

MARIA ELIZETE GONÇALVES

**Análise de sobrevivência e modelos hierárquicos logísticos
longitudinais: uma aplicação à análise da trajetória escolar
(4^a a 8^a série - ensino fundamental)**

Belo Horizonte, MG
UFMG/Cedeplar
2008

MARIA ELIZETE GONÇALVES

**Análise de sobrevivência e modelos hierárquicos logísticos
longitudinais: uma aplicação à análise da trajetória escolar
(4^a a 8^a série - ensino fundamental)**

Tese apresentada ao curso de doutorado do Centro de Desenvolvimento e Planejamento Regional da Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do Título de Doutor em Demografia.

Orientador: Prof. Dr. Eduardo Luiz Gonçalves Rios-Neto
Co-orientadora: Prof^a. Dr^a. Cibele Comini César

Belo Horizonte, MG
Centro de Desenvolvimento e Planejamento Regional
Faculdade de Ciências Econômicas - UFMG
2008

FOLHA DE APROVAÇÃO

A um amor maior: JHS

À minha mãe... *in memoriam*. Não fosse seu incentivo, eu teria ficado parada no tempo... Não teria feito o mestrado. Não teria iniciado essa tese. Não a teria concluído.

AGRADECIMENTOS

Agradeço à Deus... Fonte de todo conhecimento!

Agradeço à Maria... minha Mãe Maria! Para sempre vou te amar...

Agradeço de uma forma toda especial ao meu orientador, o Eduardo Rios-Neto. À ele, minha gratidão por inserir-me na área da educação; pela oportunidade de participação nas pesquisas; por criar as condições necessárias à realização desse trabalho; pelo imenso aprendizado adquirido ao longo da elaboração dessa tese. Se eu já o admirava antes de ser meu orientador, essa admiração aumentou muito mais durante as longas horas de orientação. Obrigada pela paciência, pelo estímulo, pela confiança; enfim, pelas oportunidades a mim oferecidas!

Agradeço à minha co-orientadora Cibele, por dois importantes motivos: primeiro, por ter me incentivado a fazer o doutorado em Demografia. Obrigada, Cibele, pelas conversas que tivemos, durante um curso de Estatística, nesse sentido. Hoje vejo que melhor aconselhamento, ao longo da minha vida acadêmica, eu não poderia ter tido. E, segundo, pela orientação dada na realização desse trabalho, não só na parte metodológica, mas também na parte teórica.

Meu agradecimento a todos os professores da Demografia, que muito contribuíram para a minha formação; em especial, à Simone Wajman, minha orientadora de curso.

Aos meus colegas da coorte 2003. Em especial, àqueles que mantivemos contato permanente ao longo desses cinco anos, que se tornaram bons amigos: Hélder, Claudinha Pereira (que mesmo nos EUA contribuiu efetivamente com esse trabalho); à Radoyka, Mirinha, Marisa e Alexandar.

Aos colegas/amigos das coortes anteriores. Em especial, Mirela, Luiza, Rofilia, Gustavo, Geovane, Kenya e Laécia.

À equipe das pesquisas do Eduardo: Almada, que sempre se mostrou disponível a auxiliar no que fosse possível, e assim o fez. À Luciana Luz, pela colaboração com a base do censo de 99. À Danielle, pela grande força, pela ajuda no manuseio da base de dados e do software MLWin. Nesses contatos, tornamo-nos amigos. E ao Marcelo, seu marido, que tive o prazer de conhecer nesse período.

Ao Departamento de Economia da Unimontes, na pessoa do prof. Rocha (chefe do Depto.) e da prof^a Ilva Ruas (coordenadora do Curso), pela compreensão por meu afastamento. Agradeço aos demais colegas do Curso. Especialmente, às amigas Gilmara Emília, Luciene Rodrigues, Sara Antunes e Fátima Maia.

Agradeço também aos meus conterrâneos: Juracy Leite, Marília Borborema, família Cavalcante, família Cardoso, família Rocha, Família Villaça e Anelisa Graciele (acadêmica do Curso, pelo auxílio na tabulação de alguns dados).

À equipe da biblioteca, da Secretaria, do laboratório de informática, da copiadora. À Lucília e à Ângela (ex-secretária e secretária do Eduardo).

Às amigas de BH, cuja amizade se consolidou ao longo desses cinco anos. Em especial, Claudinha, Rê, Marlene e Beatriz Camilo. E ao meu amigo Marcelo Carvalho. É verdade que perdemos um pouco o contato nessa minha fase de correria, mas nossa amizade, por ter uma base sólida, não foi abalada em função disso.

