

UNIVERSIDADE FEDERAL DO CEARÁ  
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E CONTABILIDADE  
CURSO DE PÓS-GRADUAÇÃO EM ECONOMIA

**LEANDRO OLIVEIRA COSTA**

**EFEITOS DA GESTÃO ESCOLAR E CARACTERÍSTICAS INDIVIDUAIS DO  
DIRETOR DETERMINANTES DO DESEMPENHO DOS ESTUDANTES DO ENSINO  
FUNDAMENTAL**

FORTALEZA

2006

**LEANDRO OLIVEIRA COSTA**

**EFEITOS DA GESTÃO ESCOLAR E CARACTERÍSTICAS INDIVIDUAIS DO  
DIRETOR DETERMINANTES DO DESEMPENHO DOS ESTUDANTES DO ENSINO  
FUNDAMENTAL BRASILEIRO**

Dissertação submetida à coordenação do Curso de Pós-Graduação em Economia, da Universidade Federal do Ceará, como requisito parcial para obtenção do grau de Mestre em Economia.

**Orientador: Prof. Dr. Ronaldo de Albuquerque  
e Arraes**

FORTALEZA

2006

Esta Dissertação foi submetida como parte dos requisitos necessários à obtenção do grau de Mestre em Economia, outorgada pela Universidade Federal do Ceará - UFC, e encontra-se à disposição dos interessados na Biblioteca do Curso de Pós-Graduação em Economia – CAEN da referida Universidade.

A citação de qualquer trecho desta dissertação é permitida, desde que seja feita em conformidade com as normas científicas.

Dissertação aprovada em \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_.

---

Leandro Oliveira Costa

---

Prof. Dr. Ronaldo de Albuquerque e Arraes  
(Orientador)

---

Prof. Dr. Paulo de Melo Jorge Neto  
(Membro da Banca Examinadora)

---

Prof. Dr. Marcos Costa Holanda  
(Membro Externo)

Ao meu colega e amigo Luciano  
(*in memoriam*).

## AGRADECIMENTOS

Devo sinceros agradecimentos aos meus pais Antônio Boanerges da Costa e Rosa das Graças Oliveira Costa, por todos os sacrifícios aos quais se submeteram para prover-me a melhor educação possível, dentro e fora do meu lar.

Ao Prof. Ronaldo de Albuquerque e Arraes, pela atenção dispensada na orientação desta pesquisa.

Aos professores Paulo de Melo Jorge Neto e Marcos Costa Holanda, por fazerem parte da banca examinadora, e por suas relevantes contribuições para o trabalho aqui desenvolvido.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão de financiamento, através da bolsa de estudo durante o curso.

Ao amigo Vitor Hugo, pela colaboração na obtenção e tratamento dos dados, dentre outros.

Aos amigos da turma de 2004 e 2003 do CAEN, em especial Antônio Germano, André Loureiro, Bruno Vinicius, Bruno Wichmann, Daniel de Pádua, Jimmy Oliveira, Rodolfo, Guilherme, Hélio, Guaraciana, Zilar, Cristiano e outros.

E a Malena, pelo apoio nos momentos mais difíceis do mestrado.

## RESUMO

Utilizando as metodologias de Modelos Lineares Hierárquicos e Modelos Estruturais com Dados em Painel sobre os dados do SAEB 2001 e 2003, foram analisados os efeitos da gestão escolar e das características individuais do diretor determinantes ao desempenho cognitivo dos estudantes do ensino fundamental brasileiro. As variáveis sobre as características individuais dos alunos apresentaram resultados semelhantes à literatura nacional, ou seja, as condições socioeconômicas e a raça são fatores relevantes para explicar a performance nos exames de proficiência em matemática. Entretanto, constatou-se um resultado distinto quanto à melhor performance dos estudantes do sexo feminino, o qual pode estar relacionado com o que se obteve para a variável repetência, visto que a maioria dos estudantes que repetem são do sexo masculino. Entre as vinte e sete variáveis ao nível da escola, somente oito divergiram os resultados entre os dois modelos. Verificou-se que a qualidade dos educadores, professores e diretores, pode ser medida tanto pelos seus níveis de escolaridade quanto por suas experiências no ambiente escolar. De onde se conclui que a estrutura física e pedagógica deve ser direcionada ao incentivo à alta frequência dos alunos e professores, ou seja, ambientes modernamente equipados e bem conservados. No cômputo geral dos resultados pode-se concluir que a escola brasileira necessita de recursos financeiros para serem aplicados em salários mais altos para diretores e professores, docentes mais qualificados, e em uma estrutura que incentiva a presença dos estudantes (como computadores com acesso à Internet) para melhorar o desempenho cognitivos dos estudantes.

Palavras-chaves: Performance estudantil, Gestão Escolar, Modelos Lineares Hierárquicos, Modelos com Dados em Painel.

## ABSTRACT

By using the methodologies of Hierarchic Linear Models and Structural Models with Panel Data on the data of SAEB in 2001 and 2003, it was analyzed the effect of school management and the principal's individual characteristics determinants to the students cognitive performance in the Brazilian primary schooling. When analyzing the individual characteristics of the students, it has been found similar results to national literature, that is, the socioeconomic conditions and the race are important factors to determine the performance in the examinations of proficiency in math. An interesting result is the best performance of the female students. This apparent contradiction may be related to what it was obtained for the schooling recurrence variable, since the majority of the repeater students are male. Among the twenty seven school variables, only eight of them diverged the results between the two models. The results obtained verified that the quality of the educators, teachers and principals, may be measured by their levels of schooling and experience in the school environment. Also signaling that the physical and pedagogical structure must be directed to the incentive the high frequency of the students and teachers, that is, well conserved and modernly equipped environments. The overall results indicate that the Brazilian schooling system needs financial resources to be applied to higher wages for teachers and principals, teachers more qualified, and a better structure that stimulates the students to go to school (as computers with access to the Internet) to improve the cognitive performance of the students.

Keywords: Student Performance, School Management, Hierarchic Linear Models, Panel Data Model.

## SUMÁRIO

Erro! Nenhuma entrada de índice analítico foi encontrada.

## LISTA DE QUADROS E TABELAS

<b>QUADRO 1:</b> Descrição das variáveis utilizadas na pesquisa.....	11
<b>TABELA 1:</b> Número de escolas de ensino fundamental e médio classificado por rede e localização.....	1
<b>TABELA 2:</b> Número de participantes nos SAEB 2001 e 2003.....	13
<b>TABELA 3:</b> Proporção de escolas públicas e privadas nos SAEB 2001 e 2003.....	13
<b>TABELA 4:</b> Proficiência dos participantes nos SAEB 2001 e 2003.....	14
<b>TABELA 5:</b> Resultado dos Testes de Hausman e de Durbin-Watson.....	29
<b>TABELA 6:</b> Modelo mais simples, somente a matriz localização e competência.....	30
<b>TABELA 7:</b> Modelo com as variáveis relativas à Administração e a Idade Média dos Estudantes.....	31
<b>TABELA 8:</b> Estimação do Modelo com variáveis relativas às dificuldades enfrentadas pelo Gestor (relativas a problemas eventuais e estruturais).....	33

<b>TABELA 9:</b> Resultados do modelo completo, com todas as variáveis do modelo.....	34
<b>TABELA 10:</b> Resultados do modelo completo considerando somente escolas públicas. ....	35
<b>TABELA 11:</b> Resultados do modelo completo considerando somente escolas privadas.....	37
<b>TABELA 12:</b> Resultados do Modelo ANOVA.....	38
<b>TABELA 13:</b> Resultados do modelo com as características individuais dos alunos.....	39
<b>TABELA 14:</b> Resultados do modelo com as características dos alunos e da escola.....	40
<b>TABELA 15:</b> Resultados do modelo com as características dos alunos e da escola pública.....	42
<b>TABELA 16:</b> Resultados do modelo com as características dos alunos e da escola privada.....	45
<b>TABELA A.1:</b> Estimação com efeitos fixos para o modelo completo.....	54
<b>TABELA B.1:</b> Análise Descritiva das Variáveis do Modelo de Dados em Painel.....	55
<b>TABELA B.2:</b> Estatísticas Descritiva das Variáveis do Modelo Linear Hierárquico.....	56
<b>TABELA C.1:</b> Resultado da análise fatorial.....	60
<b>TABELA C.2:</b> Matriz de Componentes Principais.....	60
<b>TABELA C.3:</b> Matriz de correlação.....	61

## LISTA DE FIGURAS

<b>FIGURA 1:</b> Proficiência em Matemática no SAEB 2001, discriminada por série.....	9
---	---

Costa, Leandro Oliveira.

Efeitos da gestão escolar e características individuais do diretor determinantes do desempenho dos estudantes do ensino fundamental / Leandro Oliveira Costa – Fortaleza, 2006.

65 f.

Dissertação (Mestrado) – Universidade Federal do Ceará – UFC, Curso de Pós-graduação em Economia – CAEN.

1. Econometria. 2. Educação. 3. Economia-Educação. 4. Análise de desempenho. I. Título

CDD – 330.015195

## 1 Introdução

A Constituição de 1988 consagrou o princípio da obrigatoriedade e do acesso gratuito ao ensino no Brasil, ou seja, todos os brasileiros na faixa de acesso ao ensino fundamental – de sete a quatorze anos – têm o direito a esse nível de escolaridade garantido. Mas, devido às altas taxas de retenção, reprovação e de evasão escolar, jovens em desvantagens socioeconômicas tendem a deixar o sistema educacional antes do previsto. Com isso, essas problemáticas contribuem ao acirramento da distribuição de renda brasileira, dado que cerca de 50 % dos desníveis distributivos estão associados à educação (Menezes-Filho et al, 2004)<sup>1</sup>.

Soares (2004), ao analisar os resultados do Censo Escolar de 2002, verifica que quase todas as crianças brasileiras estão matriculadas no ensino fundamental. Entretanto, há um número maior de matrículas na escola básica do que de estudantes na respectiva faixa etária, 35 milhões contra 28 milhões, pelo qual se conclui o reflexo não só nas altas taxas de repetência, mas também na entrada tardia, na evasão e no abandono de estudantes. Ainda neste panorama de estoque educacional, verifica-se, na tabela 1, que as escolas públicas brasileiras são maioria, por conseguinte, atendem a maior parte da demanda. Muito embora detenham 13% das instituições de ensino fundamental, as escolas privadas

---

<sup>1</sup> Segundo esses autores, o retorno da educação no Brasil é muito alto e o acesso aos níveis mais elevado de educação se dá apenas por uma pequena parcela da população.

absorvem somente 9% dos estudantes. Cabe observar também que, apesar do número de escolas rurais ser muito grande, são pequenas e com grande dispersão geográfica.

**Tabela 1:** Número de escolas de ensino fundamental e médio classificado por rede e localização.

Tipos de escola	Nível educacional	
	Fundamental	Médio
Pública	159,228	13,916
Particular	18,552	6,304
Urbana	70,410	19,399
Rural	107,370	821
Total	177,780	20,220

Fonte: Soares (2004) dados originais - MEC Censo Escolar, 2002.

Dada a conjuntura educacional brasileira, Soares (2004), ao analisar o desempenho cognitivo de estudantes do ensino fundamental brasileiro, constatou a existência de grandes diferenças raciais, sociais e entre regiões do país. Além disso, conclui também que embora a grande maioria dos estudantes detenha desempenho menor que o esperado para sua série, a educação básica brasileira apresenta sérios problemas de qualidade e equidade.

A pesquisa educacional em países desenvolvidos aponta que os fatores que influenciam a equidade e a efetividade dos resultados escolares estão associados a fatores externos à escola: as características socioeconômicas e culturais dos estudantes. Ou seja, as condições oferecidas pela sociedade e família antes e durante o período de escolarização é que são responsáveis pelos problemas da educação.

Porém, há evidências que a escola brasileira não superou a fase de investimentos básicos em educação, pois os recursos escolares tanto físicos quanto pedagógicos ainda produzem efeitos significativos sobre o desempenho dos estudantes. Os resultados convergem em relação ao efeito positivo da infra-estrutura, da organização e gestão da escola (baseada em liderança do diretor e em comprometimento coletivo do corpo docente com o aprendizado dos alunos) e do clima acadêmico orientado ao processo de aprendizagem. Também há evidências, ainda que mais esparsas, em favor do efeito positivo do nível educacional de professores, do salário de professores e de estilo pedagógico sintonizado com o movimento de renovação do ensino de Matemática. Ou seja, já foram identificadas algumas variáveis que influenciam o nível de desempenho cognitivo dos alunos do ensino fundamental, mas não se encontrou variáveis que possam explicar as diferenças de desempenho nos grupos definidos por raça e níveis socioeconômicos.

Contribuindo para essa linha de pesquisa, Menezes-Filho (2004) verificou que por meio da implementação do FUNDEF<sup>2</sup> registrou-se um aumento do salário relativo de professores de escola pública e, conseqüentemente, na contribuição à proficiência dos respectivos estudantes. Franco (2005), por sua vez, verificou que, além das variáveis relacionadas à composição social da escola brasileira, devem ser consideradas, como variáveis de controle, os fatores escolares associados à eficácia escolar. Esses fatores podem ser organizados em cinco categorias: recursos escolares; clima acadêmico; formação e salário docente; ênfase pedagógica; organização e gestão da escola.

Entretanto, características específicas da organização e gestão não foram ressaltadas na pesquisa dos fatores escolares relevantes a eficácia escolar. As características educacionais e a experiência de gestão do diretor, a promoção e participação de atividades continuadas, a existência de projeto pedagógico, conselhos de classe, critérios de seleção de estudantes e de formação de turmas são exemplos de importantes fatores não analisados. Logo, a constatação dessa lacuna na literatura motiva a investigação de informações sobre a gestão e características do diretor que melhorem as evidências sobre a qualidade do ensino básico brasileiro.

Com base nos dados sobre os estudantes da 8ª série do ensino fundamental dos SAEB 2001 e 2003<sup>3</sup>, pretende-se analisar qual a influência de variáveis, não enfatizadas nos trabalhos anteriores, relativas à organização e gestão escolar sobre a proficiência média dos estudantes. Para tanto, confrontam-se as metodologias de Modelos Lineares Hierárquicos (MLH) e a de Modelos Estruturais com Dados em Painel na identificação e controle das problemáticas normalmente ocorridas em microdados, como viés de agregação, endogeneidade e omissão de variáveis. Ou seja, essa pesquisa tem o intuito de melhorar as evidências que buscam entender a estrutura e qualidade da educação brasileira.

Além desta introdução, temos a seguinte estrutura: revisão bibliográfica no capítulo 2, uma análise dos dados no capítulo 3, exposição da metodologia no capítulo 4, e a posteriores a discussão dos resultados no capítulo 5, para em seguida extrair as principais conclusões.

## 2 Revisão da Literatura

---

<sup>2</sup> O FUNDEF é o Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério.

<sup>3</sup> O SAEB é o Sistema de Avaliação da Educação Básica implementado pelo governo brasileiro.

A tentativa de inserir esta pesquisa no conjunto de obras que visam melhorar as evidências sobre a estrutura e qualidade da educação brasileira, far-se-á mediante uma análise do cenário no âmbito educacional. O qual, por sua vez, está dividido em três relevantes linhas de pesquisa, teóricas e empíricas, cada uma desenvolvida independentemente. As áreas de pesquisa são relativas à comparação entre setores, análise da efetividade das escolas e o confronto das teorias burocrática e comunal.

Lee (2000), analisando a primeira corrente que compara a influência de diferentes setores ("cross-sector comparisons"), afirma que essa linha sugere que as características organizacionais da escola influenciam tanto no desempenho estudantil dentro das escolas quanto na distribuição da performance entre estudantes com diferentes características sociais ("social background"). Segundo a pesquisa de Coleman et al. (1987), estudantes americanos de escolas católicas têm melhor performance que os de escolas públicas, além de uma relação mais fraca entre as características sociais e os resultados dos estudantes em escolas católicas.

A segunda corrente ("effective schools") analisa a efetividade da escola, preocupando-se em identificar as características escolares que tornam as escolas capacitadas em qualificar estudantes desfavorecidos, cuja formulação contém dois estágios: o primeiro busca identificar as escolas que mostram sua efetividade sobre os estudantes desfavorecidos, do ponto de vista social e econômico; no segundo estágio, analisam-se as características em comum entre as escolas nas quais as efetividades foram de fato identificadas.

A terceira corrente confronta as teorias burocrática e comunal no que tange à organização das escolas. No modelo burocrático ou racional, a organização das escolas se caracteriza por uma divisão funcional do trabalho dos gestores e professores em função da especialização de cada um. As regras de ensino são definidas em questões subjetivas, considerando os tipos de estudantes. As interações sociais são definidas por regras, a afetividade é neutra, a autoridade é regida por regras e não pelo indivíduo que segue as regras. Contrariamente, a perspectiva comunitária ou pessoal observa a escola como uma pequena sociedade, onde a organização tem ênfase em relacionamentos informais e resistentes socialmente.

Como forma objetiva ao exposto, o presente trabalho se insere na linha de pesquisa que analisa a efetividade das escolas, ou seja, a "schools effective". Os resultados da literatura internacional dessa linha de pesquisa apontam que os fatores que influenciam a efetividade e a equidade dos resultados escolares estão associados a fatores extrínsecos à escola. Ou seja, os resultados almejados por uma escola são associados às características socioeconômicas e culturais dos estudantes e às condições oferecidas pela sociedade e família anterior e em curso ao processo de escolarização. Coleman et al

(1966) observou esses resultados para as escolas americanas, o relatório de Ploden para as escolas da Inglaterra e o INED<sup>4</sup> para as escolas da França (Nogueira, 1990, e Forquin, 1995).

Soares (2003) conclui que estes resultados sugerem que diferentes maneiras de organização das escolas teriam os mesmos resultados na avaliação do desenvolvimento cognitivo dos alunos, ou seja, “as escolas não fazem diferença”. Porém, há evidências que na escola brasileira não houve a superação da fase de investimentos básicos, pois a infra-estrutura escolar ainda produz efeitos significativos sobre a proficiência dos estudantes brasileiros.

