maior facilidade com línguas.

A variável independente etnia mostrou que o desempenho dos afro-descendentes é, na média, 15,86 pontos inferior em comparação aos seus colegas de outras etnias dentro da mesma escola, sinalizando uma realidade desigual também do ponto de vista étnico.

Além da origem étnica e de gênero, foram introduzidas variáveis de controle como repetência. Estas, por sua vez, mostraram um forte impacto negativo sobre o desempenho. Os alunos que apresentaram defasagem série-idade em 1 ano devido à repetência tiveram um desempenho médio inferior de 16,45 pontos em comparação àqueles que estão regularmente matriculados. O desempenho é ainda menor quando a repetência ocasiona uma defasagem de 2 anos, já que a proficiência média mostra-se inferior em mais 25,42 pontos.

A respeito das condições relacionadas à equidade, nenhuma variável da escola mostra-se estatisticamente significativa para reduzir a complementaridade do nível socioeconômico da família sobre o desempenho do estudante nas escolas de ensino fundamental.

Os efeitos aleatórios permaneceram evidenciando que o desempenho varia entre escolas com uma significância estatística de 0,1%. Contudo, como foi observado anteriormente, a influência do nível socioeconômico sobre o desempenho não é variável entre as escolas.

Como o efeito do nível socioeconômico é fixo entre escolas, não é possível distinguir, dentre o conjunto de escolas, quais teriam características promotoras de equidade.

Ainda sobre os efeitos aleatórios, estes demonstraram que a inclusão de variáveis de dependência administrativa e recurso orçamentário explica 51,5% da variância observada no modelo incondicional ( $\tau_{00} = 677,51$ ) ao reduzi-la para ( $\tau_{00} = 345,71$ ). Os resultados apresentados anteriormente confirmam as nossas expectativas ao se demonstrar que o nível socioeconômico é um determinante relevante na proficiência dos estudantes que estão nos anos iniciais de formação. Observa-se que as escolas do ensino fundamental no Paraná não apresentam características de redução dessa influência sobre

o desempenho do estudante. O resultado pode mostrar-se preocupante, pois as crianças pertencentes às famílias menos abastadas e que, na maioria dos casos, estudam em escolas públicas, podem não ser capazes de concorrer em igualdade de condições com as crianças que estudam em escolas privadas.

No entanto, de acordo com Coleman (1998), esta é a fase da formação cuja influência da família sobre a criança é mais intensa devido ao fornecimento de recursos financeiros e pedagógicos. Ademais, é o período em que a criança recebe as ferramentas básicas de instrução. No decorrer da formação, outros elementos, além da família, tornam-se importantes, como os grupos de amigos, comunidade e outros. Assim, a incorporação desses novos elementos na formação do capital humano do indivíduo torna-se necessária à investigação das escolas do ensino médio.

### Ensino Médio

Realizamos o mesmo exercício em relação aos estudantes de escolas de ensino médio. O primeiro passo foi estimar o modelo incondicional. Os resultados são mostrados na tabela A.8, em anexo. A estimativa de  $\gamma_{00}$  é 291,39, com um desvio padrão de 4,75. Isto indica que as escolas de ensino médio do Paraná obtiveram um desempenho médio de 291,39 pontos. O valor é superior ao observado para a quarta série do ensino fundamental, pois parte-se do pressuposto de que esses estudantes incorporam os conhecimentos obtidos nas séries anteriores. A seção composta pelos efeitos aleatórios mostra que a estatística  $\chi^2$  para o intercepto é 1.239,57 com 66 graus de liberdade. O  $p$ -valor de (0,000) mostra que a variância no nível da escola é estatisticamente significativa a 0,1% e evidencia que o desempenho varia de escola para escola.

O cálculo do coeficiente de correlação intraclasse,  $\rho$ , para este modelo sugere que existe uma diferença de 36,88% no desempenho entre as escolas. A estatística mostra que as diferenças de desempenho entre as escolas cresceram ligeiramente no decorrer das séries de formação.

No modelo seguinte, *random-coefficients model*, incluímos a variável que mensura a complementaridade da família, de forma a verificar se a mesma varia entre as escolas. Assim, estimamos este modelo com a especificação de  $\beta_{0j}$  e  $\beta_{1j}$  como aleatórios:

$$\begin{aligned} \text{PROFIC}_{ij} &= \beta_{0j} + \beta_{1j} (\text{NSE}_{ij} - \overline{\text{NSE}_j}) + r_{ij} \\ \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{7}$$

Os resultados são apresentados na tabela A.9, em anexo.

As estimativas dos efeitos fixos mostram que ambos os coeficientes, desempenho da escola e nível socioeconômico, são significantes ao nível de 1%. O desempenho médio das escolas foi de 290,13 pontos. A influência do nível socioeconômico, 5,64, mostra-se importante para explicar o desempenho do aluno, uma vez que o aumento em uma unidade nessa variável aumenta em 0,075 desvio padrão no desempenho do estudante dentro da mesma escola. A diferença faria um estudante na mediana ir para o 53º percentil da distribuição.

Os efeitos aleatórios indicam que a variação de desempenho entre as escolas é significativa a 1%. Porém, o efeito do nível socioeconômico sobre o desempenho não é estatisticamente diferente entre as escolas. Esta evidência é semelhante à constatada para o ensino fundamental, isto é, o efeito dessa variável sobre o desempenho é constante entre as escolas.