À minha família, minha eterna gratidão: meu pai Osvaldo, meus irmãos Eldo e Omilson. Pela força, pela compreensão durante minha ausência, pelas palavras de otimismo, pelas orações. Um obrigada especial também à Tia Eré, à minha prima Neide e seu marido Abel (pela acolhida sempre calorosa na sua residência) e aos meus queridos primos, Felipe e Matheus, que amo muito. E aos demais familiares (mineiros, paulistas, etc).

À Capes e ao CNPq, pelo financiamento obtido.

SUMÁRIO

INTRODUÇÃO	01
1. INDICADORES EDUCACIONAIS NO ENSINO FUNDAMENTAL BRASILEIRO: EVIDÊNCIAS TEÓRICAS E EMPÍRICAS RECENTES	08
1.1 Indicadores educacionais no ensino fundamental brasileiro: evidências teóricas	08
1.2 Indicadores educacionais no ensino fundamental brasileiro: evidências empíricas	17
1.3 Repetência e evasão no ensino fundamental brasileiro: em busca de uma melhor compreensão sobre estes eventos associados ao fluxo escolar	26
2. ARCABOUÇO TEÓRICO: A FUNÇÃO DE PRODUÇÃO EDUCACIONAL	31
2.1 A função de produção educacional: questões relacionadas à especificação	31
2.2 Estrutura empírica da FPE: uma adaptação aos dados / variáveis existentes	34
3. DADOS: DETERMINAÇÃO DA AMOSTRA E BASES UTILIZADAS	37
3.1 Descrição da amostra	37
3.2 Base de dados	39
4. ARCABOUÇO METODOLÓGICO: ANÁLISE DE SOBREVIVÊNCIA E MODELO HIERÁRQUICO LOGÍSTICO LONGITUDINAL	43
4.1 Análise de sobrevivência	43
4.2 O modelo de regressão hierárquico logístico longitudinal	47
4.2.1 Especificação do modelo incondicional	48
4.2.2 Especificação do modelo condicional	49
4.2.2.1 Variável-resposta: primeira repetência	49
4.2.2.2 Variável-resposta: evasão	51
5. EVENTOS BÁSICOS E COVARIÁVEIS: ESTRATÉGIAS EMPÍRICAS PARA ADEQUAÇÃO ÀS TÉCNICAS DE ANÁLISE	54
5.1 Eventos de interesse: estratégias empíricas aplicadas aos dados da “Ficha B”	54

5.2	Tempo de sobrevivência aos eventos: estratégias empíricas aplicadas	60
5.3	Variáveis dos modelos de regressão: estratégias empíricas aplicadas	61
5.3.1	Variáveis incluídas na modelagem econométrica	63
6.	ANÁLISE DESCRITIVA E ANÁLISE DE SOBREVIVÊNCIA: ACOMPANHANDO A TRAJETÓRIA DA COORTE ESCOLAR	70
6.1	Trajetoória da coorte: aspectos descritivos	70
6.2	Eventos reprovação, repetência, afastamento por abandono e evasão, entre a 4ª e a 8ª série, segundo UF's: aspectos descritivos	80
6.3	Análise de sobrevivência: resultados e discussão	85
6.3.1	Curvas de sobrevivência à primeira reprovação e à primeira repetência	86
6.3.2	Curvas de sobrevivência à evasão e ao primeiro afastamento por abandono	88
6.3.3	Curvas de sobrevivência aos eventos, segundo situação de defasagem idade-série	90
6.3.3.1	Curvas de sobrevivência à primeira reprovação e à primeira repetência, segundo situação de defasagem idade-série.	91
6.3.3.2	Curvas de sobrevivência à evasão e ao primeiro afastamento por abandono, segundo situação de defasagem idade-série.	92
7.	ANÁLISE HIERÁRQUICA LOGÍSTICA LONGITUDINAL: IDENTIFICAÇÃO E DISCUSSÃO DOS DETERMINANTES DA REPETÊNCIA E DA EVASÃO ENTRE A 4ª E A 8ª SÉRIE NO ENSINO FUNDAMENTAL	95
7.1	Probabilidade de repetência entre a 4ª e 8ª série do ensino fundamental	95
7.2	Probabilidade de evasão entre a 4ª e a 8ª série do ensino fundamental	108
7.3	Repetência e evasão no ensino fundamental: um paralelo entre os resultados obtidos e algumas considerações sobre a importância do estudo longitudinal para os eventos estudados	118
	CONSIDERAÇÕES FINAIS	123
	REFERÊNCIAS BIBLIOGRÁFICAS	134
	ANEXOS	140