Ruter et al. (1979) critica os resultados acima ao afirmar que é mais relevante avaliar o impacto sobre a proficiência dos alunos, dado o acréscimo unitário de um fator interno da escola. Franco (2005), ao fazer uma revisão do cenário da pesquisa educacional brasileira, divide os fatores escolares associados à eficácia escolar em cinco categorias: recursos escolares; clima acadêmico; formação e salário docente; ênfase pedagógica; organização e gestão da escola.

A análise da primeira categoria tem verificado que os recursos escolares no Brasil, relativos à conservação dos equipamentos e do prédio da escola, são relevantes para a performance dos alunos (Albernaz, Ferreira e Franco, 2002). Ademais, encontrou-se um efeito positivo da infra-estrutura física da escola sobre o desempenho em leitura dos alunos brasileiros (Lee, Franco e Albernaz, 2004) e resultados positivos para o efeito das condições de funcionamento de laboratórios e espaços adicionais para atividades pedagógicas (Espósito, Davis e Nunes, 2000). Entretanto, em muitos países desenvolvidos os recursos escolares não são fatores de eficácia escolar, isso ocorre em virtude do grau de conservação e qualidade das escolas e equipamento que não varia muito de escola a escola (Franco, 2005).

A categoria clima acadêmico engloba as diversas características escolares relacionadas à ênfase em atividades acadêmicas na escola, onde alguns estudos da literatura brasileira mostraram a associação com a eficácia escolar. Um exemplo é a influência da ênfase em passar e corrigir dever de casa reportada por Albernaz, Ferreira e Franco (2002), e o efeito positivo tanto do interesse e dedicação do professor quanto do nível de exigência do docente sobre o desempenho médio das escolas (Soares, 2004).

---

<sup>4</sup> O INED é o órgão governamental responsável pelos estudos demográficos.

A categoria formação e salários dos docentes observam resultados questionáveis, quando relacionados à eficácia escolar. A magnitude desses efeitos é relativamente pequena e a significância estatística dos achados pode não ser tão robusta quanto às outras variáveis (Franco, 2005). Albernaz, Ferreira e Franco (2002) reportaram efeito positivo da variável nível de formação docente sobre a eficácia escolar. Por sua vez, Franco (2005) afirma que o elemento fundamental para a significância estatística desse achado foi a existência de professores com nível de titulação aquém do mínimo de escolarização legalmente prescrito. Soares e Alves (2003) reportaram resultados convergentes com os de Albernaz, Ferreira e Franco (2002), ainda que os modelos tenham sido especificados de modo distinto. Porém, Franco (2005) recomenda que estudos sobre o assunto deveriam utilizar outras bases de dados e metodologias para melhor esclarecer esses resultados.

A categoria ênfase pedagógica aponta resultados positivos da ênfase em ensino orientado pela reforma da educação matemática, isto é, ênfase em raciocínios de alta ordem e em resolução de problemas genuínos e contextualizados (Franco, Sztajn e Ortigão, 2004). Mambrini, Pereira e Alves (2001) encontraram, antes da implementação de controle pelo status socioeconômico médio da escola, efeito positivo associado aos métodos ativos de ensino, no entanto, o efeito tornou-se estatisticamente nulo quando da implementação do controle pelo status socioeconômico médio das escolas. Franco (2005) afirma que a discrepância entre os resultados mencionados devem estimular pesquisadores a investigarem o assunto de modo mais sistemático e aprofundado.

Na categoria organização e gestão da escola, Albernaz, Ferreira e Franco (2002) e Soares e Alves (2003) mostram que o reconhecimento por parte dos professores da liderança do diretor é característica associada à eficácia escolar. Utilizando a mesma base de dados, referência da 8ª série do SAEB 2001, Soares, Mambrini, Pereira e Alves (2001) também analisaram a dedicação do diretor como fator de eficácia escolar. Lee, Franco e Albernaz (2004), utilizando dados do PISA 2000<sup>5</sup> para o Brasil, mostram que a responsabilidade coletiva dos docentes, medida pelo empenho coletivo dos docentes quanto ao aprendizado dos alunos, é fator relevante para a eficácia escolar.

Franco (2005) resume que a literatura brasileira sobre eficácia escolar apresenta achados convergentes sobre o efeito positivo dos recursos escolares (ainda que os pesquisadores entendam que

---

<sup>5</sup> PISA – Project of International Student Achievement – é um *survey* educacional que mede o conhecimento e habilidades em leitura, Matemática e Ciências de jovens de 15 anos nos principais países industrializados. Os testes são desenvolvidos para medir a capacidade dos jovens de usar seus conhecimentos e habilidades nas suas atividades cotidianas, não refletindo, portanto, domínio de um currículo escolar.

recursos só podem ser eficazes quando efetivamente utilizados), da organização e gestão da escola (baseada na liderança do diretor e no comprometimento coletivo do corpo docente com o aprendizado de seus alunos) e do clima acadêmico orientado para as exigências acadêmicas do processo de ensino e de aprendizagem. A literatura examinada produziu também evidências, ainda que mais esparsas, em favor do efeito positivo do nível educacional e salário dos professores e do estilo pedagógico sintonizado com o movimento de renovação do ensino de Matemática.

Contudo, confirma-se que características específicas da organização e gestão não foram ressaltadas na pesquisa dos fatores escolares relevantes à eficácia escolar. As características educacionais e a experiência em gestão do diretor, a promoção e participação de atividades continuadas, a existência de projeto pedagógico, conselhos de classe, critérios de seleção de estudantes e de formação de turmas são exemplos de importantes fatores ainda não analisados na literatura. Como já mencionado, propõe-se preencher esta lacuna, como será visto adiante na proposta metodológica após a exposição da série de dados.

## 3 Amostra de Dados

### 3.1 SAEB

Utiliza-se nesta pesquisa dados do Sistema Nacional de Avaliação da Educação Básica (SAEB), o qual é uma pesquisa bianual realizada pelo Ministério da Educação para monitorar a qualidade, a equidade e a eficiência do ensino e da aprendizagem no âmbito do Ensino Fundamental e Médio da educação brasileira. Os instrumentos da pesquisa são testes de Língua Portuguesa e Matemática e questionários contextuais aplicados a uma amostra de escolas, alunos, professores e diretores.

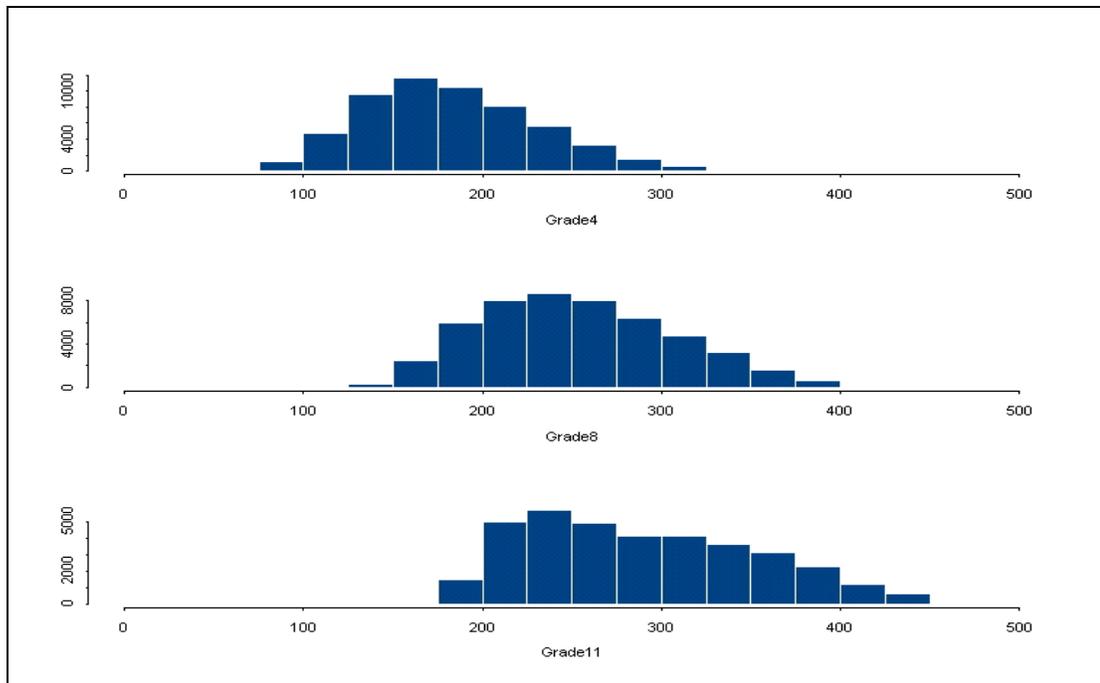
O SAEB tem informações sobre o desempenho em exames de proficiência de uma amostra de estudantes de escolas públicas e privadas referente aos anos de 1995, 1997, 1999, 2001 e 2003. Os dados contêm um conjunto detalhado de características de cada estudante, escola, professor e diretor, permitindo que controles importantes para um estudo deste tipo (como por exemplo, escolaridade dos diretores e professores, processo de seleção dos alunos etc) possam ser contemplados. Ainda, o exame do SAEB tem como referencial a Teoria de Resposta ao Item (TRI), que o torna comparável às pontuações obtidas em uma disciplina entre diferentes anos e séries (Soares, 2004).

Os testes são desenvolvidos com base em uma matriz de competências construídas para refletir não só as recomendações contidas nos Parâmetros Curriculares Nacionais, como também as matrizes utilizadas pelos diferentes sistemas estaduais. Essas matrizes de referência orientam o processo de construção das provas e dos itens que as compõem. Ou seja, traduzem a associação entre os conteúdos praticados nas escolas brasileiras de ensino fundamental e as competências cognitivas e as habilidades utilizadas pelos alunos no processo de construção do conhecimento.

Como o planejamento do teste inclui itens comuns entre as diferentes séries testadas e entre os diferentes anos, pode-se expressar a proficiência dos alunos testados nos diferentes ciclos do SAEB e nas diferentes séries na mesma escala. Naturalmente, esperam-se valores menores nessa escala para alunos da 4ª série do ensino fundamental, e maiores para alunos da 3ª série do ensino médio, visto que as primeiras diferenças observadas ao longo do tempo resultam das intervenções feitas no sistema entre os intervalos de aplicação ou são frutos de variação amostral. Para exemplificar, por meio da figura 1 temos a proficiência em matemática no SAEB 2001, discriminada por série. A metodologia de construção da proficiência está descrita em Klein e Fontanive (1995).

Em termos da medida de proficiência, o SAEB é particularmente competente e, do ponto de vista metodológico, não há dúvidas de que a proficiência medida pelo SAEB capta de forma adequada o nível de domínio das habilidades e competências cognitivas incluídas na sua matriz de especificação (Soares, 2005).

**Figura 1:** Proficiência em Matemática no SAEB 2001 discriminada por série.



Fonte: Soares (2003).

### 3.2 Dados do Modelo com Dados em Painel

No modelo com dados longitudinais, esta pesquisa utiliza dados do SAEB de 2001 e 2003 e os resultados dos exames de proficiência em Matemática para os estudantes da 8ª série do ensino fundamental. Como esses dados são “cross sections” repetidos, utilizou-se somente as escolas que participaram de ambos os SAEB (757 escolas). Os principais dados utilizados estão divididos em 5 grupos: proficiência; localização e competência; idade da turma; diretor; administração; problemas estruturais e eventuais. No Quadro 1 expomos e descrevemos todas as variáveis utilizadas na pesquisa.

**Quadro 1:** Descrição das variáveis do modelo.

Descrição das Variáveis		
Variáveis ao Nível dos Alunos	Abreviação	Descrição
<b>1. Aluno</b>		
<b>Proficiência</b>	Profic	Contínua.

<b>Gênero</b> do aluno	GE	0 = Masculino; 1 = Feminino.
<b>Raça</b> do aluno	RÇ	0 = Indígena ou Pardo ou Preto; 1 = Amarelo ou Branco.
<b>Idade</b> do aluno	ID	Varia de 12 a 19 anos.
Possui <b>Bolsa Escola</b>	BE	0 = não possui; 1 = possui.
Status Socioeconômico	SS	Continua.
Variáveis do Nível das Escolas		
<b>2. Professor</b>		
<b>Média de Idade dos Professores</b> de cada escola	med_id_prof	A média dos resultados: 0 = Até 24 anos; 1 = de 25 a 29; 2 = de 30 a 34 ; 3 = de 35 a 39; 4 = de 40 a 44; 5 = de 45 a 49; 6 = de 50 a 54; 7 = 55 anos ou mais.
<b>Média da Escolaridade Máxima dos Professores</b>	med_esc_prof	A média dos resultados: 0 = não completou o Ensino Médio; 1 = Ensino Médio – Magistério ou outros (antigo 2o Grau); 2 = Ensino Superior – Pedagogia ou Licenciatura em Letras ou Licenciatura em Matemática ou outras Licenciaturas.
<b>Média de Experiência em Ensino dos Professores</b>	med_exp_prof	A média dos resultados: 0 = há menos de 1 anos; 1 = de 1 a 2; 2 = de 3 a 5; de 6 a 9; 3 = de 10 a 15; 5 = de 15 a 20; 6 = mais de 20 anos.
<b>Média dos Salários Brutos dos Professores</b>	med_sal_bru_prof	0 = Até 1 salário mínimo; 1 = mais de 1 a 2 s. m.; 2= mais de 2 a 4 s. m.; 3= mais de 4 a 6 s. m.; 4 = mais de 4 a 9 s. m.; 5 = mais de 9 a 12 s. m.; 6 = mais de 12 a 16 s. m.; 7 = mais de 16 salários mínimos.
<b>3. Diretor</b>		
<b>Gênero</b> do diretor	GE_dir	0 = Feminino; 1 = Masculino
<b>Raça</b> do diretor	RÇ_dir	0 = Indígena ou Pardo ou Preto; 1 = Amarelo ou Branco
<b>Idade</b> do diretor	ID_dir	0 = Até 24 anos; 1 = de 25 a 29; 2 = de 30 a 34 ; 3 = de 35 a 39; 4 = de 40 a 44; 5 = de 45 a 49; 6 = de 50 a 54; 7 = 55 anos ou mais.
<b>Salário bruto do diretor</b> ( em 2001 o salário mínimo era de R\$ 180,00 e em 2003 era de 240,00).	Sal_bru_dir	0 = Até 1 salário mínimo; 1 = mais de 1 a 2 s. m.; 2= mais de 2 a 4 s. m.; 3= mais de 4 a 6 s. m.; 4 = mais de 4 a 9 s. m.; 5 = mais de 9 a 12 s. m.; 6 = mais de 12 a 16 s. m.; 7 = mais de 16 salários mínimos.
<b>Outra atividade do diretor. Máxima escolaridade do diretor</b> (completa).	Outr_ativ Esc_max	0 = não possui; 1 = possui. 0 = não completou o Ensino Fundamental (antigo 1º Grau); 1= Ensino Fundamental (antigo 1o Grau); 2 = Ensino Médio – Magistério ou outros (antigo 2o Grau); 3 = Ensino Superior – Pedagogia ou Licenciatura em Letras ou Licenciatura em Matemática ou outras Licenciaturas.
<b>Pós-graduação do diretor</b> (concluída).	Pos_grad	0 = não fez ou ainda não completou curso de pós-graduação; 1 = Aperfeiçoamento (mínimo de 180 horas); 2 = Especialização (mínimo de 360 horas); 3 = Mestrado Profissionalizante; 4 = Mestrado Acadêmico; 5 = Doutorado.
<b>Experiência em direção</b> (anos).	Exp_direç	0 = há menos de 2 anos; 1 = de 2 a 4; 2 = de 5 a 10; 3 = de 11 a 15; 4 = há mais de 15 anos.
<b>Experiência em educação</b> (anos).	Exp_educ	0 = há menos de 2 anos; 1 = de 2 a 4; 2 = de 5 a 10; de 11 a 15; 3 = de 16 a 20; 4 = há mais de 20 anos.

<b>4. Localização e Competência</b>		
<b>Rede</b>	Rede	0 = Pública; 1= Particular.
<b>Região</b>	Região	0 = Nordeste; 1 = Norte; 2 = Centro-Oeste; 3 = Sudeste; 4 = Sul.
2. Turma		
<b>Média da idade dos alunos.</b>	med_ID	Contínua
<b>5. Administração</b>		
<b>Promoção de atividade continuada</b> , como atualização, treinamento capacitação etc.	Prom_ativ	0 = não promoveu; 1 = promoveu.
<b>Participação do diretor de atividade continuada</b> nos últimos dois anos.	Part_ativ	0 = não participou ; 1 = participou.
<b>Conselho escolar</b> é um colegiado, constituído por representantes da escola e da comunidade, que tem como objetivo acompanhar as atividades escolares.	Cons_esc	0 = não tem ; 1 = tem.
<b>Projeto pedagógico.</b>	Proj_pedag	0 = não possui; 1 = possui.
<b>Critério de admissão dos alunos.</b>	C_ad_alun	0 = não existe critério pré-estabelecido; 1= sorteio ou local de moradia ou Prioridade por ordem de chegada ou outro critério; 2 = Prova de seleção.
<b>Critério de formação de turmas.</b>	C_form_turm	0 = heterogeneidade quanto à idade ou heterogeneidade quanto ao rendimento escolar ou não houve critério; 1 = homogeneidade quanto à idade; 2 = homogeneidade quanto ao rendimento escolar.
<b>Vínculo Estável</b> (estatutário ou celetista).	Vinc_estav	0 = menor ou igual a 25%; 1 = de 26% a 50%; 2 = de 51% a 75%; 3 = de 76% a 90%; 4 = de 91% a 100%.
<b>6. Estrutura e Problemas Eventuais</b>		
<b>Acesso à Internet</b> pelos alunos.	Net	0 = não possui; 1 = possui.
<b>Biblioteca.</b>	Bib	0 = não possui; 1 = possui.
<b>A infra-estrutura física e/ou pedagógica que pode dificultar a aprendizagem.</b>	Infra	0 = péssima ou ruim; 1 = razoável ou boa ou ótima.
<b>Inexistência de Professores para certas disciplinas ou séries.</b>	Sem_prof	0 = não; 1 = sim.
<b>Insuficiência de recursos Financeiros.</b>	Rec_finan	0 = não; 1 = sim.
<b>Interrupção de atividades escolares.</b>	Inerr_ativ	0 = não; 1 = sim.
<b>Alto índice de falta por parte dos professores.</b>	Falta_prof	0 = não; 1 = sim.
<b>Alto índice de falta por parte dos alunos.</b>	Falta_alun	0 = não; 1 = sim.