Para estimar o modelo condicional, incluímos variáveis no nível do aluno e da escola. O objetivo é captar se as diferenças entre as escolas são explicadas pela sua dependência administrativa e quais características são promotoras da eficácia escolar. A variável correspondente aos recursos orçamentários do município não foi incluída na análise em razão de as escolas do ensino médio serem administradas pelo Estado e não pelo município. Nos aspectos relacionados à equidade, serão investigadas quais características escolares podem tornar a escola mais igualitária, isto é, contribuir na redução da complementaridade da família no desempenho do aluno.

A estimação do modelo condicional é especificada a seguir:

$$\begin{aligned}
 \text{PROFIC}_{ij} &= \beta_{0j} + \beta_{1j} \text{SEXMASC}_{ij} + \beta_{2j} \text{REP\_UMA}_{ij} + \beta_{3j} \text{REP\_DUAS}_{ij} \\
 &+ \beta_{4j} \text{REP\_TRES}_{ij} + \beta_{5j} (\text{NSE}_{ij} - \overline{\text{NSE}}_j) + r_{ij} \\
 \beta_{0j} &= \gamma_{00} + \gamma_{01} \text{REDE}_j + \gamma_{02} \text{SLAMED}_j - \overline{\text{SLAMED}} + \gamma_{03} \text{CONSERV\_ESC}_j - \overline{\text{CONSERV\_ESC}} + u_{0j} \\
 \beta_{1j} &= \gamma_{10} \\
 \beta_{2j} &= \gamma_{20} \\
 \beta_{3j} &= \gamma_{30} \\
 \beta_{4j} &= \gamma_{40} \\
 \beta_{5j} &= \gamma_{50} + \gamma_{51} \text{EQUI\_ESC} - \overline{\text{EQUI\_ESC}} + \gamma_{52} \text{NSEMEDIO}_j - \overline{\text{NSEMEDIO}} + \gamma_{53} \text{ESCMPR\_E}_j - \overline{\text{ESCMPR\_E}} + u_{5j}
 \end{aligned} \tag{8}$$

Os resultados são mostrados na tabela A.10, em anexo.

Os efeitos fixos mostram que todas as estimativas são significantes a 5%. Estes resultados possuem implicações interessantes. Primeiramente, todos os coeficientes que explicam  $\beta_{0j}$  (a proficiência média da escola) são positivos, significando que as escolas da rede privada têm um desempenho superior às públicas em 0,8 desvios-padrões do desempenho (gráfico A.1 – em azul, as escolas da rede pública, e em vermelho as particulares). A frequência à escola privada conduz o aluno na mediana para o 79º percentil. A presença de infra-estrutura mostra-se importante para explicar o desempenho, assim como os salários dos professores, embora este último apenas marginalmente. O aumento em uma unidade no quesito conservação do edifício eleva em 0,83 desvio padrão o desempenho da escola em relação à média das escolas. As escolas com boa infra-estrutura levam o estudante da mediana para o 80º percentil. O efeito mostra-se semelhante ao observado para a variável rede.

Na explicação da equidade podemos observar que nenhuma característica escolar contribui na redução da complementaridade da família no desempenho do aluno, mesmo controlando pelo nível socioeconômico médio das escolas. As escolas frequentadas pelas camadas mais ricas (alto nível socioeconômico médio) são menos igualitárias em comparação com aquelas de menor nível socioeconômico médio, ao aumentar em 0,11 desvio padrão para cada unidade de aumento do nível socioeconômico médio. A escolaridade do professor também não contribui para reduzir a influência da família no desempenho escolar do

estudante. Este resultado é semelhante ao encontrado por Albernaz, Ferreira e Franco (2002). Os autores afirmam que a escolaridade do professor, embora aumente a eficácia da escola, tem impacto maior no desempenho daqueles alunos cujo nível socioeconômico é mais elevado.

As características dos indivíduos relacionadas a gênero mostram que os estudantes do sexo masculino têm um desempenho superior aos do sexo feminino. Isto é estatisticamente significativo a um nível de significância acima de 0,1% para as características mencionadas. Os estudantes com defasagem série-idade ocasionada pela repetência têm um desempenho inferior àqueles matriculados regularmente, com significância estatística maior do que 0,1%. Ao contrário das escolas da quarta série de ensino fundamental, as características relacionadas à cor não foram estatisticamente significativas. O resultado pode ser consequência da redução no tamanho da amostra para esses alunos, evidenciando uma evasão para o mercado de trabalho quando os estudantes se encontram no ensino médio.

A proporção da variância do erro do desempenho médio da escola foi reduzida pelas variáveis de infra-estrutura, dependência administrativa e salários em 75%. A variância estimada da proficiência média foi de  $\tau_{00} = 309,004$ , com *p*-valor menor do que 1%. Em relação à estatística  $\hat{\rho}$ , de correlação intraclasse condicional, observa-se uma redução de 36,88% para 14% entre as observações após controlar pelas características escolares. Estes resultados, assim como para o ensino fundamental, indicam que existem, ainda, níveis significantes de heterogeneidade nos dados que precisam ser considerados na análise da eficácia e equidade das escolas. A inclusão de variáveis como infra-estrutura escolar, professor e dependência administrativa capta parte dessa heterogeneidade, embora diferenças significativas permaneçam sem explicação.