LISTA DE ABREVIATURAS E SIGLAS

CEDEPLAR - Centro de Desenvolvimento e Planejamento Regional

EDURURAL - Programa de expansão e melhoria da educação no meio rural do Nordeste

EJA - Escola de Jovens e Adultos

FPE - Função de Produção Educacional

FUNDESCOLA - Fundo de Fortalecimento da Escola

IBGE - Instituto Brasileiro de Geografia e Estatística

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais

MEC - Ministério de Educação e Cultura

MQO - Mínimos Quadrados Ordinários

PDE - Plano de Desenvolvimento da Escola

PNAD - Pesquisa Nacional por Amostra de Domicílios

PPS - Probabilidade de Progressão por Série

SAEB - Sistema de Avaliação para Educação Básica

UF - Unidade de Federação

UNESCO - Organização das Nações Unidas para a educação, a ciência e a cultura

LISTA DE FIGURAS

FIGURA 1: Situação dos alunos referente eventos escolares de fluxo	27
FIGURA 2: Modelo de fluxo escolar	54
FIGURA 3: Resultados (eventos) registrados na “Ficha B”	56
FIGURA 4: Fluxo escolar, segundo registros da “Ficha B”	58
FIGURA 5: Fluxo escolar, segundo registros da “Ficha B” 1999-2000	75
FIGURA 6: Fluxo escolar, segundo registros da “Ficha B” 2000-2001	76
FIGURA 7: Fluxo escolar, segundo registros da “Ficha B” 2001-2002	77
FIGURA 8: Fluxo escolar, segundo registros da “Ficha B” 2002-2003	78
FIGURA 9: Matrículas / eventos segundo registros da “Ficha B” 2003	79

LISTA DE QUADROS

QUADRO 1: Esquema da coleta de dados da pesquisa “Avaliação de Desempenho: Fatores Associados”	40
QUADRO 2: Esquema da coleta de dados da pesquisa “Ficha B”	40
QUADRO 3: Coordenadas das categorias nas duas dimensões, sistema de segurança na Escola: Ficha B, 1999	142
QUADRO 4: Coordenadas das categorias nas duas dimensões, Estrutura básica da Escola: Censo Escolar de 1999	142
QUADRO 5: Coordenadas das categorias nas duas dimensões, Nível socioeconômico Familiar: Ficha B, 1999	143

LISTA DE GRÁFICOS

GRÁFICO 1: Taxas de escolarização das pessoas de 5 anos ou mais de idade, segundo a faixa etária: Brasil, 2006	18
GRÁFICO 2. Taxas de aprovação do ensino fundamental, por série, segundo a região geográfica: Brasil, 2005	19
GRÁFICO 3. Taxas de reprovação do ensino fundamental, por série, segundo a região geográfica: Brasil, 2005	20
GRÁFICO 4: Taxas de abandono do ensino fundamental, por série, segundo a região geográfica: Brasil, 2005	21
GRÁFICO 5: Taxas de transição por séries, ensino fundamental: Brasil, 2003.	22
GRÁFICO 6: Taxas de distorção idade-série, no ensino fundamental: Brasil, 2003	23
GRÁFICO 7: Número médio de anos de estudo das pessoas de 10 anos ou mais de idade, segundo regiões geográficas e sexo: Brasil, 2005-2006	25
GRÁFICO 8: Sistema de sequencição, Ensino Fundamental, por escolas da Ficha B	72
GRÁFICO 9: Alunos matriculados (%) por série, em 2000	73
GRÁFICO 10: Alunos matriculados (%) por série, em 2001	73
GRÁFICO 11: Alunos matriculados (%) por série, em 2002	73
GRÁFICO 12: Alunos matriculados (%) por série, em 2003	73
GRÁFICO 13: Percentual de reprovação por série, alunos da Ficha B, segundo UF's, 1999 a 2003	82
GRÁFICO 14: Evasão por série, alunos da Ficha B, segundo UF's, 1999 a 2003	85
GRÁFICO 15: Curvas de sobrevivência à 1ª reprovação entre a 4ª e 8ª série, por UF's, 1999 a 2003	87
GRÁFICO 16: Curvas de sobrevivência à 1ª repetência entre a 4ª e 8ª série, por UF's, 1999 a 2003	87
GRÁFICO 17: Curvas de sobrevivência à evasão entre a 4ª e 8ª série, por UF's, 1999 a 2003	89
GRÁFICO 18: Curvas de sobrevivência ao 1º afastamento por abandono entre a 4ª e 8ª série, por UF's, 1999 a 2003	89
GRÁFICO 19: Curvas de sobrevivência à 1ª reprovação entre a 4ª e 8ª série, por UF's, 1999-2003: alunos sem defasagem idade-série	91