O grupo proficiência agrega os resultados das proficiências dos estudantes de cada escola, utilizando a média aritmética. O grupo localização e competência, refere-se às características da escola, quais sejam: a região; caso seja urbana ou rural; pública ou privada. O grupo idade da turma é dado pela média de idade dos estudantes de cada escola. O grupo denominado problemas estruturais e eventuais tenta verificar não só se a escola tem acesso à Internet e biblioteca, isto é, estruturas que levariam a uma

melhor qualidade educacional, mas também problemas eventuais de cada escola, a saber se os alunos ou professores faltam com frequência. O grupo diretor busca analisar características individuais, sobre a renda, a escolaridade e a experiência em educação e direção. Características como os critérios de admissão de alunos, existência de projeto pedagógico, formação de turmas, a promoção e participação de atividades continuadas e a existência de conselho escolar estão no grupo administração.

O número de participantes dos SAEB 2001 e 2003 estão na Tabela 2 abaixo.

**Tabela 2:** Número de participantes nos SAEB 2001 e 2003

Participantes	Ano de Realização	
	2001	2003
Escolas	6935	6437
Diretores	6820	6628
Docentes	21754	17376
Alunos	287719	218521

Fonte: INEP– MEC SAEB, 2001 e 2003.

Verifica-se que a quantidade de escolas e diretores nas amostras são próximos, mas o número de docentes e alunos é menor em 2003. A proporção de escolas públicas e privadas na amostra é exposta na tabela 3 e as estatísticas descritivas dos exames de proficiência para os estudantes da 8ª série estão na tabela 4.

**Tabela 3:** Proporção de escolas públicas e privadas nos SAEB 2001 e 2003.

Ano	Rede	
	Pública	Privada
2001	46%	54%
2003	60%	40%

Fonte: INEP– MEC SAEB, 2001 e 2003.

**Tabela 4:** Proficiência dos participantes nos SAEB 2001 e 2003.

Proficiência	Anos	
	2001	2003
Média	253.86	255.98
Máximo	274.21	428.24
Mínimo	124.47	116.32
Desv. Padrão	53.75	55.22

Fonte: INEP– MEC SAEB, 2001 e 2003.

Comparando a tabela 1 com a tabela 2, verifica-se que a amostra não é muito representativa da quantidade das escolas, pois a quantidade na população de escolas públicas é quase dez vezes maior do que o de escolas privadas. Dado que o nível mais importante da pesquisa é o estudante, espera-se

que, devido à numerosa quantidade de pequenas escolas rurais, a amostra selecionada para a aplicação dos testes tenha sido de escolas públicas com um grande número de estudantes na região urbana, com o intuito de diminuir os custos. Ademais, a única diferença significativa nas estatísticas sobre proficiência é o resultado máximo verificado em 2003.

Na tabela B.1, Apêndice B, após analisar somente as escolas presentes nos dados dos SAEB's 2001 e 2003, apresentamos os resultados descritivos preliminares para o modelo que utiliza dados longitudinais. A amostra é constituída apenas de escolas localizadas em área urbana, logo, não se poderá analisar essa característica da escola. Verifica-se o percentual de 62% de escolas públicas, ou seja, existem 473 escolas públicas e 285 escolas privadas. Relativo ao diretor, diminuiu o número de diretores do sexo masculino, de 30% para 26%, e a faixa de idade não mudou, 60% dos diretores estão na faixa etária 35 a 50 anos. Os salários dos diretores diminuíram, visto que, em 2001, 13% dos diretores ganhavam acima de 8 salários-mínimos e 55% recebiam entre 2 e 5 salários-mínimos; em 2003 somente 5% recebiam acima de 8 salários-mínimos e 70% entre 2 e 5 salários. A promoção de atividades continuadas diminuiu de 58% para 38% e a existência de algum critério de admissão de alunos diminuiu de 92% para 58%. Além disso, cerca de 40% das escolas têm mais de 90% de professores com vínculo estável.

### **3.3 Dados do Modelo Linear Hierárquico**

A optarmos pelo Modelo Linear Hierárquico utilizamos os dados do SAEB de 2003 que abrangeu uma amostra de 218521 alunos e 6437 escolas, mas somente 36908 alunos em 2092 escolas fizeram os exames de proficiência em Matemática para a 8ª série do ensino fundamental.

O MLH é caracterizado por considerar as variáveis em diferentes níveis, atribuindo-se os valores 1 e 2 para alunos e escolas, respectivamente. Um modelo com 3 níveis não foi escolhido por duas razões: a maioria dos resultados almejados nas pesquisas que utilizam modelos em multiníveis podem ser observados em modelos básicos de 2 níveis (Raudenbush & Bryk, 2002); o número de turmas na amostra é muito pequeno em cada escola.

Ao nível dos alunos, 25 414 estudantes são de escola pública e 11121 estudantes de escolas privadas, ou seja, 70% são de escolas públicas e 30% são de escolas privadas; ao nível das escolas, 66% são de escola pública e 34% são de escolas privadas, i. e., são 710 escolas privadas e 1382 escolas públicas.

No nível dos alunos, as variáveis observadas são o sexo, a cor, a idade, se o estudante possui bolsa escola, número de reprovações e o status socioeconômico dos alunos (SS). A inclusão das variáveis referentes ao sexo e cor tem por objetivo controlar possíveis efeitos do gênero e da etnia sobre o desempenho dos alunos.

Como um dos problemas do SAEB é a impossibilidade de criar uma variável de diferencial nos testes de rendimento, ou seja, criar uma medida de valor agregado da educação que corresponda ao aprendizado do estudante entre dois períodos (Albernaz, Ferreira e Franco, 2002), incluiu-se a variável repetência, ou seja, o número de reprovações. Uma vez que a variável dependente - resultado da proficiência do estudante - é uma medida única em todo o seu percurso escolar e reflete todo o aprendizado durante os anos de sua permanência na escola até o momento do teste, a inclusão da variável repetência busca espelhar, mesmo que em parte, o percurso escolar daqueles que fizeram o teste.

A variável SS é uma *proxy* do nível socioeconômico, pois um dos problemas dos dados obtidos pelo SAEB é que não há uma variável de renda familiar capaz de discriminar as diferenças de renda entre as famílias dos estudantes que estudam na mesma escola. Esta *proxy* é criada através da extração do componente principal de nove variáveis derivadas de respostas sobre exclusão social, evidência de riqueza familiar e bens educacionais da casa e a escolaridade dos pais.

O indicador de exclusão foi construído a partir da observação da presença de água encanada e luz elétrica na residência do estudante. O indicador de riqueza familiar e de bens educacionais foi construído com base na quantidade de seis itens: rádios; televisões a cores; geladeiras; freezers; máquinas de lavar roupa; automóveis. O indicador de escolaridade dos pais foi construído a partir da máxima escolaridade do pai ou da mãe do aluno. Vide a construção da variável SS no apêndice C.

A variável bolsa escola busca avaliar se o Programa Nacional do Bolsa Escola tem causado efeitos sobre o desempenho cognitivo dos alunos. Ele foi criado em 2001 com propósito de conceder benefício monetário mensal a milhares de famílias brasileiras de baixa renda em troca da manutenção de suas crianças nas escolas, uma vez que esta variável não está incluída na base de cadastro do SAEB 2001, não será possível utilizá-la no modelo de dados em painel.

O nível das escolas é o que dá condição para avaliar as principais variáveis da pesquisa e comparar as duas metodologias econométricas, pois as variáveis neste nível são as mesmas nos dois modelos e comportam as características do diretor e da gestão da escola. A descrição das variáveis do nível 2 já foi feita na seção anterior.

Ainda, inseriu-se variáveis relativas aos docentes para confrontar com os resultados das pesquisas já citadas. Entretanto, também não foi possível determiná-las para o modelo estrutural com dados em painel, pois os dados cruzados só obtinham 360 escolas em ambos os SAEB. As variáveis são: média de idade dos professores; média da escolaridade máxima; média de experiência em ensino; média dos salários brutos.

## 4 Metodologia

Hanushek (1986) faz uma resenha da visão dos economistas e educadores sobre a “função de produção educacional”, cuja especificação econométrica pode ser representada pela seguinte equação:

$$Y_i = \beta_0 + \sum_{j=1}^5 \beta_j X_{ij} + \mu_i + \varepsilon_{it} \quad (1)$$

Sendo que  $y$  denota o desempenho dos alunos, que é função das características pessoais do aluno ( $X_1$ ), como raça e gênero; características de suas famílias ( $X_2$ ), como rendas e escolaridade; características de seus colegas na escola ( $X_3$ ), como o status socioeconômico e cultural; características de seus professores ( $X_4$ ), como escolaridade, salários e experiência; outras características escolares ( $X_5$ ); e características que não podem ser observadas ( $\mu_i$ ).

Essa função de produção educacional nos dá embasamento para utilizar as metodologias econométricas para avaliar o efeito das características relacionadas ao diretor e a gestão de uma escola sobre o desempenho cognitivo dos alunos. Essas variáveis estão inseridas em  $X_5$ , ou seja, nas outras características escolares que implementam a função de produção educacional. Mas, somente no Modelo Linear Hierárquico se faz necessário a introdução de variáveis que se inserem em  $X_1$ ,  $X_2$  e  $X_4$ .

As metodologias utilizadas baseiam-se em dois tipos de modelos econométricos: Modelos Estruturais com Dados em Painel e Modelos Lineares Hierárquicos. O confronto dessas metodologias busca identificar e solucionar problemas que usualmente são citados na literatura que analisa a efetividade das escolas sobre o desempenho cognitivo dos alunos. Os principais possíveis problemas abordados, na literatura, pela primeira metodologia, a que utiliza dados longitudinais, são: heterogeneidade não observada; endogeneidade e variáveis omitidas. A segunda metodologia envolve os seguintes possíveis problemas: viés de agregação, erro de estimação do erro padrão e heterogeneidade das regressões.

### 4.1 Modelo Estrutural com Dados em Painel

A utilização de Modelos Estruturais com Dados em Painel ocorre por conta da provável existência de efeitos não observados, ou também chamado de heterogeneidade não observada, nas variáveis que tentam explicar os resultados das proficiências dos estudantes. Essa heterogeneidade não observada normalmente causa endogeneidade, ou seja, torna as variáveis explicativas observadas correlacionadas com os distúrbios aleatórios,  $Cov(x, \varepsilon) \neq 0$ .

A heterogeneidade não observada e, conseqüentemente, a endogeneidade podem ocorrer devido: (a) variáveis omitidas; (b) erro de medida; (c) simultaneidade. Variáveis omitidas ocorrem quando, ao querer se controlar por uma ou mais variáveis adicionais, mas devido à inviabilidade dos dados, não se pode incluir essas variáveis no modelo de regressão. Suponha que  $E(y/x, \mu)$  seja a expectativa condicional de interesse, que pode ser escrita como função linear das variáveis explicativas observadas,  $x$ , e das variáveis adicionais não observadas,  $\mu$ , que em virtude da problemática de variáveis omitidas se tornam parte do distúrbio aleatório,  $\varepsilon$ . Logo, se  $\mu$  e  $x$  são correlacionados, então  $x$  torna-se estocástica. Wooldridge (2002) afirma que normalmente isso ocorre por conta do “self-selection”, ou seja, a escolha dos valores de  $x$  pelos agentes pode depender dos fatores  $\mu$  que não são observados na análise. Um exemplo é a habilidade individual dos agentes em modelos que analisam os fatores que influenciam os resultados de proficiência dos estudantes, na qual as características socioeconômicas são possivelmente correlacionadas com a habilidade. Outro exemplo é o “efeito dos pares”, em um similar estudo que agrega as proficiências dos estudantes, onde as características estruturais de uma escola podem ser correlacionadas com a habilidade do conjunto dos estudantes que dividem uma mesma estrutura escolar.

A problemática de erro de medida ocorre quando, ao analisar o efeito parcial de uma variável,  $x_k^*$ , somente observamos uma medida imperfeita,  $x_k = x_k^* + \lambda_k$ , tornando a variável explicativa estocástica. Ao utilizar  $x_k$  ao invés de  $x_k^*$ , adicionou-se outro erro ao original,  $\varepsilon$ . Isto ocorre, notadamente, quando o modelo teórico é especificado erroneamente. Um exemplo se dá quando se utiliza nível quando se deveria utilizar taxa.

A simultaneidade ocorre quando uma das variáveis explicativas é determinada simultaneamente com a variável dependente,  $y$ . Se  $x_k$  é determinado em parte por  $y$ , então  $x_k$  e  $\varepsilon$  são correlacionados.

Supomos um modelo populacional com a existência de heterogeneidade não observada ou um efeito não observado constante no tempo, fundamentado pela possível omissão de variáveis, que pode ser correlacionado ou não com as variáveis explicativas. O efeito não observado é considerado constante no tempo, pois o que interessa nesse estudo é o efeito parcial da variável explicativa observada na função de regressão populacional. Ou seja, somente o efeito da influência e não o da relação temporal. Ainda, espera-se que características como a habilidade do diretor não variem no tempo.

A existência de um efeito não observado e, conseqüentemente, de endogeneidade pode ser testada pelo teste de Breusch and Pagan. O teste baseia-se no multiplicador de Lagrange com o intuito de verificar se a variância de  $\mu_i$  é nula, esse teste é descrito em Greene (2003) e Wooldridge (2002).

O efeito não observado fundamentado na omissão de variáveis incentiva a utilização de  $\mu_i$  como uma variável aleatória inerente à população, esta observa  $y_{it}$  e  $x_{it}$  (Wooldridge, 2002). Porém, uma

questão a ser analisada previamente é se  $\mu_i$  deve ser tratado como um efeito aleatório ou um efeito fixo. Efeito aleatório é sinônimo de correlação zero entre as variáveis explicativas observadas e o efeito não observado,  $Cov(x_{it}, \mu_i) = 0$ . Efeito fixo é sinônimo de correlação entre a variável explicativa observada e o efeito não observado.

É necessário aplicar um teste para verificar se ocorrem efeitos fixos ou aleatórios. A existência de correlação entre o efeito não observado e as variáveis explicativas pode ser detectada por meio da aplicação do teste de Hausman<sup>6</sup>, que é usado para testar a ortogonalidade entre os efeitos aleatórios e os regressores. O teste de Hausman é baseado na diferença entre as estimativas dos efeitos aleatórios e fixos, cuja a idéia central é saber se o estimador de efeitos fixos é consistente quando  $\mu_i$  e  $x_{it}$  são correlacionados, mas o estimador de efeitos aleatórios é inconsistente. Uma diferença estatisticamente diferente é interpretada como uma evidência contra a suposição de efeitos aleatórios,  $E(\mu_i | x_i) = E(\mu_i) = 0$ .

Dependendo dos resultados dos testes para a presença de efeitos aleatórios ou fixos, o método de estimação será o dos Mínimos Quadrados Generalizados Factíveis (MQGF) para efeitos aleatórios ou fixos dados em Wooldridge (2002). Esse método considera a matriz de covariância de White para corrigir a possível autocorrelação.

Para verificar a influência das variáveis relativas às características do diretor, da administração e da localização da escola sobre a proficiência média dos estudantes, iniciou-se a estimação de modelos empíricos mais simples, com somente o vetor de variáveis referentes à localização e competência. Essa é uma prática comum na metodologia de dados em painel, pois, dado que os dados em painel corrigem a omissão de variáveis, insere-se um número crescente de variáveis controle a cada regressão para verificar se os resultados persistem. O primeiro modelo a ser estimado será:

$$y_{it} = L_{it}\phi + \mu_i + \varepsilon_{it} \quad (2)$$

Tal que o subscrito  $i$  refere-se à escola e  $t$  refere-se ao ano, 2001 e 2003;  $y_{it}$  é a média das proficiências dos estudantes de cada escola;  $L_{it}$  é matriz de variáveis relativas à competência e localização, ou seja, rede e região de cada escola;  $\phi$  é o vetor de parâmetros que medem os efeitos parciais da rede e região sobre a proficiência média dos estudantes.

---

<sup>6</sup> Veja Wooldridge (2003).

Espera-se que os resultados confirmem àqueles encontrados na literatura de modelos hierárquicos, a partir dos quais as escolas privadas e em regiões mais desenvolvidas tenham melhores resultados.

O modelo mais completo insere todas as variáveis no âmbito da escola que podem influenciar o resultado médio da proficiência média de cada escola. Estas variáveis são as que se inseriam na variável  $X_5$  do modelo que representa a função de produção educacional. O modelo pode ser representado por:

$$y_{it} = L_{it}\phi + T_{it}\lambda + A_{it}\delta + E_{it}\tau + D_{it}\theta + \mu_i + \varepsilon_{it} \quad (3)$$

Sendo que  $T_{it}$  é a média de idade dos estudantes da escola;  $A_{it}$  é a matriz de variáveis relativas a administração, como o critério de admissão dos alunos e a existência de projeto pedagógico;  $E_{it}$  é a matriz de variáveis relativas a estrutura da escola e  $D_{it}$  é a matriz de variáveis relativas ao diretor, com o salário, a experiência em educação e a máxima escolaridade adquirida.