## CONSIDERAÇÕES FINAIS

O objetivo deste artigo foi o de evidenciar características individuais e escolares que influenciam na formação de capacidade e conhecimento dos estudantes paranaenses dos ensinos fundamental e médio.

Os resultados mostram que os gastos em educação realizados pelos municípios impactam positivamente sobre o desempenho dos alunos do ensino fundamental. Assim, a consideração da estrutura hierárquica, os problemas relacionados à interpretação que não correspondessem efetivamente àquela hierarquia foram contornados (*ecological fallacy*). As inferências mostraram que o nível socioeconômico da família é importante e que não há variabilidade na sua influência entre as escolas do ensino fundamental.

Os resultados evidenciam que a complementaridade da família age por outros mecanismos que não estão restritos à dotação inicial, pois a decisão de matricular a criança em uma escola privada depende de recursos. Portanto, a transmissão da desigualdade realiza-se por mecanismos indiretos. Vale ressaltar que as escolas da rede estadual mostraram-se melhores, comparativamente às municipais, embora tenhamos controlado pelo volume orçamentário destinado à educação. O resultado sugere que a concentração de recursos nos municípios pode ter reduzido a qualidade do sistema

educacional nas localidades cujos gastos são menores. De acordo com Dias (2000), o Paraná é o estado cujo processo de municipalização mostra-se mais avançado, com 86% das escolas sendo de responsabilidade das prefeituras.

O desempenho de alunos negros e pardos nos exames de proficiência para as escolas de ensino fundamental é muito inferior em comparação ao dos alunos que se autodenominaram brancos dentro da mesma escola. O resultado parece corroborar situações de desigualdade racial presentes em distintas esferas da sociedade brasileira em geral desfavoráveis à população de cor.

Todavia, esse quadro não é observado no ensino médio, uma vez que as estimativas mostraram-se insignificantes. Acreditamos que o resultado pode ser decorrente da evasão escolar neste período de formação, devido à entrada precoce no mercado de trabalho. Esse fato dificulta a mobilidade intergeracional para esses jovens. Ademais, a baixa escolaridade dos seus pais ocasiona menores salários, dificultando o acesso dessas crianças a uma educação de qualidade, já que a complementaridade via financiamento à educação é fundamental.

Verificou-se que no ensino fundamental a complementaridade da família é muito mais forte em comparação à observada no ensino médio. Assim, a educação fornecida por estas escolas não consegue romper o círculo vicioso e, dessa forma, transmite-se a desigualdade para as gerações futuras.

Por fim, a importância do desenvolvimento de políticas públicas que visam à elevação da qualidade da educação pública básica pode se mostrar eficaz no combate à desigualdade de condições. De acordo com o arcabouço apresentado, uma educação de qualidade é usufruída, em geral, por crianças cujas famílias possuem uma boa dotação inicial de recursos. Crianças e jovens oriundos de famílias com baixa dotação de recursos têm acesso a uma educação de baixa qualidade. Ao assumir que são indivíduos que se encontram inviabilizados de competir em igualdade de condições no mercado de trabalho e ter acesso a melhores salários, uma política pública de acréscimo de qualidade da escola pública pode lograr sucesso no rompimento desse círculo vicioso, transmitido de geração em geração.

## REFERÊNCIAS

ALBERNAZ, A.; FERREIRA, F.; FRANCO, C. Qualidade e equidade na educação fundamental brasileira. **Pesquisa e Planejamento Econômico**, Rio de Janeiro: IPEA, v. 32, n. 3, p.453-476, dez. 2002.

ALKER, H. A Typology of ecological fallacies. In: DOGAN, M.; ROKKAN, S. **Social ecology**. Cambridge, Mass., MIT Press, 1969. p. 69-86.

BARROS, R. P.; HENRIQUES, R.; MENDONÇA, R. **Pelo fim das décadas perdidas: educação e desenvolvimento sustentado no Brasil**. Rio de Janeiro: IPEA, 2002. (Texto para discussão, 857).

BARROS, R. P. et al. **Os determinantes da desigualdade educacional no Brasil**. Rio de Janeiro: IPEA, 2001. (Texto para discussão, 834).