GRÁFICO 20: Curvas de sobrevivência à 1ª reprovação entre a 4ª e 8ª série, por UF's, 1999-2003: alunos com defasagem idade-série.	91
GRÁFICO 21: Curvas de sobrevivência à 1ª repetência entre a 4ª e 8ª série, por UF's, 1999-2003: alunos sem defasagem idade-série	91
GRÁFICO 22: Curvas de sobrevivência à 1ª repetência entre a 4ª e 8ª série, por UF's, 1999-2003: alunos com defasagem idade-série	91
GRÁFICO 23: Curvas de sobrevivência à evasão entre a 4ª e 8ª série, por UF's, 1999-2003: alunos sem defasagem idade-série	93
GRÁFICO 24: Curvas de sobrevivência à evasão entre a 4ª e 8ª série, por UF's, 1999-2003: alunos com defasagem idade-série	93
GRÁFICO 25: Curvas de sobrevivência ao 1º abandono entre a 4ª e 8ª série, por UF's, 1999-2003: alunos sem defasagem idade-série	93
GRÁFICO 26: Curvas de sobrevivência ao 1º abandono entre a 4ª e 8ª série, por UF's, 1999-2003: alunos com defasagem idade-série	93
GRÁFICO 27: Probabilidades estimadas da primeira repetência entre a 4ª e a 8ª série, segundo a série cursada: alunos da Ficha B, 1999-2003	103
GRÁFICO 28: Efeito das variáveis de aluno e escola sobre a probabilidade média de repetência entre a 4ª e a 8ª série do ensino fundamental: Ficha B, 1999-2003.	107
GRÁFICO 29: Probabilidades estimadas de evasão entre a 4ª e a 8ª série, segundo a série cursada: alunos da Ficha B, 1999-2003	114
GRÁFICO 30: Efeito das variáveis de aluno e escola sobre a probabilidade média de evasão entre a 4ª e a 8ª série do ensino fundamental: Ficha B, 1999-2003.	117
GRÁFICO 31: Probabilidades estimadas da primeira repetência e da evasão entre a 4ª e 8ª série do ensino fundamental, segundo as variáveis relevantes (modelos 6): Ficha B, 1999-2003	118
GRÁFICO 32: percentual de casos de repetência entre a 4ª e 8ª série: Ficha B, 1999-2003	120
GRÁFICO 33: Probabilidades de 1ª repetência entre a 4ª e a 8ª série: Ficha B, 1999-2003	120
GRÁFICO 34: Percentual de casos de evasão entre a 4ª e 8ª série: Ficha B, 1999-2003	121
GRÁFICO 35: Probabilidades de evasão entre a 4ª e a 8ª série: Ficha B, 1999-2003	121