Têm-se a expectativa que todas as variáveis relacionadas à administração das escolas tenham efeitos positivos sobre a performance média dos estudantes, ou seja, escolas um maior aparato administrativo tenham melhores resultados. Analisando a variável idade média dos estudantes, espera-se que escolas com a maioria dos alunos na respectiva faixa etária da série, tenham menos problemas de aprendizado e, conseqüentemente, melhores resultados. Por último, espera-se que diretores mais bem pagos em uma estrutura adequada tenham a capacidade de administrar uma escola, fazendo do ambiente escolar adequado ao aprendizado e ao desenvolvimento cognitivo dos alunos.

Caso se verifique a significância da variável rede, é necessário, num segundo momento, analisar como se comportam as variáveis, relativas aos estudantes e a escolas, quando se considera as diferenças entre as escola pública e privada. Ou seja, estimar-se-á separadamente tanto o modelo completo de dados em painel e o MLH para escolas públicas e privadas.

Uma das hipóteses para essas diferentes é que na administração das escolas privada se supõe que os gestores e professores podem ter qualquer iniciativa desde que ela não seja proibida por lei. Já na escola pública, pode-se fazer somente o que está previsto na lei. Outras hipóteses decorrem desde o perfil dos estudantes, perfil dos funcionários públicos, a relativa estabilidade no serviço público, até a importância da estrutura e ambiente escolar na educação das crianças.

#### **4.2 Modelos Lineares Hierárquicos - MLH**

Os Modelos Lineares Hierárquicos (Raudenbush & Bryk, 2002), também chamados de modelos de multinível (Goldstein, 2003) ou modelos de coeficientes aleatórios (Longford, 1993), são modelos de coeficientes aleatórios que a literatura educacional considera estatisticamente apropriado para a análise de variáveis aleatórias que não são independentes e identicamente distribuídas. As variáveis que caracterizam as características educacionais são não-aleatórias e agrupadas em subgrupos, como aluno agrupados em classes, classes agrupadas em escolas, escolas agrupadas em municípios e assim por diante. Logo, é insensato considerar que os coeficientes que relacionam as variáveis independentes às variáveis explicativas sejam constantes entre subgrupos, ou seja, que os resíduos sejam identicamente e independentemente distribuídos – iid.

Entretanto, o pesquisador pode ignorar os multiníveis naturais dos dados, desagregando ou agregando, e utilizar o método do MQO ou , considerando as categorias, a técnica ANOVA<sup>7</sup>. Ao desagregar as variáveis de nível superior, como turmas e escolas, ao nível dos indivíduos leva as características dos professores, da turma e da escola serem analisadas ao nível dos estudantes. O problema com essa aproximação é que os estudantes de uma mesma classe têm os mesmos valores das variáveis da turma ou da escola, então não é possível assumir a independência das observações que é uma suposição básica do MQO. Outra alternativa é agregar as variáveis referente aos estudantes em um nível superior, como turmas e classes, e fazer a análise sobre esse nível superior. Porém, o principal problema é descartar ou desperdiçar todas as informações internas dos grupos e, conseqüentemente, as relações entre as variáveis agregadas tornam-se mais fortes. Esse efeito da agregação deve ser esperado no modelo estrutural de dados em painel, que considera somente o nível das escolas.

As problemáticas que freqüentemente ocorrem na análise de dados em multiníveis quando utilizamos métodos de estimação que ignoram esses níveis são: viés de agregação, erro padrão estimado erroneamente e a heterogeneidade das regressões. O problema de viés de agregação pode ocorrer quando as variáveis seguem diferentes significados e, conseqüentemente, têm diferentes efeitos em diferentes níveis de agregação. Um exemplo ocorre quando as características sociais e econômicas dos estudantes de uma mesma escola são agregadas em uma variável agregada de nível socioeconômico. Essa agregação pode influenciar os resultados acadêmicos individuais dos estudantes, desconsiderando suas próprias características socioeconômicas, ou seja, seu próprio status socioeconômico.

O segundo problema é o erro na estimação do erro padrão utilizado nos testes de hipóteses. Isso ocorre quando, ao utilizar dados hierarquizados, considera-se a independência entre os casos individuais quando de fato isso não ocorre. Por exemplo: resultados de estudantes na mesma escola dividem, no mínimo, alguma dependência com seus colegas de sala. Essa dependência ocorre porque os estudantes

---

<sup>7</sup> Veja Haudenbush e Bryk (2002).

da mesma escola partilharem experiências educacionais juntos (efeito dos pares), e os estudantes são alocados nas escolas de formas não aleatórias. A principal forma de alocação é a localização ou proximidade da residência e, quando particular, os preços das mensalidades.

O terceiro problema refere-se à heterogeneidade das inclinações, ou seja, na aferição dos efeitos parciais. Como na relação entre a performance dos estudantes e suas características socioeconômicas que podem variar de escola a escola ou pode ser função de um grupo de variáveis de um certo nível. Um exemplo é a diferença entre resultados de estudantes de diferentes grupos raciais varia consideravelmente entre escolas, cuja explicação se baseia no fato do coeficiente dessa regressão depender da composição racial dessas escolas.

Entretanto, a vantagem do MLH é que ele combina a flexibilidade das estimações de equações separadas para cada subgrupo com o fato das variações desses coeficientes entre os subgrupos poderem ser estimados em um outro nível do modelo.

As seguintes equações representam o modelo hierárquico geral para dois níveis:

Nível 1

$$Y = \beta_{0j} + \sum_{q=1}^Q \beta_{qj} X_{qij} + v_{ij} \quad v_{ij} \sim N(0, \tau^2) \quad (4)$$

Nível 2

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + \eta_{qj} \quad \eta_{qj} \sim N(0, \tau_{qq}) \quad (5)$$

De modo que  $i$  indexa os alunos;  $j$  indexa as escolas;  $q$  indexa os  $Q$  coeficientes aleatórios do nível 1;  $s$  indexa as  $S_q$  variáveis explicativas;  $X_{qij}$  é a matriz de variáveis a nível dos alunos;  $W_{sj}$  é a matriz de variáveis a nível das escolas. Para qualquer par de coeficientes aleatórios, a covariância entre os erros do nível 2 é  $\tau_{qq}$ . Supõe-se que os erros do nível 1 e 2 são independentes,  $\text{cov}(v_{ij}, \eta_{qj}) = 0, \forall q, i, j$ .

Note que ao combinar os dois níveis do modelo não se encontra um típico modelo linear assumido na metodologia dos Mínimos Quadrados Ordinários (MQO) padrão. Isso ocorre principalmente por causa do distúrbio aleatório ser uma combinação dos erros dos dois níveis. As propriedades desejáveis requeridas para os estimadores através da aplicação do MQO é que os erros aleatórios sejam independentes, normalmente distribuídos e tenham variância constante. Porém, a complexidade do erro dos MLH torna os distúrbios aleatórios dependentes, o que confronta com as propriedades derivadas do MQO.

Devido à estrutura do modelo impossibilitar a estimação por MQO, estima-se o modelo através de um algoritmo baseado na maximização da função de verossimilhança conjunta dos dois níveis, com os parâmetros sendo estimados a partir do inverso da matriz de informação.

O processo iterativo do algoritmo inicia-se gerando razoáveis valores iniciais para todos os parâmetros, normalmente as estimativas do MQO sobre um único nível. Depois de uma interação reestima-se por MQG. Quando o processo de interações converge obtêm-se o estimador de Máxima Verossimilhança (MV). Uma síntese da técnica de estimação é dada no apêndice D.

Os resultados do MLH, na avaliação da proficiência estudantil, serão construídos a partir de um modelo mais simples para em seguida se montar um modelo completo, no qual são identificadas as características da administração e do diretor capazes de tornar a escola brasileira mais eficaz e equitativa. Tal modelo pode ser representado pela seguinte equação:

$$Y = \beta_{0j} + v_{ij} \quad v_{ij} \sim N(0, \tau^2) \quad (6)$$

Percebe-se neste modelo simples do MLH que ele é equivalente ao ANOVA com efeitos aleatórios, em que a proficiência do aluno é representada pela soma do desempenho médio da escola<sup>8</sup>,  $\beta_{0j}$ ; e de um resíduo,  $v_{ij}$ , que depende do aluno e da escola. Ou seja,  $\beta_{0j}$  no modelo do nível 1 é considerado zero para todo  $j$ . É um pressuposto do modelo que esse resíduo siga uma distribuição normal com média zero e variância constante  $\tau^2$ .

No nível da escola, representa-se a proficiência média da escola em Matemática,  $\beta_{0j}$ , como função da média geral de todas as escolas, ( $\gamma_{0j}$ ), mais um componente aleatório,  $\eta_{0j}$ . A equação de nível 2 é expressa como se segue:

$$\beta_{0j} = \gamma_{0j} + \eta_{0j} \quad \eta_{0j} \sim N(0, \tau_{00}) \quad (7)$$

Sendo que o erro  $\eta_{ij}$  segue o pressuposto do modelo acima.

O modelo ANOVA é freqüentemente utilizado como um passo preliminar na análise de dados hierarquizados. Ele produz uma estimação pontual e um intervalo de confiança para a média dos resultados da escola,  $\gamma_{0j}$ , além de prover informações sobre a variabilidade dos resultados em cada um

---

<sup>8</sup> Este intercepto representa a proficiência média dos estudantes uma vez que o método de estimação considera o desvio das variáveis em relação à média, maiores esclarecimentos em Raudenbush e Bryk (2002).

dos dois grupos. O parâmetro  $\tau^2$  representa a variabilidade dentro do grupo e  $\tau_{00}$  captura a variabilidade entre grupos (Raudenbush & Bryk, 2002).

O modelo a seguir introduz as características individuais dos alunos que têm relação com o desempenho no teste de proficiência em matemática. Inicialmente são incluídas seis variáveis de controle ao nível dos alunos, quais sejam: gênero (GE); raça (RÇ); bolsa escola (BE); repetência (RE); status socioeconômico (SS). A inclusão das variáveis gênero e raça ao nível dos alunos tem por objetivo controlar possíveis efeitos do sexo e da etnia sobre o desempenho, contribuindo, assim, com resultados escassos na literatura, e provendo aos legisladores suscetíveis para implementação de políticas discriminárias na legislação educacional brasileira. O modelo do nível 1 é representado pela equação abaixo:

$$Y = \beta_{0j} + \beta_{1j}(GE) + \beta_{2j}(RÇ) + \beta_{3j}(ID) + \beta_{4j}(BE) + \beta_{5j}(RE) + \beta_{6j}(SS) + v_{ij} \quad (8)$$

O modelo de nível 2 incorpora agora o efeito do status socioeconômico médio dos alunos da escola (SSMÉDIO), sobre o desempenho médio da escola,  $\beta_{0j}$ , e sobre o status socioeconômico dos alunos,  $\beta_{6j}$ . A equação de nível 2 pode ser assim expressa:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}(SSMEDIO)_{.j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} \\ \beta_{3j} &= \gamma_{30} \\ \beta_{4j} &= \gamma_{40} \\ \beta_{5j} &= \gamma_{50} \\ \beta_{6j} &= \gamma_{60} + \gamma_{61}(SSMEDIO)_{.j} + u_{6j} \end{aligned} \quad (9)$$

Espera-se que um percentual significativo da variância total dos resultados no teste seja entre médias escolares, que confirmará a assimetria entre as escolas brasileiras em termos de preparo para um exame de proficiência. Como Albernaz, Ferreira e Franco (2002) citaram que ocorre um fenômeno de estratificação ("*sorting*") dos alunos por critérios econômicos, o que reforça o impacto positivo das condições socioeconômicas. Não obstante, normalmente verifica-se, ainda que a variância seja controlada pelo o SS, ela permanece estatística e economicamente significativa. Dessa forma, incluir-se-á no modelo anterior as variáveis do nível das escolas referentes à localização e competência, média de idade da turma, características do diretor, da administração da escola e de problemas eventuais e estruturais para testar o poder explicativo destas sobre o desempenho dos estudantes. Assim, o modelo

de nível 1 continua a ser representado pela equação (8), enquanto o modelo de nível 2 passa a ser representado pela equação (6):

$$\begin{aligned}
 \beta_{0j} = & \gamma_{00} + \gamma_{01}(SSMEDIO)_{.j} + \gamma_{02}(rede) + \gamma_{03}(região) + \gamma_{04}(med\_ID) + \\
 & \gamma_{05}(prom\_ativ) + \gamma_{06}(part\_ativ) + \gamma_{07}(conse\_esc) + \gamma_{08}(proj\_pedag) + \\
 & \gamma_{09}(c\_ad\_alun) + \gamma_{10}(c\_for\_turm) + \gamma_{11}(vinc\_estav) + \gamma_{12}(GE\_dir) + \\
 & \gamma_{13}(ID\_dir) + \gamma_{14}(sal\_bru) + \gamma_{15}(outr\_ativ) + \gamma_{16}(esc\_max) + \\
 & \gamma_{17}(pos\_grad) + \gamma_{18}(exp\_direç) + \gamma_{19}(exp\_educ) + \gamma_{20}(inf\_ra) + \gamma_{21}(recu\_finan) \\
 & + \gamma_{22}(sem\_prof) + \gamma_{23}(interr\_ativ) + \gamma_{24}(falta\_prof) + \gamma_{25}(falta\_alu) + \\
 & \gamma_{26}(net) + \gamma_{27}(bib) + \gamma_{28}(med\_ID\_pro) + \gamma_{29}(med\_esc\_pro) + \gamma_{30}(med\_exp\_pro) + \\
 & \gamma_{31}(med\_sal\_bru\_pro) + u_{0j} \\
 \beta_{1j} = & \gamma_{10} \\
 \beta_{2j} = & \gamma_{20} \\
 \beta_{3j} = & \gamma_{30} \\
 \beta_{4j} = & \gamma_{40} \\
 \beta_{5j} = & \gamma_{50} \\
 \beta_{6j} = & \gamma_{60} + \gamma_{61} * (SSmédio)_{.j} + u_{6j}
 \end{aligned} \tag{10}$$

Espera-se que os resultados convirjam, em sua maioria, com os da metodologia que considera os dados em painel. Entretanto, sabe-se que os modelos utilizam amostras diferentes, impossibilitando uma comparação metodológica concisa.

## 5 Resultados

São apresentados separadamente os resultados dos dois modelos nas seções seguintes. No modelo com dados longitudinais foi necessário expor alguns resultados de testes que comprovassem a escolha do método de estimação.

### 5.1 Resultado do Modelo Com Dados em Painel

A tabela 5 apresenta os resultados do teste de Breusch e Pagan, para verificar se ocorre heterogeneidade não observada, o teste de Hausman, para verificar se ocorrem efeitos fixos ou aleatórios, e a estatística  $d$  de Durbin-Watson, para verificar a possível ocorrência de autocorrelação.

**Tabela 5: Resultado dos Testes.**

Teste de Breusch e Pagan	176.51
Teste de Hausman	-256,70
Teste de Durbin-Watson	0,4878

O teste de Breusch e Pagan confirma a necessidade da utilização de dados em painel, pois esse afirma a presença de um efeito não observado. Ainda, pode-se concluir sobre a hipótese assintótica do teste de Hausman que ocorre efeitos aleatórios<sup>9</sup>. Ainda, verificou-se que nas estimações por efeito fixo os resultados divergem daqueles resultados da literatura já citada. Um exemplo é a não significância de variáveis ligadas à estrutura da escola - acesso à internet e existência de biblioteca. Mas, para uma simples comparação, as estimações considerando efeito fixo estão na tabela A.1, Apêndice A. Entretanto, a hipótese de efeito fixo já deveria ser questionada, pois possivelmente variáveis não observadas em cada escola que influenciam a proficiência, como o empenho e a habilidade do diretor, não são correlacionadas com as variáveis explicativas. Logo, os resultados analisados relativo à relação entre proficiência e gestão serão os da estimativa considerando efeito aleatório, reportados na tabela 6, 7, 8 e 9.

Pelo teste de Durbin-Watson verificou-se a presença de correlação positiva nos dados, logo, o modelo empírico foi estimado pelo método dos Mínimos Quadrados Generalizados factíveis (MQGF) para efeitos aleatórios, o qual considera a matriz de covariância de White para corrigir a correlação.

Na tabela 6 apresentamos os resultados do modelo mais simples (equação 2), com somente as variáveis relativas à localização e rede de ensino – se a escola é pública ou privada (dummy igual a 1 para a escola privada) e se a escola está localizada numa região mais desenvolvida do país (região Sul e

<sup>9</sup> A estimação por Efeitos Aleatórios considera as hipóteses: exogeneidade estrita, ortogonalidade entre  $\mu_i$  e  $x$  e média nula e homocedasticidade de  $\mu_i$  (Wooldridge, 2002).

Sudeste)<sup>10</sup>. Os resultados mostram-se extremamente significativos, ou seja, as escolas privadas e em regiões mais desenvolvidas têm resultados de proficiência média dos estudantes melhores que as escolas públicas e nas regiões menos desenvolvidas (Norte, Nordeste e Centro-Oeste). Esse resultado deve ocorrer devido a melhor estrutura propiciada ou também aos professores e os administradores serem mais bem remunerados.