- BENABOU, R. Heterogeneity, stratification, and growth: macroeconomic implication of community structure and school finance. **American Economic Review**, v. 86, n. 3, p. 584-609, 1996.
- BRYK, A. S.; RAUDENBUSH, S. A hierarchical model for studying school effects. **Sociology of Education**, n. 1, v. 59. p. 1-17, 1986.
- BRYK, A. S.; RAUDENBUSH, S. **Hierarchical linear models: applications and data analysis methods**. London: Sage Publications, 1992.
- COLEMAN, J. Social capital in the creation of human capital. **American Journal of Sociology**, v. 94, 1988.
- COLEMAN, J. et al. **High school achievement: public, catholic, and other private schools compared**. New York: Basic, 1982.
- CRONBACH, L. J.; WEBB, N. Between and within-class effects in a reported aptitude-by-treatment interaction: reanalysis of a study by G. L. Anderson. **Journal of Education Psychology**, n. 6, p. 717-724, 1975.
- DIAS, M. L. M. S. Mudanças em curso no ensino público do Paraná. **Revista Paranaense de Desenvolvimento**, Curitiba: IPARDES, n. 98, p. 45-65, jan./jun. 2000.
- ENGERMAN, S.; SOKOLOFF, K. **Factor endowments, inequality and paths of development among new world economics**. Cambridge, 2002. (NBER Working Papers Series, 9259). Disponível em: <<http://www.nber.org/papers/w9259>>. Acesso em: 12/05/2006.
- GALOR, O.; ZEIRA, J. Income distribution and macroeconomics. **Review of Economic Studies**, v. 60, p. 35-52, 1993.
- GLOMM, G.; RAVIKUMAR, B. Public versus private investment in human capital: endogenous growth and income inequality. **Journal of Political Economy**, v. 100, n. 4, 1992.
- HANUSHEK, E. The economics of schooling: production and efficiency in public schools, **Journal of Economic Literature**, v. 24, p. 1141-1177, 1986.
- HENDRY, D. **Dynamic econometrics**. New York: Oxford University Press, 1995.
- HOX, J. J. **Applied multilevel analysis**. Amsterdam: TT-Publikaties, 1995. Disponível <<http://www.soziolegie.uni-halle.de/langer/multilevel/books/hox95mla.pdf>>. Acesso em: 03/03/2006.
- HOX, J. J. **Hierarchical models for survey data**. Ann Arbor: Summer Institute in Survey Research Techniques, 2000.
- HOX, J. J.; KREFT, I. G. G. Multilevel analysis methods. **Sociological Methods & Research**, v. 22, n. 3, p. 283-299, 1994.
- HWANG, Dae-Yeop. **A hierarchical linear modeling approach to higher education research: the influences of student and institutional characteristics**. S. I.: Education Research Information Center, 2002. Disponível em: <[http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?\\_nfpb=true&\\_ERICExtSearch\\_SearchValue\\_0=ED466778&ERICExtSearch\\_SearchType\\_0=no&accno=ED466778](http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED466778&ERICExtSearch_SearchType_0=no&accno=ED466778)>. Acesso em: 14/09/2006.
- JOHNSON, R.; WICHERN, D. **Applied multivariate statistical analysis**. 4ª. ed. New Jersey: Prentice Hall, 1998.
- KISH, L. **Statistical design for research**. New York: Wiley, 1987.
- MAAS, C. J. M.; HOX, J. J. Robustness issues in multilevel regression analysis. **Statistica Neerlandica**, v. 58, n. 2, p. 127-137, 2004.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Belo Horizonte: UFMG, 2005.

NICOLAU, J. **A participação eleitoral no Brasil.** S.l: University of Oxford Centre for Brazilian Studies, 2002. (Working papers series, 26). Disponível em: <www.brazil.ox.ac.uk/workingpapers/Nicolau26.pdf>. Acesso em: 14/06/2006.

OECD. **PISA 2000: first results.** Paris, 2001.

## ANEXO

TABELA A.1 - VARIÁVEIS ESTUDANTIS USADAS NA ESTIMAÇÃO DAS ESCOLAS DE QUARTA SÉRIE DO ENSINO FUNDAMENTAL

VARIÁVEL	DESCRIÇÃO	MÉDIA	DESVIO PADRÃO	MÍNIMO	MÁXIMO
PROFIC	Resultados no exame de matemática	183,13	43,05	66,92	369,98
SEXMASC	1 = meninos; 0 = meninas				
REP_UMA	1 = 1 ano de repetência; 0 = outros				
REP_DUAS	1 = 2 anos de repetência; 0 = outros				
NEGRO	1 = afro-descendentes; 0 = outros				
NSE	Nível socioeconômico da família	0,69	1,13	-2,18	3,54

TABELA A.2 - VARIÁVEIS ESCOLARES USADAS NA ESTIMAÇÃO DAS ESCOLAS DE QUARTA SÉRIE DO ENSINO FUNDAMENTAL

VARIÁVEL	DESCRIÇÃO	MÉDIA	DESVIO PADRÃO	MÍNIMO	MÁXIMO
ESTAD	1 = Dependência administrativa do estado, 0 = outros				
PARTIC	1 = setor privado, 0 = outros				
LNDESPED	Logaritmo das despesas do município com educação	17,12	1,66	12,97	19,55

TABELA A.3 - VARIÁVEIS ESTUDANTIS USADAS NA ESTIMAÇÃO DE ENSINO MÉDIO

VARIÁVEL	DESCRIÇÃO	MÉDIA	DESVIO PADRÃO	MÍNIMO	MÁXIMO
PROFIC	Resultados no exame de matemática	292,15	58,25	155,89	457,55
SEXMASC	1 = meninos ; 0 = meninas				
REP_UMA	1 = 1 ano de repetência; 0 = outros				
REP_DUAS	1 = 2 anos de repetência; 0 = outros				
REP_TRES	1 = 3 anos de repetência; 0 = outros				
NSE	Nível socioeconômico da família	0,41	1,03	-2,14	2,70