LISTA DE TABELAS

TABELA 1: Total de alunos segundo o número de escolas e de turmas e UF's, Ficha B	70
TABELA 2: Fluxo de alunos segundo matrículas, evasão e transferências: Ficha B, 1999 a 2003	71
TABELA 3: Taxas de reprovação por ano, série e UF's: Ficha B, 1999 a 2003	81
TABELA 4: Total e percentual de reprovações por UF, segundo alunos matriculados em 1999, presentes durante o período de acompanhamento: Ficha B, 1999-2003	82
TABELA 5: Total e percentual de alunos segundo a situação de repetência entre a 4ª e 8ª série, por UF: Ficha B, 1999 a 2003	83
TABELA 6: Total e percentual de alunos segundo a série de ocorrência do afastamento por abandono entre a 4ª e 8ª série, por UF: Ficha B, 1999-2003	84
TABELA 7: Total e percentual de alunos segundo a situação de evasão, entre a 4ª e 8ª série, por UF: Ficha B, 1999-2003.	84
TABELA 8: Resultado do modelo incondicional para a probabilidade da primeira repetência entre a 4ª e a 8ª série do ensino fundamental: alunos da Ficha B, 1999-2003	96
TABELA 9: Modelos estimados para a probabilidade da primeira repetência entre a 4ª e 8ª série do ensino fundamental, segundo o Tempo: Ficha B, 1999-2003	97
TABELA 10: Modelos estimados para a probabilidade da primeira repetência entre a 4ª e 8ª série do ensino fundamental segundo a série cursada: alunos da Ficha B, 1999-2003	101
TABELA 11: Coeficientes significativos do modelo 6 e respectivas probabilidades da primeira repetência entre a 4ª e a 8ª série do ensino fundamental: alunos da Ficha B, 1999-2003	104
TABELA 12: Resultado do modelo incondicional para a probabilidade de evasão entre a 4ª e a 8ª série do ensino fundamental: alunos da Ficha B, 1999-2003	108
TABELA 13: Modelos estimados para a probabilidade de evasão entre a 4ª e a 8ª série do ensino fundamental, segundo o Tempo: alunos da Ficha B, 1999-2003	109
TABELA 14: Modelos estimados para a probabilidade de evasão entre a 4ª e a 8ª série do ensino fundamental, segundo a Série cursada: alunos da Ficha B, 1999-2003	113
TABELA 15: Coeficientes significativos do modelo 6 e respectivas probabilidades de evasão entre a 4ª e a 8ª série do ensino fundamental: alunos da Ficha B, 1999-2003	115
TABELA 16: IDH por UF's selecionadas e posição em relação às demais UF's brasileiras, 2000	125

TABELA 17: Percentual de alunos segundo variáveis fixas no tempo e Região. Ficha B, 1999-2003	141
TABELA 18: Percentual de alunos segundo a situação de trabalho e Região: Ficha B, 1999-2003	141
TABELA 19: Percentual de alunos por nível socioeconômico, proficiência e Região: Ficha B, 1999-2003	141
TABELA 20: Variáveis de escola segundo Regiões: Ficha B e Censo Escolar de 1999	142

RESUMO

Há concordância que, na atualidade, dois grandes problemas do sistema de ensino brasileiro são a evasão e a repetência, apesar de ainda existir um hiato em relação ao acesso à escola. Neste sentido, a principal pretensão dessa tese é a identificação e análise dos determinantes da probabilidade de ocorrência da primeira repetência e da evasão, entre a 4ª e a 8ª série do ensino fundamental. Outro objetivo refere-se à estimação das curvas de sobrevivência dos alunos a esses eventos, e também aos eventos primeira reprovação e primeiro afastamento por abandono. Foram utilizados dois métodos: o de modelos hierárquicos logísticos longitudinais e da análise de sobrevivência.

O desenvolvimento da tese está baseado na análise da trajetória escolar de uma coorte de alunos matriculados na 4ª série do ensino fundamental, em 1999, pertencentes a escolas situadas nas áreas urbanas de microrregiões localizadas em alguns estados das regiões Norte, Nordeste e Centro-Oeste do país. Estes alunos foram acompanhados até 2003, ano em que deveriam concluir a 8ª série.

É possível inferir que os dois métodos de análise se complementam e que seus resultados são compatíveis. A análise de sobrevivência revela que a probabilidade de sobrevivência do aluno à primeira repetência entre a 4ª e a 8ª série é muito inferior à probabilidade de sobrevivência à evasão. Em contrapartida, os modelos estimados mostram que a probabilidade média de repetência é muito superior à probabilidade média de evasão, para uma escola típica. Ambos os métodos evidenciam que o resultado escolar passado impacta de forma expressiva a trajetória escolar presente do aluno, sendo que os piores resultados educacionais foram observados para os alunos matriculados nas escolas da região Nordeste.

A análise hierárquica teve o mérito de revelar que nas escolas com uma menor proporção de repetentes é mais evidenciada a relação entre a ocorrência da repetência e a 5ª série. Supõe-se ser possível associar menor ocorrência do evento às melhores escolas (com melhores recursos físicos e humanos). Os resultados mostram que mesmo nessas escolas a repetência é mais expressiva na 5ª série. Essa série marca a transição entre os métodos de organização curricular. Percebe-se que o aluno é fortemente afetado pelas mudanças na grade curricular ocorridas entre a 4ª e a 5ª série. A sugestão é que haja uma revisão no projeto pedagógico de cada escola de forma a minimizar o efeito dessa mudança sobre o desempenho acadêmico. Por outro lado, as evidências apontam que a probabilidade de evasão aumenta nas últimas séries do ensino fundamental. Como nessas séries o aluno se encontra com uma idade relativamente mais avançada, é possível atribuir a saída da escola ao ingresso no mercado de trabalho. Nesse caso, é importante que sejam adotadas políticas públicas educacionais que favoreçam a permanência do aluno na escola. Mais importante, permanência na condição de aprovado.