**Tabela 6:** Modelo mais simples, somente a matriz localização e competência.

Variáveis Explicativas	Coefficiente	Erro Padrão	Z	P >  z
Rede	58.290	1.629	35.780	0.000
Região	1.442	0.384	3.760	0.000
Constante	232.809	1.095	212.700	0.000

Ao analisar as estimações do modelo que introduz a matriz das variáveis relativas à administração e a idade média dos estudantes, tabela 7, a variável idade média dos estudantes mostra-se um efeito negativo sobre a proficiência média dos estudantes de cada escola, indicando que escolas que têm estudantes em média mais velhos tendem a reduzir o escore da proficiência, ou possivelmente essas escolas têm altas taxas de reprovação ou evasão. A variável promoção de atividade continuada (entende-se como as atividades de formação continuadas as atualizações, treinamentos e capacitações promovidas pelo diretor) e a participação de atividade continuada pelo diretor mostram-se não serem significantes, ou seja, não têm efeito sobre a média do desempenho cognitivo dos alunos. A existência de conselho escolar e a de projeto pedagógico mostraram ter efeito nulo sobre o desempenho. A variável critério de admissão dos alunos (relativo aos critérios de admissão dada a escassez de vagas em cada escola, como: sorteio, local de moradia e por prova de seleção) mostrou-se um efeito negativo e significativo, isso deve ocorrer devido às escolas públicas normalmente utilizarem esses métodos de seleção para alocar os estudantes não priorizando, talvez, os mais capacitados. A variável que denota se as escolas possuem critérios para formação de turmas (relativos à homogeneidade quanto à idade ou quanto ao rendimento escolar) não tem efeito sobre o desempenho médio dos estudantes, ou seja, turmas seletivas podem ter efeitos individuais, mas não para a coletividade. Ou seja, a prática das escolas particulares criarem “turmas especiais” pode ser considerado mais uma questão de marketing do que uma medida de efetivação do aprendizado. Por outro lado, a variável vínculo estável (avalia a proporção dos professores com vínculo estável) apresenta com elevada significância efeito positivo sobre a média de proficiência das escolas. Isso deve ocorrer devido a escolas que não têm uma estrutura adequada para o aprendizado também não dão condições trabalhistas aos seus professores. As variáveis relativas à rede e região mantiveram seus efeitos e o alto poder de explicação sobre o desempenho, tal qual observou-se no modelo anterior (tabela 6).

<sup>10</sup> Considerou-se o IDH para classificar o desenvolvimento das regiões. Os IDH's de 2000 das regiões sul, sudeste, centro-oeste, norte e nordeste são, respectivamente 0,80, 0,78, 0,75, 0,71 e 0,67 .

**Tabela 7:** Modelo com as variáveis relativas à Administração e a Idade Média dos Estudantes

Variáveis Explicativas	Coefficiente	Erro Padrão	Z	P >  z
Rede	53.947	1.786	30.210	0.000
Região	2.548	0.421	6.050	0.000
Idade Média dos Estudantes	-4.978	0.601	-8.280	0.000
Promoção de atividade continuada	1.514	1.032	1.470	0.143
Participação de atividade continuada	1.258	1.595	0.790	0.430
Conselho escolar	-0.587	1.793	-0.330	0.744
Projeto pedagógico	-1.674	1.193	-1.400	0.160
Critério de admissão dos alunos	-3.085	1.019	-3.030	0.002
Critério de formação de turmas	0.481	0.555	0.870	0.386
Vínculo Estável	1.075	0.381	2.820	0.005
Constante	308.345	9.714	31.740	0.000

Pela tabela 8 temos os resultados do modelo considerando as variáveis relativas às dificuldades enfrentadas pelo gestor, isto é, relativas a problemas eventuais e a estrutura. A variável que verifica se a infra-estrutura influencia a aprendizagem mostrou-se não significativa. Como essa variável é muito ampla, pois a infra-estrutura é tanto física quanto pedagógica, torna-se difícil verificar a relação com o desempenho dos alunos. As escolas com insuficiência de recursos financeiros têm efeitos negativos e significativos sobre o desempenho cognitivo dos alunos, reafirmando que a questão financeira é um fator de significativa importância para as escolas brasileiras. As variáveis relativas à inexistência de professores para certas disciplinas ou séries e a ocorrência de interrupção de atividades escolares mostram-se não ter efeito sobre a proficiência média dos estudantes. O efeito negativo do alto índice de falta por parte dos professores ainda não é significativo. Mas à variável alto índice de faltas por parte dos alunos verifica-se resultados negativos, ou seja, altas taxas de falta são prejudiciais ao desempenho cognitivo dos alunos tal como corroborado pelo efeito da variável que exprime a insuficiência de recursos financeiros. A existência de computadores com acesso à Internet têm efeito positivo sobre a proficiência média dos estudantes, consolidando a Internet como um meio para o melhor desenvolvimento educacional dos alunos, pois escolas com melhores recursos financeiros têm a capacidade de prover acesso a Internet para os estudantes, tal como pelo efeito da variável que exprime a interferência de recursos financeiros. A existência de biblioteca não se mostra significativa no desempenho médio dos estudantes, sinalizando que não basta as escolas possuírem o espaço tido como tal se ele não é bem equipado e incentivada sua utilização. Esses resultados indicam que a condição financeira da escola é o fator preponderante nos resultados obtidos pelas escolas. Os resultados dos modelos mais simples prevalecem, ou seja, a rede privada mantém um hiato significativo sobre a escola pública quanto ao desempenho dos alunos.

**Tabela 8:** Estimaco do Modelo com variveis relativas s dificuldades enfrentadas pelo Gestor (relativas a problemas eventuais e estruturais).

Variveis Explicativas	Coefficiente	Erro Padro	Z	P >  z
Rede	48.871	1.948	25.080	0.000
Regio	2.638	0.424	6.220	0.000
Idade Mdia dos Estudantes	-5.250	0.610	-8.610	0.000
Promoo de atividade continuada	1.063	1.039	1.020	0.306
Participao de atividade continuada	1.074	1.594	0.670	0.500
Conselho escolar	-0.247	1.736	-0.140	0.887
Projeto pedaggico	-1.623	1.193	-1.360	0.174
Crterio de admisso dos alunos	-2.682	1.029	-2.610	0.009
Crterio de formao de turmas	0.383	0.553	0.690	0.488
Vnculo estvel	1.051	0.380	2.770	0.006
A infra-estrutura dificulta a aprendizagem	1.640	1.322	1.240	0.215
Insuficincia de recursos Financeiros	-3.748	1.135	-3.300	0.001
Inexistncia de Professores para certas disciplinas ou sries	1.310	1.197	1.090	0.274
Interrupo de atividades escolares	1.185	1.404	0.840	0.398
Alto ndice de falta por parte dos professores	-2.052	1.314	-1.560	0.118
Alto ndice de falta por parte dos alunos	-3.199	1.240	-2.580	0.010
Acesso  Internet	6.306	1.272	4.960	0.000
Biblioteca	2.316	1.613	1.440	0.151
Constante	310.396	10.062	30.850	0.000

Na tabela 9 temos os resultados do modelo que introduz as variveis relativas ao diretor. As variveis que se apresentaram significantes nos outros modelos, como rede, regio, idade mdia dos estudantes, crterio de admisso de alunos, vnculo estvel, insuficincia de recursos financeiros, alto ndice de falta por parte dos alunos e acesso  Internet permaneceram com elevada significncia, bem como seus efeitos sobre o desempenho. A varivel alto ndice de faltas por parte dos professores tornou-se significativa a 9,1%, mostrando-se ter um efeito negativo sobre os resultados da escola e seguindo o mesmo sentido da varivel alto ndice por parte dos alunos, sinalizando escolas que incentivam ou motivam a presena dos estudantes e professores obtm resultados melhores. O sexo e a idade dos diretores mostram-se irrelevantes sobre a proficincia dos estudantes, pois seus coeficientes no so estatisticamente significantes. A varivel salrio bruto do diretor denota ser extremamente relevante para explicar a proficincia mdia dos estudantes, refletindo novamente que as escolas que tm melhores condioes financeiras e, conseqentemente, tm diretores mais bem remunerados obtm resultados melhores. Por outro lado, as variveis outra atividade do diretor, escolaridade mxima do diretor e se o diretor j fez um ps-graduao no so significantes. Esses resultados assemelhasse com a no

significância da prática de atividades continuadas por professores e diretores, ou seja, atividades educacionais direcionadas aos educadores, durante o período letivo, não geram resultados na performance dos estudantes. Ainda, a variável experiência em educação é significativa, sinalizando que diretores que têm maior experiência no cotidiano escolar é que estão mais qualificados a melhorar o ambiente de aprendizado dos alunos.

**Tabela 9:** Resultados do modelo completo, com todas as variáveis do modelo.

Variáveis Explicativas	Coefficiente	Erro Padrão	Z	P >  z
Rede	48.041	1.990	24.140	0.000
Região	2.603	0.434	6.000	0.000
Idade Média dos Estudantes	-4.955	0.614	-8.070	0.000
Promoção de atividade continuada	0.572	1.044	0.550	0.584
Participação de atividade continuada	0.713	1.600	0.450	0.656
Conselho escolar	-0.635	1.723	-0.370	0.713
Projeto pedagógico	-1.693	1.196	-1.420	0.157
Critério de admissão dos alunos	-2.306	1.041	-2.220	0.027
Critério de formação de turmas	0.429	0.551	0.780	0.436
Vínculo estável	0.851	0.381	2.240	0.025
A infra-estrutura dificulta a aprendizagem	1.250	1.318	0.950	0.343
Insuficiência de recursos Financeiros	-3.547	1.134	-3.130	0.002
Inexistência de Professores para certas disciplinas ou séries	1.274	1.191	1.070	0.285
Interrupção de atividades escolares	1.144	1.396	0.820	0.412
Alto índice de falta por parte dos professores	-2.210	1.308	-1.690	0.091
Alto índice de falta por parte dos alunos	-3.311	1.235	-2.680	0.007
Acesso à Internet	5.635	1.271	4.430	0.000
Biblioteca	2.050	1.604	1.280	0.201
Sexo do diretor	-0.885	1.242	-0.710	0.476
Idade do diretor	-0.282	0.400	-0.710	0.480
Salário bruto do diretor	1.402	0.351	3.990	0.000
Outra atividade do diretor	1.624	1.142	1.420	0.155
Máxima escolaridade do diretor	1.422	1.829	0.780	0.437
Pós-graduação do diretor	0.287	0.412	0.700	0.486
Experiência em direção	0.494	0.472	1.050	0.295
Experiência em educação	1.267	0.617	2.050	0.040
Constante	293.574	11.905	24.660	0.000

Esses resultados apontam para o que já vinha se confirmando; os recursos financeiros aplicados em salários mais altos para diretores experientes, com vínculo estável, e em uma estrutura que incentiva a presença dos estudantes (como computadores com acesso à Internet) são fatores que de fato são relevantes para o desempenho dos estudantes. Ainda, escolas públicas, que praticam políticas educacionais que distinguem os estudantes que serão admitidos, são relacionadas com resultados desfavoráveis.

Verificada a significância da variável rede, sinalizando que as escolas públicas obtêm resultados piores, espera-se que diferentes resultados ocorram ao estimar o modelo completo considerando somente as escolas públicas e, consecutivamente, somente as escolas privadas. A tabela 10 apresenta os resultados da estimação do modelo considerando somente as escolas públicas e a tabela 11 considera o modelo utilizando somente as escolas privadas.

**Tabela 10:** Resultados do modelo completo considerando somente escolas públicas.

Variáveis Explicativas	Coefficiente	Erro Padrão	Z	P >  z
Região	2.717	0.476	5.714	0.000
Idade Média dos Estudantes	-6.465	0.665	-9.727	0.000
Promoção de atividade continuada	0.456	1.117	0.408	0.683
Participação de atividade continuada	1.744	1.780	0.980	0.327
Conselho escolar	-1.719	1.931	-0.890	0.374
Projeto pedagógico	2.023	1.481	1.366	0.172
Critério de admissão dos alunos	-5.102	1.298	-3.930	0.000
Critério de formação de turmas	0.715	0.601	1.189	0.235
Vínculo estável	0.756	0.474	1.596	0.111
A infra-estrutura dificulta a aprendizagem	1.090	1.298	0.839	0.401
Insuficiência de recursos Financeiros	-1.767	1.250	-1.414	0.158
Inexistência de Professores para certas disciplinas ou séries	1.252	1.145	1.093	0.275
Interrupção de atividades escolares	0.334	1.270	0.263	0.793
Alto índice de falta por parte dos professores	-0.133	1.232	-0.108	0.091
Alto índice de falta por parte dos alunos	-4.379	1.134	-3.860	0.000
Acesso à Internet	4.147	1.232	3.365	0.001
Biblioteca	2.719	1.511	1.799	0.072
Sexo do diretor	-1.598	1.300	-1.229	0.219
Idade do diretor	-0.458	0.441	-1.040	0.299
Salário bruto do diretor	1.969	0.497	3.961	0.000
Outra atividade do diretor	0.957	1.278	0.749	0.454
Máxima escolaridade do diretor	2.089	1.895	1.102	0.271
Pós-graduação do diretor	0.382	0.469	0.815	0.415
Experiência em direção	-0.032	0.511	-0.062	0.951
Experiência em educação	0.583	0.654	0.890	0.374
Constante	316.256	12.776	24.754	0.000

Na estimação que considera somente escolas públicas 4 resultados divergiram, as variáveis referentes são: vínculo estável, insuficiência de recursos financeiros, experiência em educação do diretor e biblioteca. Analisando a variável vínculo estável, somente a significância da estimativa é que se modificou, deixando de ser significativa a 10%. Isso deve ser reflexo da seleção via concurso dos professores de escolas públicas, o que garante o vínculo estável à grande maioria dos docentes. Dessa

forma, não se pode considerar a condição de vínculo estável como variável motivadora dos professores em promover o desempenho educacional dos alunos.

Em relação a variável insuficiência de recursos financeiros, também, a significância da estimativa é que se modificou, deixando de ser significativa a 10%. Isso deve ser reflexo da estrutura da gestão das escolas públicas, que não atribui ao diretor à responsabilidade de administrar os recursos financeiros da escola.

A não significância da variável experiência em direção pode sinalizar o método de admissão normalmente utilizado pelas escolas públicas, pois somente 18% dos diretores da amostra foram admitidos por indicações técnicas. Ou seja, não consideram a produtividade e a qualificação adequada para administrar uma escola.

Por último, a existência de biblioteca confirma-se ser um fator relevante no desempenho médio dos estudantes de escolas públicas. Isso afirma a hipótese que a escola brasileira não passou da fase de investimentos básicos.

Na regressão do modelo com as características das escolas privadas, tabela 11, somente um resultados não se assemelham aos da regressão que não separa a escola pública da privada. A variável que divergiu foi a alto índice de falta por parte dos alunos. Esse resultado não era esperado, podendo ter ocorrido pela restrita amostra de escolas privadas ou sinalizando que o ambiente escolar não é o mais importante no desenvolvimento educacional de estudantes mais privilegiados economicamente.

**Tabela 11:** Resultados do modelo completo considerando somente escolas privadas.

Variáveis Explicativas	Coefficiente	Erro Padrão	Z	P >  z
Região	2.712	1.052	2.578	0.010
Idade Média dos Estudantes	-4.323	1.653	-2.614	0.009
Promoção de atividade continuada	-1.922	2.436	-0.789	0.430
Participação de atividade continuada	0.822	3.407	0.241	0.809
Conselho escolar	1.465	2.714	0.540	0.590
Projeto pedagógico	-4.018	2.470	-1.626	0.104
Critério de admissão dos alunos	-3.406	2.040	-1.670	0.096
Critério de formação de turmas	0.207	1.216	0.171	0.865
Vínculo estável	1.266	0.704	1.799	0.073
A infra-estrutura dificulta a aprendizagem	1.532	4.175	0.367	0.714
Insuficiência de recursos Financeiros	-7.184	2.400	-2.993	0.003
Inexistência de Professores para certas disciplinas ou séries	3.976	3.577	1.112	0.267
Interrupção de atividades escolares	0.924	7.818	0.118	0.906
Alto índice de falta por parte dos professores	-10.881	4.132	-2.633	0.009

Alto índice de falta por parte dos alunos	-4.576	5.048	-0.906	0.365
Acesso à Internet	11.781	3.290	3.581	0.000
Biblioteca	1.220	5.273	0.231	0.817
Sexo do diretor	1.919	2.725	0.704	0.482
Idade do diretor	-0.113	0.821	-0.138	0.891
Salário bruto do diretor	1.094	0.551	1.984	0.048
Outra atividade do diretor	2.491	2.328	1.070	0.285
Máxima escolaridade do diretor	1.559	4.318	0.361	0.718
Pós-graduação do diretor	0.356	0.834	0.427	0.669
Experiência em direção	-0.086	1.065	-0.081	0.935
Experiência em educação	3.252	1.463	2.223	0.027
Constante	323.882	28.867	11.220	0.000

## 5.2 Resultados do Modelo Linear Hierárquico

Os resultados do modelo mais simples, que regride a variável dependente sobre as médias, está exposto na tabela 12.

**Tabela 12:** Resultados do Modelo ANOVA<sup>11</sup>.

Efeitos Fixos	Coefficiente	Erro-padrão	Teste-t	Valor-p
Média das escolas em Matemática, $\hat{\gamma}_{0j}$	255.182	0.830	307.623	0.000
Efeitos Aleatórios	Variância do erro	Graus de Liberdade	$\chi^2$	Valor-p
Média da escola, $\eta_{0j}$	1308.934	2090.000	32548.13	0.000
Efeito do Aluno, $v_{ij}$	1662.715			

A tabela 12 contém dois resultados relevantes. O primeiro é que a variância estimada de  $\beta_{0j}$  em torno de  $\hat{\gamma}_{0j}$  é alta ( $\hat{\tau}_{00} = 1308.934$ ) e estatisticamente significativa. Dessa forma, a hipótese nula de nenhuma diferença entre os coeficientes de intercepto das escolas é descartada, ou seja, a eficácia das escolas varia significativamente em torno da média de todas as escolas. Em segundo lugar, a correlação

<sup>11</sup> O conceito de efeitos fixos e aleatórios segue a terminologia habitual na literatura de MLH, ou seja, efeitos fixos correspondem à estimação dos coeficientes e efeitos aleatórios a estimação das variâncias. O que é distinto da metodologia de dados em painel.

intra-classe, dada por  $\rho = \tau_{00}/(\tau_{00} + \tau^2)$ , resultou em uma estimativa de  $1308.934/(1308.934 + 1662.715) = 0.44$ . Ou seja, 44% da variância total do desempenho estudantil ocorre entre as escolas<sup>12</sup>, o que torna recomendável a utilização dos Modelos Hierárquicos Lineares na estimação da função de produção educacional brasileira.