TABELA A.4 - VARIÁVEIS ESCOLARES USADAS NA ESTIMAÇÃO DE ENSINO MÉDIO

VARIÁVEL	DESCRIÇÃO	MÉDIA	DESVIO PADRÃO	MÍNIMO	MÁXIMO
SLAMED	Salário do professor	2.068,66	1.068,03	240,00	4000,00
ESCMR	Escolaridade média do professor	16,61	1,14	11,00	18,00
EQUI_ESC	Infra-estrutura escolar: equipamentos	0,55	0,15	0,22	1,00
CONSERV_ESC	Infra-estrutura escolar: conservação	0,89	0,16	0,34	1,00
REDE	1 = Setor privado, 0 = outros				
NSEMEDIO	Nível socioeconômico médio	0,71	0,71	-0,62	2,44

TABELA A.5 - MODELO INCONDICIONAL DE VERIFICAÇÃO DA EFICÁCIA ESCOLAR - QUARTA SÉRIE DO ENSINO FUNDAMENTAL

EFEITOS FIXOS	COEFICIENTE	ERRO PADRÃO		p-VALOR
Desempenho da escola, $\beta_{0j}, \gamma_{00}$	183,54	3,39		0,000
EFEITOS ALEATÓRIOS	VARIÂNCIA	G.L.	$\chi^2$	p-VALOR
Média das escolas, $u_{0j}$	671,74	150	1.414,25	0,000
Efeito aluno, $r_{ij}$	1.190,22			

*Deviance = 24.425,98*

TABELA A.6 - RANDOM-COEFFICIENTS MODEL PARA A VERIFICAÇÃO DA EFICÁCIA E EQUIDADE ESCOLAR - QUARTA SÉRIE DO ENSINO FUNDAMENTAL

EFEITOS FIXOS	COEFICIENTE	ERRO PADRÃO		p-VALOR
Desempenho da escola, $\beta_{0j}, \gamma_{00}$	183,52	3,39		0,000
NSE, $\beta_{1j}, \gamma_{10}$	6,41	1,41		0,000
EFEITOS ALEATÓRIOS	VARIÂNCIA	G.L.	$\chi^2$	p-VALOR
Média das escolas, $u_{0j}$	677,51	150	1431,38	0,000
Efeito aluno, $r_{ij}$	1165,11			

*Deviance = 23.827,53*

TABELA A.7 - MODELO CONDICIONAL PARA A VERIFICAÇÃO DA EFICÁCIA E EQUIDADE ESCOLAR - QUARTA SÉRIE DO ENSINO FUNDAMENTAL

EFEITO FIXO	COEFICIENTE	ERRO PADRÃO		p-VALOR
Desempenho da escola, $\beta_{0j}, \gamma_{00}$	180,30	4,67		0,000
LNDSPED $\gamma_{01}$	1,92	1,004		0,057
Estadual $\gamma_{02}$	17,80	5,21		0,001
Particular $\gamma_{03}$	46,70	5,36		0,000
SEXMASC $\beta_{1j}, \gamma_{10}$	6,18	2,05		0,003
Negro, $\beta_{2j}, \gamma_{20}$	-12,29	3,37		0,000
Repet. 1 ano, $\beta_{4j}, \gamma_{40}$	-18,23	3,28		0,000
Repet. 2 anos, $\beta_{5j}, \gamma_{50}$	-27,06	4,29		0,000
NSE $\beta_{6j}, \gamma_{60}$	3,43	1,24		0,006
EFEITOS ALEATÓRIOS	VARIÂNCIA	G.L.	$\chi^2$	p-VALOR
Média das escolas, $u_{0j}$	330,17	144	871,39	0,000
Efeito aluno, $r_{ij}$	1.047,79			

*Deviance = 22.756,102*

TABELA A.8 - MODELO INCONDICIONAL DE VERIFICAÇÃO DA EFICÁCIA ESCOLAR - ENSINO MÉDIO

EFEITOS FIXOS	COEFICIENTE	ERRO PADRÃO		p-VALOR
Desempenho da escola, $\beta_{0j}, \gamma_{00}$	291,40	4,75		0,000
EFEITOS ALEATÓRIOS	VARIÂNCIA	G.L.	$\chi^2$	p-VALOR
Média das escolas, $u_{0j}$	1.239,57	66	1.116,06	0,000
Efeito aluno, $r_{ij}$	2121,96			

Deviance = 17.239,52

TABELA A.9 - RANDOM-COEFFICIENTS MODEL PARA A VERIFICAÇÃO DA EFICÁCIA E EQUIDADE ESCOLAR - ENSINO MÉDIO

EFEITOS FIXOS	COEFICIENTE	ERRO PADRÃO		P-VALOR
Desempenho da escola, $\beta_{0j}, \gamma_{00}$	291,39	4,75		0,000
NSE, $\beta_{1j}, \gamma_{10}$	5,25	1,94		0,007
EFEITOS ALEATÓRIOS	VARIÂNCIA	G.L.	$\chi^2$	P-VALOR
Média das escolas, $u_{0j}$	1.240,15	66	1122,14	0,000
Efeito aluno, $r_{ij}$	2.110,46			