ABSTRACT

There is agreement that currently, two large problems that the Brazilian educational system faces are dropouts and grade repeats, although there is also a gap in access to schools. In this sense, the main goal of this dissertation is to identify and analyze the determinants of the probability of occurrence of the first repeating grade or dropout, between 4th and 8th grades of what is called, fundamental school. The other objective is to estimate survival curves of students to these events, and also of initial repeating grade or initial dropout. Two methods were utilized: hierarchical logistic models and survival analysis.

The development of this dissertation is based upon the schooling trajectories of a cohort of students that were enrolled in 4th grade in 1999, in schools located in urban areas of micro-regions located in certain states of the North, Northeast, Central-west region of Brazil. These students were followed until 2003, year in which they should have completed the 8th grade.

It is possible to infer that the two methods of analysis are complementary and that their results are compatible. The survival analysis shows that the probability of survival of the student to the first repeating grade between 4th and 8th grade is much inferior to the probability of survival to dropout. On the other hand, the estimated models show that the mean probability of repeating a grade is much higher than the mean probability of dropout, for a typical school. Both methods make evident that the past performance in school expressively impacts the present trajectory of the student, and the worst results were observed for students enrolled in schools of the Northeast region.

The hierarchical analysis was able to show that in schools with smaller proportions of grade repeats; the relationship between occurrence of repeating grade and the 5th grade is more evident. It is possible to associate lower event occurrence to the best schools (with more physical and human resources). The results show that the even in those schools, repeating grades is more pronounced in 5th grade, which is the grade that marks the transitions between the methods of curricular organization. It is noticed that the students are heavily affected by changes in the curriculum that occur between 4th and 5th grade. A revision in the pedagogical project of each school is suggested, so as to minimize the effect of such transition on the school performance. Conversely, the evidence shows that the probability of dropout increases in the last grades of the fundamental school. As in these grades the student is at a relatively older age, it is possible to attribute dropping out of school to the entrance in the job market. In that case, it is important that public policies target adherence to school. More importantly, adherence given being approved in the grade.

INTRODUÇÃO

Tem havido um consenso na literatura internacional que a falta de acesso à escola, a evasão e a repetência constituem três grandes problemas enfrentados pelos sistemas educacionais contemporâneos. São problemas cuja magnitude atinge, sobretudo, as primeiras séries da educação fundamental dos diversos países, principalmente daqueles menos desenvolvidos.

No Brasil, o problema do acesso à escola, no ensino fundamental, está em vias de ser solucionado. Em 2000, aproximadamente 95% das crianças cuja idade estava entre 7 e 14 anos freqüentavam a escola (Censo Demográfico de 2000). Seis anos mais tarde, a taxa de escolarização para as crianças dessa faixa etária correspondeu a aproximadamente 98% (PNAD, 2006). Entretanto, o país ainda tem taxas de repetência e evasão que estão entre as mais altas do mundo.

Em 2003, a taxa nacional de repetência para o ensino fundamental equivaleu a 19,2%, sendo a taxa de evasão equivalente a 6,8%. Enquanto a taxa de repetência é mais elevada na 1ª e na 5ª série, a taxa de evasão aumenta continuamente ao longo das séries cursadas. Importante destacar que estes dois fatores resultam em altos valores para o indicador distorção idade-série. Na realidade, este indicador é fruto da ocorrência de três fatores: i) entrada tardia na escola; ii) repetência e; iii) abandono (com reingresso). Muitos estudos têm mencionado a existência de estreita ligação entre os dois últimos fatores.

Vários autores (COSTA-RIBEIRO, 1993; KLEIN, 1995; FLETCHER, 1997; SOUZA, 2001) têm atribuído a problemática da repetência e da evasão à baixa qualidade do ensino. Neste sentido, as políticas voltadas à área da educação estão sendo concentradas, basicamente, na questão da qualidade educacional. A qualidade tem sido mensurada, principalmente, através de testes padronizados de rendimento ou através do desempenho escolar do aluno. Neste contexto, nos anos recentes tem crescido o número de estudos, no país, analisando o desempenho acadêmico dos alunos matriculados no ensino fundamental (entre eles, RIOS-NETO, CÉSAR e RIANI, 2002; FERRÃO et al. 2002 a, b; CERQUEIRA, 2004; MACEDO, 2004; RIANI, 2005; PEREIRA, 2006).