Os resultados do modelo com as características individuais dos alunos encontram-se na Tabela 13. Os dois principais resultados desta regressão referem-se ao efeito do status socioeconômico sobre a proficiência individual do aluno e sobre a média da escola. Ou seja, o valor médio do status socioeconômico em cada escola é significativamente relacionado com o desempenho médio da escola. Mas o SSMEDIO não influencia o SS de cada aluno, confirmando a independência entre as condições socioeconômicas dos alunos.

**Tabela 13:** Os resultados do modelo com as características individuais dos alunos.

Efeitos Fixos	Coefficiente	Erro-padrão	Teste-t	Valor-p
<b>Média da escola, (<math>\beta_{0j}</math>)</b>				
Intercepto, ( $\gamma_{00}$ )	265.682	2.649	100.312	0.000
SSMEDIO, ( $\gamma_{01}$ )	35.629	0.705	50.555	0.000
Gênero, ( $\gamma_{10}$ )	-11.903	0.447	-26.605	0.000
Raça, ( $\gamma_{20}$ )	2.770	0.471	5.886	0.000
Idade, ( $\gamma_{30}$ )	-0.020	0.163	-0.125	0.901
Aluno com Bolsa escola, ( $\gamma_{40}$ )	0.894	0.953	0.938	0.349
Repetências, ( $\gamma_{50}$ )	-10.659	0.309	-34.523	0.000
<b>Inclinação do SS, (<math>\beta_{0j}</math>)</b>				
Intercepto, ( $\gamma_{60}$ )	5.059	0.349	14.483	0.000
SSMEDIO, ( $\gamma_{61}$ )	-0.115	0.464	-0.247	0.805
<b>Efeitos Aleatórios</b>	<b>Variância</b>	<b>Grau de Liberdade</b>	<b><math>\chi^2</math></b>	<b>Valor-p</b>
Média da escola, ( $u_0$ )	354.466	2085.000	10359.891	0.000
Inclinação do SS, ( $u_{6j}$ )	20.870	2085.000	2178.964	0.074
Efeito do Aluno, ( $r_{ij}$ )	1556.366			

Controlando pelo status socioeconômico dos alunos, os resultados mostram a existência de três fatores individuais com efeitos estatisticamente significativos sobre o desempenho estudantil: as repetências, o gênero e a raça. No que diz respeito às repetências, as mesmas mostram-se ter um efeito negativo sobre o desempenho dos alunos, ou seja, alunos que apresentam uma trajetória de aprendizado

<sup>12</sup> As pesquisas que analisaram os dados do SAEB de 2001 verificaram que cerca de 30% da variância total do desempenho estudantil ocorre entre as escolas.

com resultados negativos obtêm resultados desfavoráveis nos exames de proficiência, de acordo com o esperado. Este impacto negativo da repetência sobre o desempenho dos alunos confirma a importância de tentarmos controlar para o estoque de capital humano com o qual o aluno chega ao teste, de forma a não atribuímos erroneamente ao ensino recente os problemas advindos da trajetória passada.

Em termos de gênero, nota-se um desempenho superior dos estudantes do sexo masculino em Matemática, convergindo com os resultados encontrados na literatura brasileira e internacional. Ainda, os dados da tabela 13 indicam que os alunos não brancos (pardos, mulatos e negros) têm resultados inferiores àqueles observados para os alunos brancos. Este efeito da raça sobre o rendimento escolar, mesmo após o controle pelo status socioeconômico, constitui um resultado preocupante para aqueles interessados em reduzir a desigualdade de oportunidades no Brasil. Como Albernaz, Ferreira e Franco (2002) afirmam, o negro brasileiro parece não só ter menos chance de estar na escola, mas, além disso, os que chegam à escola, e aí conseguem permanecer, parecem ter um desempenho pior do que seus colegas brancos, mesmo controlando pelo status socioeconômico. Pimentel e Zucchi (2006) relatam que a probabilidade de um indivíduo estudar é sistematicamente menor para indivíduos de cor negra do que para indivíduos de cor branca. Isto mostra que a discriminação contra o negro está presente na aquisição do insumo educação, refletindo no fato desse indivíduo não apresentar suficiente sinalização de produtividade do trabalho para ingressar no mercado de trabalho.

Sabendo-se que cerca de 44% da variância total dos resultados no teste do SAEB de 2003 era entre médias escolares, confirma que há uma assimetria entre as escolas brasileiras em termos de preparo para um exame de proficiência. Afirmando a necessidade da utilização do MLH, pois há significativas diferenças no nível das escolas que devem ser considerados no seu respectivo nível.

**Tabela 14:** Resultados do modelo com as características dos alunos e da escola.

Efeitos Fixos	Coefficiente	Erro-padrão	Teste-t	Valor-p
Média da escola ( $\beta_{0j}$ )				
Intercepto ( $\gamma_{01}$ )	260,979	6,411	40,708	0,000
Rede ( $\gamma_{02}$ )	23,418	1,945	12,043	0,000
Região ( $\gamma_{03}$ )	2,122	0,379	5,604	0,000
Idade Média dos Estudantes ( $\gamma_{04}$ )	-0,914	0,822	-1,112	0,267
SSMEDIO ( $\gamma_{05}$ )	19,812	1,205	16,441	0,000
Promoção de atividade continuada ( $\gamma_{06}$ )	-0,248	0,901	-0,276	0,783
Participação de atividade continuada ( $\gamma_{07}$ )	2,469	1,444	1,711	0,087
Conselho escolar ( $\gamma_{08}$ )	-2,259	1,100	-2,053	0,040
Projeto pedagógico ( $\gamma_{09}$ )	-1,269	1,245	-1,020	0,309
Critério de admissão dos alunos	-3,516	0,836	-4,207	0,000

(Y <sub>10</sub> )				
Critério de formação de turmas(Y <sub>11</sub> )	0,449	0,463	0,971	0,332
Vínculo estável (Y <sub>12</sub> )	0,361	0,291	1,240	0,216
Sexo do diretor (Y <sub>13</sub> )	-1,035	1,015	-1,019	0,309
Idade do diretor (Y <sub>14</sub> )	0,232	0,329	0,704	0,481
Salário bruto do diretor (Y <sub>15</sub> )	-0,270	0,347	-0,779	0,436
Outra atividade do diretor (Y <sub>16</sub> )	0,297	0,931	0,320	0,749
Máxima escolaridade do diretor (Y <sub>17</sub> )	-4,398	1,771	-2,483	0,013
Pós-graduação do diretor (Y <sub>18</sub> )	0,138	0,473	0,292	0,770
Experiência em direção (Y <sub>19</sub> )	0,698	0,421	1,659	0,097
Experiência em educação (Y <sub>20</sub> )	-0,033	0,483	-0,069	0,945
A infra-estrutura dificulta a aprendizagem (Y <sub>21</sub> )	0,579	0,982	0,590	0,555
Insuficiência de Recursos Financeiros (Y <sub>21</sub> )	-1,701	0,973	-1,749	0,080
Inexistência de Professores para certas disciplinas ou séries (Y <sub>22</sub> )	-1,016	0,927	-1,096	0,274
Interrupção de atividades escolares (Y <sub>23</sub> )	0,917	1,200	0,765	0,445
Alto índice de falta por parte dos professores (Y <sub>24</sub> )	-2,245	0,961	-2,336	0,020
Alto índice de falta por parte dos alunos (Y <sub>25</sub> )	-1,942	0,946	-2,053	0,040
Acesso à Internet (Y <sub>26</sub> )	2,220	0,955	2,325	0,020
Biblioteca (Y <sub>27</sub> )	1,902	1,278	1,489	0,137
Média de Idade dos Professores (Y <sub>28</sub> )	-0,438	0,551	-0,795	0,427
Média de Escolaridade dos Professores (Y <sub>29</sub> )	-0,836	1,313	-0,637	0,524
Média de Experiência dos Professores (Y <sub>30</sub> )	0,790	0,404	1,954	0,050
Salário Bruto Médio dos Professores (Y <sub>31</sub> )	2,528	0,434	5,823	0,000
Sexo, (Y <sub>10</sub> )	-11,847	0,447	-26,513	0,000
Raça, (Y <sub>20</sub> )	2,686	0,469	5,731	0,000
Idade, (Y <sub>30</sub> )	-0,159	0,166	-0,957	0,339
Aluno com Bolsa escola, (Y <sub>40</sub> )	0,332	0,942	0,352	0,724
Repetências, (Y <sub>50</sub> )	-10,489	0,308	-34,079	0,000
Inclinação do SS, ( $\beta_{0j}$ )				
Intercepto, (Y <sub>60</sub> )	5,076	0,348	14,592	0,000
SSMEDIO, (Y <sub>61</sub> )	-0,117	0,464	-0,253	0,800
Efeitos Aleatórios	Variância	Grau de Liberdade	$\chi^2$	Valor-p
Média da escola, (u <sub>0</sub> )	263.602	2055.000	8115.896	0,000
Inclinação do SS, (u <sub>6j</sub> )	21.145	2085.000	2179.121	0,074

Os resultados da estimação do modelo mais completo encontram-se na Tabela 14. A rede de ensino e a região permanecem sendo fatores significativos na média das performances das escolas, ou seja, alunos de escolas privadas em regiões mais desenvolvidas, em média, obtêm resultados melhores. Resultados divergentes do modelo com dados em painel são observados na variável idade média dos estudantes, mesmo que o sinal tenha sido o esperado. A participação de atividades continuadas e a escolaridade máxima do diretor são agora fatores determinantes no desempenho dos alunos. O que pode sinalizar que administradores mais bem qualificados desenvolvem mecanismos para o aprendizado dos alunos. Promoção de atividades continuadas e a existência de projeto pedagógico confirmam-se enquanto atividades que não são relacionadas com o desempenho cognitivo dos alunos. A variável critério de admissão dos alunos permanece causando um efeito negativo sobre a performance dos estudantes, confirmando que as escolas públicas que utilizam esses critérios são relacionadas a baixos desempenhos. A variável vínculo estável e salário bruto do diretor não são significantes, isso poderia ser uma evidência que não é correta a relação entre estrutura financeira das escolas e as condições necessárias ao desenvolvimento educacional, mas a variável insuficiência de recursos financeiros afirma que a condição econômica é um fator relevante. As variáveis sexo e idade do diretor e a interrupção das atividades novamente mostram-se não significativas. Caso o diretor tenha feito pós-graduação ou exerça outra atividade, não há evidências em nenhuma das metodologias que são importantes para o desempenho da escola. Divergindo do modelo com dados em painel, no lugar da experiência em educação, a variável significativa foi a experiência em direção. A variável que considera a infra-estrutura física ou pedagógica e a existência de biblioteca permanece sendo não significativa, ao contrário da variável que considera o provimento de computadores com acesso à Internet. Entre as variáveis relativas aos professores, a experiência em educação e o salário bruto médio dos professores são significativas, confirmando que educadores com experiência em educar e que são bem pagos são relacionados com resultados melhores. Os resultados das variáveis no âmbito dos estudantes não se modificaram.

A tabela 15 apresenta os resultados da estimação do modelo considerando somente as escolas públicas e a tabela 16 considera o modelo utilizando somente as escolas privadas. Essas estimações foram feitas pelo mesmo motivo do modelo de dados em painel, a variável rede foi extremamente significativa.

**Tabela 15:** Resultados do modelo com as características dos alunos e da escola pública.

Efeitos Fixos	Coefficiente	Erro-padrão	Teste-t	Valor-p
Média da escola ( $\beta_{0j}$ )				
Intercepto ( $\gamma_{01}$ )	287.272	15.894	18.074	0.000
Região ( $\gamma_{03}$ )	2.333	0.420	5.559	0.000

Idade Média dos Estudantes ( $Y_{04}$ )	-1.422	0.885	-1.606	0.108
SSMEDIO ( $Y_{05}$ )	13.497	1.305	10.340	0.000
Promoção de atividade continuada ( $Y_{06}$ )	0.000	0.911	0.000	1.000
Participação de atividade continuada ( $Y_{07}$ )	1.603	1.593	1.006	0.315
Conselho escolar ( $Y_{08}$ )	0.297	1.331	0.223	0.824
Projeto pedagógico( $Y_{09}$ )	-0.464	1.350	-0.343	0.731
Critério de admissão dos alunos ( $Y_{10}$ )	-3.653	1.012	-3.609	0.001
Critério de formação de turmas( $Y_{11}$ )	0.061	0.515	0.118	0.906
Vínculo estável ( $Y_{12}$ )	0.588	0.335	1.756	0.079
Sexo do diretor ( $Y_{13}$ )	-2.524	1.115	-2.265	0.024
Idade do diretor ( $Y_{14}$ )	0.446	0.360	1.238	0.216
Salário bruto do diretor ( $Y_{15}$ )	0.189	0.504	0.375	0.707
Outra atividade do diretor ( $Y_{16}$ )	0.562	0.987	0.569	0.569
Máxima escolaridade do diretor ( $Y_{17}$ )	-4.680	1.839	-2.545	0.011
Pós-graduação do diretor ( $Y_{18}$ )	0.024	0.555	0.044	0.965
Experiência em direção ( $Y_{19}$ )	-0.070	0.478	-0.147	0.884
Experiência em educação ( $Y_{20}$ )	-0.626	0.509	-1.230	0.219
A infra-estrutura dificulta a aprendizagem ( $Y_{21}$ )	-0.429	0.987	-0.435	0.663
Insuficiência de Recursos Financeiros ( $Y_{21}$ )	0.251	1.069	0.235	0.814
Inexistência de Professores para certas disciplinas ou séries ( $Y_{22}$ )	-1.587	0.938	-1.692	0.090
Interrupção de atividades escolares ( $Y_{23}$ )	0.017	1.182	0.014	0.989
Alto índice de falta por parte dos professores ( $Y_{24}$ )	-1.733	0.972	-1.783	0.074
Alto índice de falta por parte dos alunos ( $Y_{25}$ )	-1.914	0.965	-1.985	0.047
Acesso à Internet ( $Y_{26}$ )	1.770	0.934	1.895	0.058
Biblioteca ( $Y_{27}$ )	1.536	1.289	1.191	0.234
Média de Idade dos Professores ( $Y_{28}$ )	-0.448	0.605	-0.740	0.459
Média de Escolaridade dos Professores ( $Y_{29}$ )	0.242	1.321	0.183	0.855
Média de Experiência dos Professores	0.728	0.449	1.620	0.105

(Y <sub>30</sub> )				
Salário Bruto Médio dos Professores	1.160	0.507	2.287	0.022
(Y <sub>31</sub> )				
Sexo, (Y <sub>10</sub> )	-12.685	0.513	-24.710	0.000
Raça, (Y <sub>20</sub> )	2.601	0.533	4.885	0.000
Idade, (Y <sub>30</sub> )	-0.181	0.189	-0.959	0.338
Aluno com Bolsa escola, (Y <sub>40</sub> )	-9.044	0.318	-28.396	0.000
Repetências, (Y <sub>50</sub> )	5.578	0.395	14.138	0.000
Inclinação do SS, ( $\beta_{0j}$ )				
Intercepto, (Y <sub>60</sub> )	5.076	0.348	14.592	0.000
SSMEDIO, (Y <sub>61</sub> )	-0.117	0.464	-0.253	0.800
Efeitos Aleatórios	Variância	Grau de Liberdade	X <sup>2</sup>	Valor-p
Média da escola, (u <sub>0</sub> )	263.602	2055.000	8115.896	0.000
Inclinação do SS, (u <sub>6j</sub> )	195.956	1350.000	4775.580	0.000
Efeito do Aluno, (r <sub>ij</sub> )	1436.014			

Na estimação que considera somente escolas públicas 4 resultados divergiram, as variáveis referentes são: promoção de atividade continuada, conselho escolar, recursos financeiros e inexistência de professores para certas disciplinas ou séries. Dado que as variáveis promoção de atividades continuadas e conselho escolar não foram significativas, sinaliza que essas práticas na escola pública brasileira não são relacionadas com o desempenho dos estudantes.

A variável insuficiência de recursos financeiros novamente mostrou-se insignificante, como no modelo de dados em painel. Ou seja, esse resultado é mais uma evidência da estrutura da gestão das escolas públicas, que não atribui ao diretor a responsabilidade de administrar os recursos financeiros da escola.

Um resultado extremamente importante é o efeito negativo da variável inexistência de professores para certas disciplinas ou séries. Sinalizando que a falta de professores nas escolas públicas geram resultados indesejáveis do ponto de vista desenvolvimentista, ou seja, leva a ambiente de baixa performance estudantil. Isso reflete a evidente necessidade de educadores para o aprendizado dos alunos e que escolas privadas têm mecanismos de punir ou suprir a falta de professores.

Na estimação que considera somente as escolas privadas, veja gráfico 15, ocorreram 7 divergências entre as variáveis do modelo completo. As divergências foram: a insignificância da variável participação de atividade continuada pelo diretor, o mesmo resultado verificado na estimação com dados

em painel; a significância da variável critério de formação de turmas, afirmando que turmas seletivas é um mecanismo eficiente em melhorar o resultado, somente, de estudantes de escolas privadas; a não significância da máxima escolaridade do diretor, o mesmo resultado verificado na estimação com dados em painel; a significância da variável relativa à infra-estrutura física e pedagógica, não verificada em nenhum outra estimação; a insignificância da variável computadores com acesso à internet, também não verificada em nenhum outra estimação; a não significância da variável média de anos de estudo, já verificado na estimação com dados em painel.