Deviance = 17.228,98

TABELA A.10 - MODELO CONDICIONAL PARA A VERIFICAÇÃO DA EFICÁCIA E EQUIDADE ESCOLAR - ENSINO MÉDIO

EFEITO FIXO	COEFICIENTE	ERRO PADRÃO		P-VALOR
Desempenho da escola, $\beta_{0j}, \gamma_{00}$	291,48	5,39		0,000
Rede, $\gamma_{01}$	46,84	10,20		0,000
Salários, $\gamma_{02}$	0,00802	0,0037		0,036
Conservação. Prédio, $\gamma_{03}$	48,52	16,21		0,004
Masculino, $\beta_{1j}, \gamma_{10}$	16,85	2,36		0,000
Repetência 1 ano, $\beta_{2j}, \gamma_{20}$	-31,41	4,39		0,000
Repetência 2 anos, $\beta_{3j}, \gamma_{30}$	-28,46	5,69		0,000
Repetência 3 anos, $\beta_{4j}, \gamma_{40}$	-34,23	17,03		0,044
NSE, $\beta_{5j}, \gamma_{50}$	4,36	1,42		0,003
Equipamentos, $\gamma_{51}$	16,11	7,02		0,022
Nível socioeconômico médio, $\gamma_{52}$	6,91	2,57		0,008
Escolaridade professor, $\gamma_{53}$	5,74	1,79		0,002
EFEITOS ALEATÓRIOS	VARIÂNCIA	G.L.	$\chi^2$	P-VALOR
Média das escolas, $u_{0j}$	309,004	63	323,13	0,000
Efeito aluno, $r_{ij}$	1885,48			

Deviance = 16.806,31

## ELABORAÇÃO DAS VARIÁVEIS ATRAVÉS DE ANÁLISE FATORIAL

A literatura que aborda as funções de produção educacional verifica que o nível socioeconômico da família é um dos condicionantes que afetam o desempenho escolar, além de ser um indicador dos anos de estudo a serem atingidos pela criança. Entretanto, os questionários que compõem o SAEB não possuem um único item de forma que o nível socioeconômico da família do estudante possa ser extraído, ou seja, não existe uma pergunta que possibilite inferir a renda da família.

O problema se repete para variáveis que visam captar as condições de infraestrutura existentes nas escolas que estão relacionadas à conservação, equipamentos e condições de segurança do edifício. Apesar de as estimativas de Hanushek (1986) não se mostrarem significativas para as variáveis escolares, Albernaz, Ferreira e Franco (2002) acreditam que o resultado encontrado por Hanushek (1986) pode ser decorrente de um tratamento não adequado das informações, uma vez que o autor não considera uma estrutura hierárquica nas relações. Assim, ocorre a subestimação das estimativas em razão da inferência de resultados que não correspondem exatamente à hierarquia estudada (*atomistic fallacy*).

O trabalho de Albernaz, Ferreira e Franco (2002) mostra que as características escolares que se referem às condições das salas de aula, informadas no SAEB 1999, influenciaram positivamente o desempenho do estudante para as quartas e oitavas séries do ensino fundamental. Assim, para o desenvolvimento deste trabalho, decidiu-se incluir novas variáveis escolares na análise para as quartas séries e para os terceiros anos do ensino médio das escolas paranaenses. Acredita-se que existam diferenças de infraestrutura entre as escolas, mesmo ao isolar os efeitos oriundos da dependência administrativa e, dessa forma, com impactos sobre a eficácia e a equidade escolar.

Assim, para a construção das variáveis empregou-se a análise fatorial, que é uma técnica estatística da análise multivariada cujo principal objetivo é a descrição da variabilidade original em termos de um número menor,  $m$ , de variáveis aleatórias, chamadas de fatores comuns e que estão relacionadas com o vetor original  $X$  através de um modelo linear (MINGOTI, 2005). Ou seja, são as associações existentes entre um grande número de variáveis por intermédio da elaboração de um número pequeno de construtos, denominados fatores. A técnica é muito utilizada nas áreas de educação, geologia e marketing.

Os fatores são dimensões latentes que se manifestam de forma redundante em algumas variáveis originais, isto é, as variáveis de uma determinada base de dados são agrupadas em função da correlação existente entre elas. A técnica é baseada em um modelo explícito em que as variáveis da matriz de dados são expressas como uma função linear de um número reduzido de fatores latentes. No caso das variáveis que correspondem ao nível socioeconômico do estudante e das variáveis de infraestrutura escolar (conservação, equipamentos e segurança), os itens de questionários referentes a este construto são expressos de acordo com Johnson e Wichern (1998, p.515):

$$(X - \mu)_{(px1)} = L_{(pxm)} F_{(mx1)} + \varepsilon_{(px1)} \quad (1)$$

onde:

$$(X - \mu)_{px1} = \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix}, \quad \varepsilon_{px1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}, \quad F_{mx1} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}, \quad L_{pxm} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pm} \end{bmatrix}$$

Onde:  $X$  é um item de questionário,  $m$  é o vetor da média da variável,  $F_{mx1}$  é o vetor de fatores comuns ou latentes, que descreve os elementos de uma população em estudo.  $\varepsilon$  é um vetor de erros aleatórios e corresponde aos erros de medida e à variação de  $Z_i$  (variáveis originais do banco de dados), que não é explicada pelos fatores de variação,  $Z_j$ , em que  $j = 1, 2, \dots, m$ .  $l_{ij}$  é denominado de *loading* e é o coeficiente da  $i$ -ésima variável padronizada  $Z_i$  no  $j$ -ésimo fator  $F_j$  e representa o grau de relação entre  $Z_i$  e  $F_j$ .  $L$  é a matriz de *loadings* ou de carga dos fatores, isto é, a correlação entre as variáveis e o fator. Para estimar o modelo considera-se que os fatores latentes formam um conjunto de variáveis padronizadas e não-correlacionadas, isto é:

- $E[F_{mx1}] = 0$ , ou seja, todos os fatores têm média zero;

- $\text{var}[F_{mx1}] = I_{mxm} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \vdots & \vdots \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$ , logo todos os fatores são

não-correlacionados e têm variância igual a 1;

- $E[\varepsilon_{px1}] = 0$ , os resíduos têm média zero;

- $\text{var}[\varepsilon_{pxp}] = \Psi_{pxp} = \begin{bmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & \vdots \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & 0 & \Psi_p \end{bmatrix}$ , ou seja,  $\text{var}[\varepsilon_j] = \Psi_j$  e

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j;$$

- Os vetores  $\varepsilon_{px1}$  e  $F_{mx1}$  são independentes, logo,  $\text{cov}(\varepsilon_{px1}, F_{mx1}) = E(\varepsilon F') = 0$ .

O modelo (1) implica uma estrutura de covariância para o vetor aleatório,  $X$ . Dessa forma:

$$\Sigma = \text{cov}(X) = LL' + \Psi \tag{2}$$

ou

$$\begin{aligned} \text{var}(X_i) &= \sigma_i^2 = l_{i1}^2 + \dots + l_{im}^2 + \Psi_i \\ \text{cov}(X_i, X_k) &= l_{i1}l_{k1} + \dots + l_{im}l_{km} \end{aligned} \tag{2'}$$

Onde:  $h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2$  é chamado de comunalidade, isto é, a parcela da variância da variável  $X_i$  que é distribuída pelos  $m$  fatores comuns.  $\psi_i$  é a parcela da var ( $X_i$ ), denominada de variância específica.

Neste trabalho, utilizaram-se nove itens do questionário do aluno para a elaboração do nível socioeconômico da família do estudante para as quartas séries do ensino fundamental e terceiros anos do ensino médio. As questões são as que se seguem.

A disponibilidade dos seguintes itens na casa do estudante (respostas de 0 a 4): i) televisão em cores, ii) rádio, iii) quartos para dormir, iv) automóvel/carro, v) diarista. Além desses, incluímos uma variável que visa captar a presença de uma pequena biblioteca em casa, cuja resposta varia entre 0 e 3.<sup>1</sup>

A existência ou não destes itens na casa do estudante (não=0 e sim=1): i) videocassete e ii) aspirador de pó. E, por fim: até que série os pais estudaram. As ponderações assumiram os critérios estabelecidos pela ABEP (Associação Brasileira de Empresas de Pesquisa - 2003).<sup>2</sup>

Ressalta-se que aqueles itens dos questionários que permitiam aos estudantes assinalarem a alternativa "não sei" foram desconsiderados, já que esta resposta não agrega qualquer tipo de informação na análise.

A técnica utilizada para a extração dos fatores foi a de Componente Principal, cujos objetivos, de acordo com Mingoti (2005), são a redução dos dados e a interpretação dos componentes. Isto é, semelhante à análise fatorial, realizar-se-á uma transformação linear das variáveis originais, em geral correlacionadas, resultando em novas variáveis, denominadas de componentes principais.

O primeiro componente principal é a combinação linear com a maior variância explicada possível. O segundo componente procura explicar a segunda maior parcela da variância restante e assim sucessivamente. É importante destacar que as novas variáveis geradas são não-correlacionadas entre si e apresentam a mesma variação total das variáveis originais. De acordo com Johnson e Wichern (1998), é o método mais indicado quando o objetivo é encontrar um número mínimo de fatores necessários para substituir as variáveis originais. Logo, sendo um dos objetivos do trabalho a mensuração da influência do nível socioeconômico e das variáveis de infra-estrutura sobre o desempenho do estudante, realizou-se apenas a extração da primeira componente principal.

A seguir, têm-se as tabelas que mostram o total de variância original das variáveis explicadas após a extração da primeira componente principal na seguinte ordem: o nível socioeconômico para as quartas séries do ensino fundamental e terceiros anos do ensino médio.

<sup>1</sup> 0 = não existem livros em casa, 1 = de um a 20 livros, 2 = entre 20 e 100 livros, e 3 = mais de 100 livros.

<sup>2</sup> 0 = analfabeto e séries iniciais do ensino fundamental incompleto; 1 = séries iniciais do ensino fundamental completo e fundamental incompleto; 2 = fundamental completo e médio incompleto; 3 = médio completo e superior incompleto; 5 = superior completo. Mais informações no site: [www.abep.org](http://www.abep.org).

TABELA A.11 - TESTES KMO E BARTLETT (4.º SÉRIE - ENSINO FUNDAMENTAL)

CRITÉRIOS		VALORES
Medida de adequação de Kaiser-Meyer-Olkin		0,874
Teste de esfericidade de Bartlett	Chi-quadrado aprox.	101585,6
	G.L.	36
	Sig.	0,000

TABELA A.12 - TESTES KMO E BARTLETT (3.º ANO - ENSINO MÉDIO)

CRITÉRIOS		VALORES
Medida de adequação de Kaiser-Meyer-Olkin		0,839
Teste de esfericidade de Bartlett	Chi-quadrado aprox.	42926,719
	G.L.	28
	Sig.	0,000

Segundo os critérios de esfericidade de Bartlett e de Kaiser-Meyer-Olkin (KMO), com um nível de significância (,000), pode-se rejeitar a hipótese nula de que os dados não são adequados para a aplicação da técnica de análise fatorial.