Diversas têm sido as formas de medida das variáveis-resposta e os métodos utilizados pelos autores brasileiros na análise do desempenho escolar. CERQUEIRA (2004) desenvolveu seu estudo adotando, como metodologia de análise, o modelo tradicional de regressão linear (método dos Mínimos Quadrados Ordinários), utilizando como variáveis-resposta uma transformação logit das taxas de distorção idade-série,

taxas de repetência e taxas de abandono. MACEDO (2004) utilizou o modelo de especificação de valor adicionado, visando a identificação e a quantificação dos fatores associados ao rendimento dos alunos da 5ª série do ensino fundamental. Os demais autores citados estimaram modelos de regressão hierárquicos, considerados mais apropriados à dados educacionais. As variáveis-resposta mediram a progressão escolar (RIOS-NETO, CÉSAR e RIANI, 2002), a proficiência nos testes de português e de matemática (FERRÃO et al. 2002 a, b; PEREIRA, 2006) e o acesso ao sistema de ensino, eficiência e rendimento escolar (RIANI, 2005).

Contudo, devido à especificidade dos dados trabalhados e das variáveis-resposta utilizadas, esses estudos não captam a dimensão e a profundidade dos problemas da evasão e da repetência. Deve-se ressaltar que o estudo realizado por GOMES-NETO & HANUSHEK (1996) baseou-se numa base de dados longitudinal, sendo estimados modelos de regressão logística visando identificar os fatores determinantes da repetência em escolas da região Nordeste. Porém, os próprios autores afirmaram que a base não foi desenhada de forma a responder às questões relacionadas às causas e aos efeitos da repetência. Ademais, o estudo foi feito para os primeiros anos da década de 80, contemplando as primeiras séries do ensino fundamental (1ª a 4ª) e sendo restrito às escolas localizadas na área rural da região Nordeste (Ceará, Pernambuco e Piauí).

Além disso, a maioria dos estudos realizados não dá a ênfase necessária às três grandes regiões brasileiras que apresentam os piores resultados em termos de importantes indicadores educacionais: Nordeste, Norte e Centro-Oeste, respectivamente.

O desenvolvimento desta tese está baseado na análise da trajetória escolar de uma coorte de alunos matriculados na 4ª série do ensino fundamental, em 1999, pertencentes a escolas situadas nas áreas urbanas de microrregiões localizadas em alguns estados das regiões Norte, Nordeste e Centro-Oeste do país. Estes alunos foram acompanhados até 2003, ano em que deveriam concluir a 8ª série.

Ao longo da trajetória escolar, mais especificamente ao término do ano letivo t , os seguintes resultados podem constar na ficha do aluno: aprovação, reprovação, afastamento por abandono, transferência (para outra escola ou para a EJA /Supletivo Seriado) e falecimento. Os integrantes da coorte que foram aprovados no ano t estiveram expostos às seguintes situações no ano $t+1$: promoção, repetência (apesar da aprovação), transferência, evasão e falecimento. Os alunos reprovados ou afastados por abandono no ano t , por sua vez, estiveram sujeitos à repetência, evasão, transferência e falecimento no ano $t+1$.

Os principais objetivos da tese são a identificação e análise dos determinantes da probabilidade de ocorrência de dois desses eventos: a repetência e a evasão, entre a 4ª e a 8ª série. Como alguns alunos repetiram a série mais de uma vez, no período, a análise se restringe à primeira repetência. Outro objetivo refere-se à estimação das curvas de sobrevivência dos alunos a esses eventos, e também à primeira reprovação e ao primeiro afastamento por abandono.

Deve-se ressaltar que, nas pesquisas do ciclo de vida, estudam-se os processos sociais relacionados ao ciclo de vida dos indivíduos; mais comumente, sobre parte desse ciclo. No caso deste estudo, o aluno não foi acompanhado durante toda a sua trajetória escolar no ensino fundamental, mas durante uma parte muito relevante desta trajetória: desde a sua matrícula na 4ª série (1999) até o ano em que ele supostamente concluiria a 8ª série (2003). Apesar dessa limitação relacionada ao período de acompanhamento da coorte, trata-se de uma proposta inédita, na área da educação, no país.