**Tabela 16:** Resultados do modelo com as características dos alunos e da escola privada.

Efeitos Fixos	Coeficiente	Erro-padrão	Teste-t	Valor-p
Média da escola ( $\beta_{0j}$ )				
Intercepto ( $\gamma_{01}$ )	266.112	28.071	9.480	0.000
Região ( $\gamma_{03}$ )	1.278	0.707	1.809	0.070
Idade Média dos Estudantes ( $\gamma_{04}$ )	0.069	1.579	0.044	0.965
SSMEDIO ( $\gamma_{05}$ )	18.940	2.436	7.777	0.000
Promoção de atividade continuada ( $\gamma_{06}$ )	-1.691	2.430	-0.696	0.486
Participação de atividade continuada ( $\gamma_{07}$ )	3.474	2.939	1.182	0.238
Conselho escolar ( $\gamma_{08}$ )	-4.672	1.718	-2.720	0.007
Projeto pedagógico ( $\gamma_{09}$ )	-2.310	3.030	-0.762	0.446
Critério de admissão dos alunos ( $\gamma_{10}$ )	-3.392	1.368	-2.479	0.014
Critério de formação de turmas ( $\gamma_{11}$ )	1.622	0.913	1.776	0.076
Vínculo estável ( $\gamma_{12}$ )	0.162	0.486	0.334	0.738
Sexo do diretor ( $\gamma_{13}$ )	1.848	1.891	0.978	0.329
Idade do diretor ( $\gamma_{14}$ )	-0.404	0.632	-0.640	0.522
Salário bruto do diretor ( $\gamma_{15}$ )	-0.563	0.443	-1.271	0.204
Outra atividade do diretor ( $\gamma_{16}$ )	-0.686	1.866	-0.367	0.713
Máxima escolaridade do diretor ( $\gamma_{17}$ )	-3.385	3.689	-0.917	0.360
Pós-graduação do diretor ( $\gamma_{18}$ )	-0.283	0.831	-0.340	0.734
Experiência em direção ( $\gamma_{19}$ )	1.664	0.822	2.024	0.043
Experiência em educação ( $\gamma_{20}$ )	0.785	1.150	0.683	0.495
A infra-estrutura dificulta a aprendizagem ( $\gamma_{21}$ )	5.328	2.948	1.807	0.071
Insuficiência de Recursos Financeiros ( $\gamma_{21}$ )	-3.361	1.804	-1.863	0.062
Inexistência de Professores para	4.537	3.059	1.483	0.138

certas disciplinas ou séries ( $Y_{22}$ )				
Interrupção de atividades escolares ( $Y_{23}$ )	9.261	6.690	1.384	0.167
Alto índice de falta por parte dos professores ( $Y_{24}$ )	-5.319	3.270	-1.626	0.104
Alto índice de falta por parte dos alunos ( $Y_{25}$ )	-5.965	3.335	-1.789	0.074
Acesso à Internet ( $Y_{26}$ )	2.881	3.036	0.949	0.343
Biblioteca ( $Y_{27}$ )	8.866	4.903	1.808	0.071
Média de Idade dos Professores ( $Y_{28}$ )	0.320	1.094	0.292	0.770
Média de Escolaridade dos Professores ( $Y_{29}$ )	-2.402	3.312	-0.725	0.468
Média de Experiência dos Professores ( $Y_{30}$ )	0.555	0.824	0.673	0.501
Salário Bruto Médio dos Professores ( $Y_{31}$ )	3.758	0.740	5.077	0.000
Sexo, ( $Y_{10}$ )	-9.897	0.870	-11.379	0.000
Raça, ( $Y_{20}$ )	2.579	0.952	2.709	0.007
Idade, ( $Y_{30}$ )	-17.141	0.806	-21.266	0.000
Aluno com Bolsa escola, ( $Y_{40}$ )	5.147	0.706	7.287	0.000
Repetências, ( $Y_{50}$ )	-9.897	0.870	-11.379	0.000
Inclinação do SS, ( $\beta_{0j}$ )				
Intercepto, ( $Y_{60}$ )	5.076	0.348	14.592	0.000
SSMEDIO, ( $Y_{61}$ )	-0.117	0.464	-0.253	0.800
Efeitos Aleatórios	Variância	Grau de $\chi^2$ Liberdade		Valor-p
Média da escola, ( $u_0$ )	351.976	679.000	2689.940	0.000
Inclinação do SS, ( $u_{6j}$ )	195.956	1350.000	4775.580	0.000
Efeito do Aluno, ( $r_{ij}$ )	1436.014			

## 6 Conclusão

Através do confronto entre as metodologias de Modelos Lineares Hierárquicos (MLH) e a de Modelos Estruturais com Dados em Painel, sobre a única base de dados oficiais atualmente disponível (exames de proficiência e questionários dos SAEB 2001 e 2003), tentou-se identificar alguns determinantes da efetividade da escola do ensino fundamental. Ou seja, essa pesquisa teve o intuito de contribuir com a literatura ao incorporar evidências adicionais sobre gestão escolar, que busca melhor entender a estrutura e qualidade da educação brasileira.

Os resultados relativos às características individuais dos alunos, observados somente no MLH aplicado ao dados do SAEB 2003, convergem, em sua maior parte, com as pesquisas que utilizaram os dados do SAEB 2001. Ou seja, o resultado principal é o efeito do status socioeconômico sobre a proficiência individual do aluno e sobre a média da escola. Seguindo o que as pesquisas internacionais já afirmavam, os fatores que influenciam a equidade e a efetividade dos resultados escolares estão associados a fatores externos à escola, como as características socioeconômicas e culturais dos estudantes, além de observar-se os resultados para as variáveis repetência, gênero e raça. O efeito negativo das repetências confirma que os resultados dos exames avaliam a trajetória de aprendizado. O resultado da variável gênero é o único divergente da literatura, quando comprova-se a melhor performance dos estudantes do sexo feminino. A contradição destes resultados pode estar relacionada com o que se obteve para a variável repetência, visto que a maioria dos estudantes que repetem serem do sexo masculino, embora esse resultado incentiva a uma investigação mais específica. Por último, constata-se efeito negativo dos alunos não brancos (pardos, mulatos e negros) sobre o rendimento escolar. Esse efeito adverso da raça sobre a performance escolar ratifica o problema discricionário histórico sobre o negro brasileiro, pois, mesmo após conseguir ter acesso à educação, demonstra ter desempenho pior do que seus colegas brancos, ainda que controlando pelo nível socioeconômico. Isto mostra que a mencionada democracia racial não está presente na aquisição do insumo educação.

Entre as vinte sete variáveis ao nível da escola, somente oito divergiram os resultados entre os dois modelos. Primeiramente, os resultados ao nível da escola não relacionados às características individuais e da gestão do diretor só não convergem sobre o efeito da idade média dos estudantes. Logo, conclui-se que as escolas privadas e em regiões mais desenvolvidas têm melhores condição ao desenvolvimento educacional e, conseqüentemente, do desempenho dos estudantes. Esse resultado

deve ocorrer devido a melhor estrutura propiciada ou devido os professores e os administradores serem mais bem remunerados.

Entre as variáveis relativas a administração da escola, a variável promoção de atividades continuadas e participação de atividades continuadas pelo diretor mostram-se insignificantes somente no modelo com dados em painel, no MLH a segunda variável apresenta um efeito positivo sobre o desempenho médio dos alunos. A existência de conselho escolar diverge seus resultados entre os modelos, significativa no MLH. A existência de projeto pedagógico nas escolas mostra-se insignificante em ambos os modelos, ou seja, é uma variável que não tem relação com a performance dos estudantes. A ocorrência de critério de admissão dos alunos, como sorteio, local de moradia e por prova de seleção, mostrou-se um efeito negativo. Isso deve ocorrer devido às escolas públicas normalmente utilizarem esses métodos de seleção para alocar os estudantes. Ambos os modelos confirmam que escolas que possuem critérios de formação de turmas não geram efeitos sobre o desempenho médio dos estudantes, ou seja, turmas seletistas podem ter efeitos individuais, mas não para a coletividade. A variável vínculo estável é divergente, no modelo com dados em painel apresenta efeito positivo sobre a média de proficiência das escolas, mas no MLH é insignificante.

Entre as variáveis relativas às dificuldades enfrentadas pelo gestor, isto é, problemas eventuais e a estrutura da escola, todos os resultados convergem entre os dois modelos. Concluindo-se que a variável relativa à infra-estrutura tanto física quanto pedagógica é não significativa. As escolas com insuficiência de recursos financeiros têm efeitos negativos sobre o desempenho dos alunos, afirmando que a questão financeira é um fator de significativa importância para as escolas brasileiras. As variáveis relativas à inexistência de professores para certas disciplinas ou séries e a ocorrência de interrupção de atividades escolares mostram-se não terem efeito sobre a proficiência média dos estudantes. Observa-se o efeito negativo do alto índice de falta por parte dos professores e dos estudantes, ou seja, altas taxas de falta são prejudiciais à performance dos alunos. Sinalizando a necessidade das escolas de incentivar ou motivar a presença dos estudantes e professores para obter melhores resultados. O acesso à Internet consolida-se como um meio de desenvolvimento educacional dos alunos. Ou seja, escolas com melhores recursos financeiros, que têm a capacidade de prover acesso à Internet, obtêm melhores resultados. A existência de biblioteca não se mostra significativa no desempenho médio dos estudantes, sinalizando que não basta às escolas possuírem espaços considerados bibliotecas se eles não são bem equipados e não há incentivo à sua utilização.

Entre as variáveis relativas às características do diretor, a metade divergiu. O sexo e a idade dos diretores mostram-se não terem efeito sobre a proficiência dos estudantes em ambos os modelos. A variável salário bruto do diretor mostrou-se um efeito positivo somente no modelo com dados em painel. As variáveis outra atividade do diretor e se o diretor já fez um pós-graduação não são significativas em ambos os modelos. Esses resultados seguem o sentido da não significância da prática de atividades

continuadas por professores e diretores, ou seja, atividades educacionais direcionadas aos educadores não são significativas para os desempenho dos estudantes. Entretanto, a variável escolaridade máxima do diretor é significativa no MLH. As variáveis experiência em educação e em direção alternam suas significâncias nos dois modelos, ou seja, o fator experiência não deve ser deixado de se levar em conta ao analisar um diretor, pois a experiência no cotidiano escolar pode qualificar o educador a melhorar o ambiente de aprendizado dos alunos. Entre as variáveis relativas aos professores, a experiência em educação e o salário bruto médio dos professores são significativas, confirmando que educadores com experiência em educar e que são bem pagos estão relacionados com resultados melhores.

Esses resultados apontam que as qualidades dos educadores, professores e diretores, podem ser medidas tanto pelos seus níveis de escolaridade e suas experiências no ambiente escolar. Sinalizando também que a estrutura física e pedagógica deve ser direcionada ao incentivo a alta frequência dos alunos e professores, ou seja, ambientes modernamente equipados e bem conservados. Ou seja, a escola brasileira necessita de recursos financeiros para serem aplicados em salários mais altos para diretores e professores experientes com vínculo estável, e em uma estrutura que incentiva a presença dos estudantes (como computadores com acesso à Internet) para melhorar o desempenho cognitivos dos estudantes. Ainda, escolas públicas, que praticam políticas educacionais que distinguem os estudantes que serão admitidos, são relacionadas com resultados desfavoráveis.

### Referências Bibliográficas

ALBERNAZ, A. FERREIRA, F. e FRANCO, C. 2002. Qualidade e equidade no ensino fundamental brasileiro. **Pesquisa e Planejamento Econômico**, v.32, n.3. Rio de Janeiro, IPEA.

COLEMAN, J. S.; CAMPBELL E.; HOBSON C.; McPARTLAND J.; MOOD A.; WEINSFIELD F.; YORK R. **Equality of Educational Opportunity**. Washington D. C.: US Government Printing Office, 1966.

COLEMAN, J. S.; HOFFER, T. Public and private high schools: The impact of communities. New York: Basic Books,(1987).

ESPÓSITO, Y.L., DAVIS, C. e NUNES, M.M.R. 2000. Sistema de Avaliação do Rendimento Escolar – O modelo adotado pelo estado de São Paulo. **Revista Brasileira de Educação**, n. 13, pp.25-53.

FERNANDES, C.O. 2003. A Escolaridade em Ciclos: práticas que conformam a escola dentro de uma nova lógica - A transição para a escola do século XXI. Rio de Janeiro: PUC-Rio, (**Tese de Doutorado**).

FORQUIN, J. C. Sociologia das desigualdades de acesso à educação: principais orientações, principais resultados desde 1965. In: FORQUIN, J. C. (Org.). **Sociologia da educação: dez anos de pesquisas**. Petrópolis: Vozes, 1995, p. 19-78.

FRANCO, C. 2005. A Pesquisa sobre Característica de Escolas Eficazes no Brasil Breve Revisão dos Principais Achados e Alguns Problemas em Aberto. Texto para análise do Ministério da Educação.

FRANCO, C., SZTAJN, P., e ORTIGÃO, M.I. 2005. Mathematics Teachers, Reform and Equity: results from the Brazilian National Assessment. Global Conference on Education Research for Developing Countries. Prague, March 31 to April 2, 2005.

FRANCO, C., FERNANDES, C, SOARES, J.F., BELTRÃO, K., BARBOSA, M. E, ALVES, M. T. G. 2003. O referencial teórico na construção dos questionários contextuais do SAEB 2001. **Estudos em Avaliação Educacional**. São Paulo: , n.28, p.39 - 71, 2003.

GOLDSTEIN, H. Multilevel statistical models. London: E. Arnold, 2003.

KLEIN, R.; FONTANIVE, N. (1995). Avaliação em Larga Escala: uma proposta inovadora. Em Aberto, 29-35.

LEE, V. E. Using hierarchical linear modeling to study social contexts: the case of school effects. **Educational Psychologist**, n. 35, p. 125-141, 2000.

LEE, V., FRANCO, C. e ALBERNAZ, A. 2004. Quality and Equality in Brazilian Secondary Schools: A Multilevel Cross-National School Effects Study. Paper presented at the 2004 Annual Meeting of the American Educational Research Association, San Diego, CA.

MACHADO SOARES, T.. 2004. Influência do Professor e do Ambiente em Sala de Aula sobre a Proficiência Alcançada pelos Alunos Avaliados no Simave-2002.

MENEZES-FILHO, N. e PAZELLO, E. (2004). Does Money in Schools Matter? Evaluating the Effects of a Funding Reform on Wages and Test Scores in Brazil. PREAL (em <http://www.preal.org/FIE/pdf/FUNDEF%20BRASIL.pdf>).

NOGUEIRA, M. A. A sociologia da educação do final dos anos 60/ início dos anos 70: o nascimento do paradigma da reprodução. Em Aberto, v. 9, n. 46, p. 49-59, 1990.

RAUDENBUSH, S. W.; BRYK, A. (2002). Hierarchical Linear Models: Applications and Data Analysis

Methods (2 ed.). Thousand Oaks; London; New Dalhi: Sage Publications.

RUTTER, M. Fifteen thousand hours: secondary schools and their effects on children. Cambridge, Mass: Harvard University Press, 1979.

SOARES, J. F. (2004). Qualidade e equidade na educação básica Brasileira: A evidência do SAEB-2001. **Arquivos Analíticos de Políticas Educativas**, 12 (38). <http://epaa.asu.edu/epaa/v12n38>

SOARES, J.F.; ALVES, M.T. 2003. Desigualdades Raciais no Sistema Brasileiro de Educação Básica. **Educação e Pesquisa**, São Paulo, v. 29, p. 147-165.