O resultado da Análise fatorial para o nível socioeconômico está resumido nas tabelas A.13 e A.14, a seguir:

TABELA A.13 - ANÁLISE FATORIAL (4.º SÉRIE - ENSINO FUNDAMENTAL)

Componente	VARIÂNCIA TOTAL EXPLICADA					
	Autovalores iniciais			Extração da Soma dos Quadrados		
	Total	% de Variância	% Acumulada	Total	% de Variância	% Acumulada
1	3,699	41,104	41,104	3,699	41,104	41,104
2	1,017	11,295	52,399			
3	0,787	8,740	61,138			
4	0,765	8,503	69,642			
5	0,707	7,856	77,497			
6	0,621	6,895	84,392			
7	0,514	5,710	90,102			
8	0,476	5,291	95,394			
9	0,415	4,606	100,000			

Método de extração: Análise de Componente Principal

TABELA A.14 - ANÁLISE FATORIAL (3.º ANO - ENSINO MÉDIO)

Componente	VARIÂNCIA TOTAL EXPLICADA					
	Autovalores iniciais			Extração da Soma dos Quadrados		
	Total	% de Variância	% Acumulada	Total	% de Variância	% Acumulada
1	3,239	41,486	40,486	3,239	40,486	40,486
2	1,008	12,599	53,085			
3	0,920	11,501	64,586			
4	0,715	8,933	73,519			
5	0,656	8,203	81,722			
6	0,571	7,132	88,854			
7	0,484	6,054	94,908			
8	0,407	5,092	100,000			

Método de extração: Análise de Componente Principal

O software utilizado para a realização da análise fatorial foi o SPSS, e a variância total explicada pelas 9 variáveis no primeiro caso foi de 41,104%. No segundo caso, a análise fatorial foi executada sobre 8 variáveis e a variância total explicada foi de 40,486%. Extraíu-se somente a primeira componente principal. A tabela a seguir apresenta as cargas fatoriais dessas variáveis. Os valores referem-se às correlações de cada variável com a componente principal e o valor de cada carga corresponde à quantidade de variância que é expressa pelo fator. Portanto, uma carga de 0,66 (nível de escolaridade da mãe) significa que 36% da variância original é explicada por esse fator. Assim, o nível socioeconômico fica elaborado da seguinte forma:

TABELA A.15 - MATRIZ DE COMPONENTES (4.º SÉRIE DO ENSINO FUNDAMENTAL)<sup>(1)</sup>

	1.º COMPONENTE
QT_TV	0,732
QT_RADIO	0,566
VIDEO_K7	0,68
QUARTO	0,495
CARRO	0,711
LIVRO	0,578
DIARISTA	0,655
MÃE	0,668
PAI	0,648

Método de extração: Análise de Componente Principal

(1) 1 componente extraído.

$$NSE_{ij} = qt\_tv(0,732) + qt\_radio(0,566) + video\_k7(0,68) + quarto(0,495) + carro(0,711) + livro(0,578) + diarista(0,655) + mae(0,668) + pai(0,648)$$

TABELA A.16 - MATRIZ DE COMPONENTES (3.º ANO DO ENSINO MÉDIO)<sup>(1)</sup>

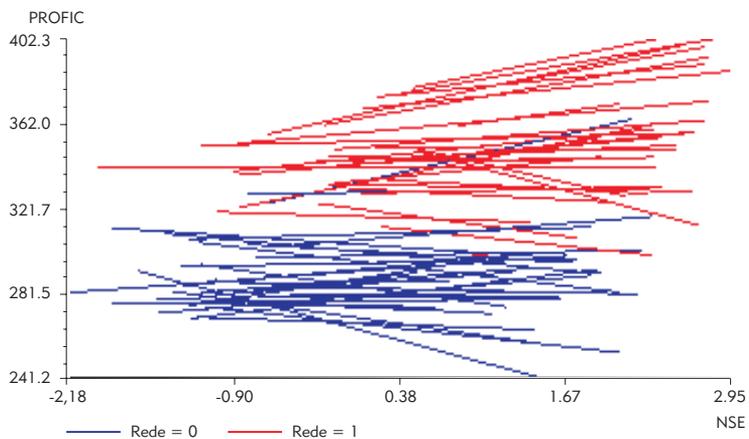
	1º COMPONENTE
QT_TV	0,743
QT_RADIO	0,654
ASPIRA_PÓ	0,667
CARRO	0,644
LIVRO	0,591
DIARISTA	0,203
MÃE	0,720
PAI	0,700

Método de extração: Análise de Componente Principal

(1) 1 componente extraído.

$$NSE_{ij} = qt\_tv(0,743) + qt\_radio(0,654) + aspira\_po(0,667) + carro(0,644) + livro(0,591) + diarista(0,203) + mae(0,720) + pai(0,700)$$

GRÁFICO A.1 - DESEMPENHO DAS ESCOLAS PÚBLICAS E PRIVADAS



FONTE: Os autores