## Anexos

### Apêndice A Estimação do modelo com efeitos Fixos

**Tabela A.1:** Estimação com efeitos fixos para o modelo completo.

Variáveis Explicativas	Coefficiente	Erro Padrão	Z	P >  z
Rede	9.307	14.072	0.660	0.509
Região	0.131	0.481	0.270	0.786
Idade Média dos Estudantes	-2.202	0.655	-3.360	0.001
Promoção de atividade continuada	0.501	1.150	0.440	0.663
Participação de atividade continuada	-1.042	1.777	-0.590	0.558
Conselho escolar	(dropped)			
Projeto pedagógico	-2.621	1.262	-2.080	0.038
Critério de admissão dos alunos	0.944	1.150	0.820	0.412
Critério de formação de turmas	-0.391	0.644	-0.610	0.544
Vínculo estável	1.125	0.441	2.550	0.011
A infra-estrutura dificulta a aprendizagem	0.514	1.426	0.360	0.719
Insuficiência de recursos Financeiros	-1.087	1.334	-0.810	0.415
Inexistência de Professores para certas disciplinas ou séries	0.884	1.354	0.650	0.514
Interrupção de atividades escolares	1.063	1.527	0.700	0.487
Alto índice de falta por parte dos professores	-2.317	1.508	-1.540	0.125
Alto índice de falta por parte dos alunos	-0.742	1.374	-0.540	0.589
Acesso à Internet	0.803	1.611	0.500	0.618
Biblioteca	-1.806	1.901	-0.950	0.342
Sexo do diretor	0.799	1.578	0.510	0.613
Idade do diretor	-0.305	0.482	-0.630	0.527
Salário bruto do diretor	0.661	0.508	1.300	0.194
Outra atividade do diretor.	0.481	1.366	0.350	0.725
Máxima escolaridade do diretor	-0.202	2.070	-0.100	0.922
Pós-graduação do diretor	-0.103	0.474	-0.220	0.828
Experiência em direção	0.705	0.557	1.270	0.206
Experiência em educação	0.538	0.705	0.760	0.446
Constante	282.626	13.871	20.380	0.000

## Apêndice B Descrição dos Dados

**Tabela B.1:** Análise Descritiva das Variáveis do Modelo de Dados em Painel.

	Dados de 2001				Dados de 2003			
	Méd.	Max	Min	Desv. Padr.	Méd.	Max	Min	Desv. Padr.
Média da proficiência das escolas	256.26	361.67	186.15	37.26	257.21	370.00	165.66	38.26
Dependência administrativa	1.05	2.00	0.00	0.84	1.04	2.00	0.00	0.84
Localização	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rede	0.38	1.00	0.00	0.49	0.38	1.00	0.00	0.49
Região	1.79	4.00	0.00	1.34	1.79	4.00	0.00	1.34
Média da idade dos alunos	14.93	18.50	13.33	1.08	15.12	17.52	13.93	0.52
Sexo do diretor	0.69	1.00	0.00	0.46	0.74	1.00	0.00	0.44
Idade do diretor	4.24	7.00	0.00	1.77	4.28	7.00	0.00	1.70
Salário bruto do diretor	3.78	7.00	0.00	1.88	3.42	7.00	0.00	1.62
Outra atividade do diretor	0.68	1.00	0.00	0.47	0.67	1.00	0.00	0.47
Máxima escolaridade do diretor	2.93	3.00	1.00	0.26	2.93	3.00	2.00	0.25
Pós-graduação do diretor	1.67	5.00	0.00	1.52	1.16	4.00	0.00	0.99
Experiência em direção	1.77	4.00	0.00	1.35	1.68	4.00	0.00	1.34
Experiência em educação	3.52	4.00	0.00	0.87	3.90	5.00	0.00	1.19
A infra-estrutura dificulta a aprendizagem	0.90	1.00	0.00	0.29	0.76	1.00	0.00	0.43
Insuficiência de Recursos Financeiros	0.63	1.00	0.00	0.48	0.68	1.00	0.00	0.47
Inexistência de Professores para certas disciplinas ou séries	0.36	1.00	0.00	0.48	0.35	1.00	0.00	0.48
Interrupção de atividades escolares	0.17	1.00	0.00	0.38	0.16	1.00	0.00	0.36
Alto índice de falta por parte dos professores	0.24	1.00	0.00	0.43	0.29	1.00	0.00	0.45
Alto índice de falta por parte dos alunos	0.28	1.00	0.00	0.45	0.33	1.00	0.00	0.47
Acesso à internet pelos alunos	0.56	1.00	0.00	0.50	0.58	1.00	0.00	0.49
Biblioteca	0.82	1.00	0.00	0.38	0.91	1.00	0.00	0.29
Promoção se atividade continuada	0.42	1.00	0.00	0.49	0.61	1.00	0.00	0.49
Participação de atividade continuada	0.87	1.00	0.00	0.34	0.89	1.00	0.00	0.32
Conselho escolar	0.69	1.00	0.00	0.46	0.71	1.00	0.00	0.46
Projeto pedagógico	0.85	1.00	0.00	0.36	0.85	1.00	0.00	0.36
Critério de admissão dos alunos	0.96	2.00	0.00	0.29	1.37	2.00	0.00	0.59
Critério de formação de turmas	1.06	2.00	0.00	0.96	1.01	2.00	0.00	0.97
Vínculo estável	2.02	3.00	0.00	1.13	2.48	4.00	0.00	1.53

**Tabela B.2:** Estatísticas Descritiva das Variáveis do Modelo Linear Hierárquico.

Estatísticas Descritivas do Nível 1				
Variáveis	Média	Desvio-Padrão	Mínimo	Máximo
GE	0.52	0.50	0.00	1.00
RÇ	0.47	0.50	0.00	1.00
ID	15.09	1.39	12.00	19.00
BE	0.95	0.23	0.00	1.00
RE	0.52	0.81	0.00	3.00
SS	-0.00	1.00	-3.73	2.69
PROFIC	255.92	55.21	116.32	428.24
Estatísticas Descritivas do Nível 2				
Variáveis	Média	Desvio-Padrão	Mínimo	Máximo
REDE	0.34	0.47	0.00	1.00
REGIAO	2.06	1.40	0.00	4.00
MED_ID	15.12	0.54	13.00	18.00
SSMEDIO	-0.02	0.78	-2.96	2.31
PROM_ATI	0.63	0.48	0.00	1.00
PART_ATI	0.88	0.32	0.00	1.00
CONS_ESC	0.70	0.46	0.00	1.00
PROJ_PED	0.86	0.35	0.00	1.00
C_AD_ALU	1.37	0.57	0.00	2.00
C_FOR_TU	1.06	0.97	0.00	2.00
VINC_EST	2.44	1.55	0.00	4.00
GE_DIR	0.74	0.44	0.00	1.00
ID_DIR	4.37	1.71	0.00	7.00
SAL_BRU	3.36	1.63	0.00	7.00
OUT_ATIV	0.65	0.48	0.00	1.00
MAX_ESC	2.93	0.27	0.00	3.00
POS_GRA	1.09	1.00	0.00	4.00
EXP_DIR	1.74	1.33	0.00	4.00
EXP_EDUC	3.97	1.17	0.00	5.00
INFRA	0.75	0.43	0.00	2.24
REC_FIN	0.69	0.46	0.00	1.00
SEM_PROF	0.34	0.47	0.00	1.00
INTE_ATV	0.17	0.37	0.00	1.00
FALT_PRO	0.27	0.45	0.00	1.00
FALT_ALU	0.34	0.47	0.00	1.00
NET	0.55	0.50	0.00	1.00
BIB	0.88	0.32	0.00	1.00
ME_ID_PROF	2.20	1.11	0.00	5.00
ME_ESC_PROF	1.91	0.30	0.00	2.00
ME_EXP_PROF	3.75	1.54	0.00	6.00
SAL_BR_PROF	2.86	1.26	0.00	7.00

## Apêndice C Definindo a Variável Status Socioeconômico<sup>13</sup>

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

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

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

$$X - \mu = LF + \varepsilon \quad (11)$$

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

$$E(F) = \mathbf{0}, \text{ cov}(F) = E(ff') = \mathbf{I} \quad (12)$$

$F$  e  $\varepsilon$  são independentes, logo:  $\text{cov}(F, \varepsilon) = \mathbf{0}$ . A covariância do modelo é dada por:

$\Sigma = \text{Cov}(X) = LL' + \psi$ , onde  $\psi$  é a matriz diagonal ( $p \times p$ ) dos fatores individuais. Desta forma as variâncias de cada variável podem ser decompostas em comunalidade ( $h^2$ ), devido aos fatores comuns e variância específica  $\psi$ , devida aos fatores individuais.

$$\sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \psi_i \quad (13)$$

<sup>13</sup> Essa metodologia é a mesma utilizada em Albernaz, Ferreira e Franco (2002), somente substituiu-se a variável número de telefones por número automóveis, pois não há mais essa variável nos questionários do SAEB 2003.

$$\text{Var}(X_i) = \text{comunalidades} + \text{variância específica} \quad (14)$$

$$h^2_i = I^2_{i1} + I^2_{i2} + \dots + I^2_{im} \quad (15)$$

$$\text{Var}(X_i) = h^2_i + \psi_i \quad (16)$$

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

Neste estudo, nove itens foram considerados relevantes na investigação do status socioeconômico do aluno. São elas:

Onde você mora existe:

1) água encanada?

2) eletricidade?

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

3) rádios?

4) televisões a cores?

5) geladeiras?

6) freezers?

7) máquinas de lavar roupa?

8) automóveis?

9) Até que série seu pai estudou? Até que série sua mãe estudou?

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

As respostas dos itens de nº 1 e 2 são sim ou não, codificadas para 0 se não e 1 se sim. As respostas das perguntas de nº 3 a 8 revelam a disponibilidade destes itens na casa dos alunos. As respostas podem variar entre 0 e 6. O último item considera duas perguntas presentes no questionário do aluno. A resposta deste item corresponde a maior série cursada pelo seu pai e/ou sua mãe. Cada pergunta apresenta 7 possibilidades de respostas. Estas variam de 0: nunca estudou; até 6: freqüentou a faculdade.

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

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

**Tabela C.1:** Resultado da análise fatorial.

Component	Total Variance Explained					
	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,917	32,411	32,411	2,917	32,411	32,411
2	1,398	15,528	47,939			
3	,832	9,246	57,185			
4	,792	8,804	65,988			
5	,709	7,876	73,865			
6	,672	7,472	81,336			
7	,665	7,391	88,727			
8	,581	6,461	95,188			
9	,433	4,812	100,000			

Extraction Method: Principal Component Analysis.

A análise fatorial apresentada foi realizada a partir do software *SPSS*. Selecionou-se a opção de se obter somente um fator nesta análise. A partir da tabela C.1, observa-se que a “componente principal” explica 32.41 % da variância total das 9 variáveis. Esta análise também nos fornece a tabela C.2, que mostra a carga das 9 variáveis neste fator. Estes valores correspondem às correlações de cada variável com a “componente principal”.

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

**Tabela C.2:** Matriz de Componentes Principais.

**Component Matrix** <sup>a</sup>

	Component
	1
AGUA_ENC	,364
ELETRIC	,243
RADIO	,660
TELEV	,786
GELAD	,433
FREEZER	,460
MQ_LAV_R	,609
AUTOM	,682
MAX_EDU	,661

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

De forma que

$$SS_{ij} = .364 (\text{água encanada}) + .243 (\text{eletricidade}) + .660 (\text{n}^\circ \text{rádios}) + .786 (\text{n}^\circ \text{TV a cores}) + .433 (\text{n}^\circ \text{geladeiras}) + .460 (\text{n}^\circ \text{freezer}) + .609 (\text{máq. lavar roupa}) + .682 (\text{automóveis}) + .661 (\text{maxedu}).$$

**Tabela C.3:** Matriz de correlação.

**Correlation Matrix**

	VAR00001	VAR00002	VAR00003	VAR00004	VAR00005	VAR00006	VAR00007	VAR00008	VAR00009
Correlation VAR00001	1,000	,294	,123	,170	,245	,047	,164	,102	,173
VAR00002	,294	1,000	,043	,097	,264	,006	,108	,014	,093
VAR00003	,123	,043	1,000	,508	,149	,225	,278	,385	,318
VAR00004	,170	,097	,508	1,000	,219	,292	,352	,499	,455
VAR00005	,245	,264	,149	,219	1,000	,071	,252	,124	,206
VAR00006	,047	,006	,225	,292	,071	1,000	,217	,290	,188
VAR00007	,164	,108	,278	,352	,252	,217	1,000	,309	,306
VAR00008	,102	,014	,385	,499	,124	,290	,309	1,000	,371
VAR00009	,173	,093	,318	,455	,206	,188	,306	,371	1,000

#### Apêndice D                    Algoritmo da Estimação de Máxima Verossimilhança (ML) por Maximização das Expectativas (EM)

Considere o modelo do nível 1:

$$Y_j = X_j \beta_j + r_j; \quad r_j \sim N(0, \sigma^2 I); \quad (17)$$

onde  $Y_j$  é vetor de variáveis dependentes  $n_j$  por 1,  $X_j$  é uma matriz  $n_j$  por  $Q$  de variáveis explicativas do nível 1,  $\beta_j$  é um vetor  $Q$  por 1 de coeficientes do nível 1 e  $r_j$  é um vetor  $n_j$  por 1 efeitos aleatórios assumido que é normalmente distribuído com média zero e matriz de covariância igual a  $\sigma^2 I$ , onde  $I$  é uma matriz identidade  $n_j$  por  $n_j$ . As unidades do nível 2 são independentes.

No nível 2 os coeficientes do nível 1 tornam-se um vetor de variáveis dependentes:

$$\beta_j = W_j \gamma + u_j; \quad u_j \sim N(0, T); \quad (18)$$

onde  $W_j$  é uma  $Q$  por  $f$  de variáveis explicativas do nível 2,  $\gamma$  é um vetor  $f$  por 1 de efeitos fixos e  $u_j$  é vetor  $q$  por 1 de efeitos aleatórios do nível 2. Assume-se que os efeitos aleatórios são normalmente distribuídos com média zero e matriz de covariância  $T$ .

Substituindo a equação do nível 2 na equação do nível 1 temos o modelo combinado:

$$Y_j = X_j W_j \gamma + X_j u_j + r_j. \quad (19)$$

U caso especial do modelo misturado pode ser

$$Y_j = A_{jf} \theta_f + A_{rj} \theta_{rj} + r_j; \quad (20)$$

$$\theta_{rj} \sim N(0, T);$$

$$r_j \sim N(0, \sigma^2 I);$$

onde  $A_{fj} = X_j W_j$ ,  $\theta_f = \gamma$ ,  $A_{rj} = X_j$ ,  $\theta_{rj} = u_j$ . O modelo misturado é mais geral do que o modelo combinado, pois ele não requer todos os coeficientes do nível 1 tenham o componente aleatório. Então, em muitas aplicações  $A_{rj}$  vai ser um subconjunto de  $X_j$  com efeitos aleatórios e em outras aplicações  $A_{rj}$  pode conter variáveis que não têm efeitos fixos.

Passo M

A maximização das expectativas concebe  $Y_j$  como sendo dados observados com  $\theta_{rj}$  como dados perdidos. Então, os dados completos são  $(Y_j, \theta_{rj})$ ,  $j = 1, \dots, J$ , enquanto  $\theta_f$ ,  $\sigma^2$  e  $T$  são os parâmetros a serem estimados.

Para encontra  $\theta_f$  pode-se deduzir  $A_{rj} \theta_{rj}$  de ambos os lados da equação do modelo misturado:

$$Y_j - A_{rj} \theta_{rj} = A_{jf} \theta_f + r_j \quad (21)$$

e justificando a estimação por Mínimos Quadrados Ordinários (OLS)

$$\hat{\theta}_f = (\sum A_{jf}^T A_{jf})^{-1} \sum A_{jf}^T (Y_j - A_{rj} \theta_{rj}) \quad (22)$$

como o estimador de dados completos de Máxima Verossimilhança de  $\theta_f$ .

O estimador de dados completos de Máxima Verossimilhança de  $T$  e  $\sigma^2$  são diretamente similares:

$$\hat{T} = J^{-1} \sum \theta_{rj} \theta_{rj}^T; \quad (23)$$

$$\hat{\sigma}^2 = N^{-1} \sum r_j r_j^T; \quad (24)$$

$$= N^{-1} \sum (Y_j - A_{jf} \theta_f - A_{rj} \theta_{rj})^T (Y_j - A_{jf} \theta_f - A_{rj} \theta_{rj}); \quad (25)$$

onde  $N = \sum n_j$ .

Esse raciocínio define as exatas estatísticas suficientes de dados completos (CDSS), ou seja, estatísticas que podem ser suficientes  $\theta_f$ ,  $\sigma^2$  e  $T$  se os dados completos fossem observados. Estes são:

$$\sum A_{ff}^T A_{rj} \theta_{rj}, \quad \sum \theta_{rj} \theta_{rj}^T, \quad \sum Y_j^T A_{rj} \theta_{rj}, \quad \sum \theta_{rj}^T A_{rj}^T A_{rj} \theta_{rj}. \quad (26)$$

Passo E

Os CDSS não são observados, mas eles podem ser estimados por suas expectativas condicionais, dados os dados de  $Y$  e os parâmetros nas interações anteriores. Em seu paper seminal, Dempster et al. (1977) mostrou que substituindo o esperado CDSS para as fórmulas do passo M poderia produzir novas estimações dos parâmetros com uma maior verossimilhança do que as estimações correntes.

Para encontrar  $E(\text{CDSS} | Y, \theta_f, \sigma^2, T)$  requer derivar a distribuição condicional dos dados ausentes,  $\theta_r$ , dado  $Y$ ,  $\theta_f$ ,  $\sigma^2$  e  $T$ . Da equação (4), a distribuição conjunta dos dados completos é

$$\begin{pmatrix} Y_j \\ \theta_{rj} \end{pmatrix} \sim N \left[ \begin{pmatrix} A_{ff} \theta_f \\ 0 \end{pmatrix}, \begin{pmatrix} A_{rj} T A_{rj}^T + \sigma^2 I & A_{rj} T \\ T A_{rj}^T & T \end{pmatrix} \right], \quad (27)$$

da distribuição condicional do dados ausentes chega-se aos seguintes dados completos:

$$\theta_{rj} | Y, \theta_f, T, \sigma^2 \sim N(\theta_{rj}^*, \sigma^2 C_j^{-1}), \quad (28)$$

com

$$\begin{aligned} \theta_{rj}^* &= C_j^{-1} A_{rj}^T (Y - A_{ff} \theta_f), \\ C_j &= A_{rj}^T A_{rj} + \sigma^2 T^{-1}. \end{aligned} \quad (29)$$

A prova dessa distribuição encontra-se em Raudenbush e Bryk (2002).

Juntando os dois passos

Tendo identificado o CDSS necessário para o passo M, estamos prontos para definir o algoritmo EM.

1. Estima-se o CDSS. Os resultados encontrados são

$$E(A_{ff}^T A_{rj} \theta_{rj} | Y, \theta_f, \sigma^2, T) = \sum A_{ff}^T A_{rj} \theta_{rj}^* \quad (30)$$

$$E(\theta_{ij}\theta_{ij}^T | Y, \theta_f, \sigma^2, T) = \sum \theta_{ij}^* \theta_{ij}^{*T} + \sigma^2 \sum C_j^{-1} \quad (31)$$

$$E(r_j^T r_j | Y, \theta_f, \sigma^2, T) = \sum r_j^{*T} r_j^* + \sigma^2 \text{tr} \sum C_j^{-1} A_{ij}^T A_{ij} \quad (32)$$

Onde  $r_j^* = Y_j - A_{jj}\theta_f - A_{ij}\theta_{ij}^*$ . Os valores dessas expectativas são baseadas nas estimações de  $\theta_f$ ,  $\sigma^2$  e T nas interações prévias.

2. Substituindo as estimativas do CDSS dentro das fórmulas do passo M para obter novas estimativas dos parâmetros.
3. Substitua as novas estimativas dos parâmetros dentro do passo 1.
4. Continue até: (a) as mudanças na log-verossimilhança tornassem suficientemente pequenas ou (b) a maior mudança no valor de algum dos parâmetros seja suficientemente pequena.

A convergência do algoritmo pode ser monitorada pelo calculo da log-verossimilhança em cada interação.