Para o desenvolvimento do estudo foi utilizado o banco de dados “Avaliação de desempenho: fatores associados”, resultado de uma parceria firmada entre o Centro de Desenvolvimento e Planejamento Regional (CEDEPLAR) e o Instituto Nacional de Estudos e Pesquisas Educacionais (INEP). É a primeira base de dados longitudinais sobre o rendimento escolar, no Brasil, contendo informações anuais para o período de 1999 a 2003, para os alunos da 4ª a 8ª série do ensino fundamental, respectivamente. Além disso, utilizou-se também um banco complementar, fruto da mesma parceria, denominado Ficha Histórico Escolar, ou “Ficha B”. Neste banco, existem registros sobre o resultado final (aprovação, reprovação) do aluno a cada ano letivo, além de informações sobre transferência, afastamento por abandono, evasão e falecimento. No país, é a primeira vez que este banco de dados é utilizado. É possível inferir que o uso conjunto dessas bases resultou num grande avanço no estudo da problemática da evasão e da repetência escolar no sistema de ensino brasileiro. Informações sobre variáveis adicionais relacionadas às escolas da amostra foram extraídas do Censo Escolar de 1999.

São contempladas as áreas urbanas das microrregiões de seis estados brasileiros, localizados nas regiões Nordeste (Pernambuco e Sergipe), Norte (Pará e Rondônia) e Centro-Oeste (Goiás e Mato Grosso do Sul). Ou seja, o estudo abrange escolas das regiões que apresentam os piores indicadores educacionais do país.

Os métodos analíticos utilizados foram a análise de sobrevivência e os modelos hierárquicos logísticos longitudinais. O primeiro método teve por objetivo a estimação e análise das funções de sobrevivência do aluno à primeira reprovação, à primeira

repetência, ao primeiro afastamento por abandono e à evasão, desde sua matrícula na 4ª série no ano de 1999 até o ano de 2003. Como era de se esperar, nem todos os alunos permaneceram no estudo durante todo o período de acompanhamento. Daí a proposta de utilização dessa metodologia que permite a incorporação desse aspecto dos dados na modelagem.

Uma hipótese a ser testada é se as funções de sobrevivência são iguais para os alunos das diversas escolas/UF's. Pretende-se comparar as curvas de sobrevivência estimadas, visando verificar se os alunos matriculados nas escolas situadas nas UF's mais desenvolvidas sobrevivem mais tempo aos eventos de interesse. Pretende-se verificar, também, se os alunos matriculados na 4ª série, em 1999, que tinham idade defasada em relação a essa série, apresentam uma probabilidade de sobrevivência diferenciada em relação aos alunos matriculados na idade ideal. A suposição é que o tempo de sobrevivência aos eventos é maior para os alunos das UF's mais desenvolvidas e para os alunos cuja idade era adequada à 4ª série, em 1999.

O segundo método (modelos hierárquicos logísticos longitudinais) visou identificar os determinantes da probabilidade de ocorrência da repetência e da evasão nas séries finais do ensino fundamental (4ª a 8ª). Até então, o estudo desses problemas tem sido bastante limitado, devido à escassez de uma base de dados apropriada. A estimação dos modelos se baseou na Função de Produção Educacional - FPE - cuja especificação indica que o desempenho do aluno é uma função dos insumos familiares, dos insumos acumulativos das escolas/pares e de outros insumos relevantes, além da sua habilidade natural. A FPE tem sido muito utilizada na literatura internacional e nacional nas análises de desempenho escolar. Na maioria desses estudos, foi necessário fazer adaptações na especificação dos modelos, devido à natureza não longitudinal dos dados. Em geral, foram utilizados modelos de valor-adicionado ou modelos cujas variáveis foram coletadas em apenas dois pontos no tempo, entre outras adaptações.

No caso específico desta tese, a natureza dos dados utilizados (longitudinais) permitiu a incorporação, nos modelos, tanto de variáveis relacionadas à trajetória escolar passada quanto de variáveis contemporâneas do aluno. Alternativamente, permitiu incorporar tanto variáveis consideradas fixas como aquelas que variam no tempo. Esse aspecto dos dados proporciona inúmeros ganhos em relação aos modelos usualmente estimados que buscam identificar os fatores relacionados ao desempenho escolar.

Além disso, diferentemente das variáveis-resposta comumente utilizadas (resultados de testes aplicados), na tese o resultado educacional foi medido pelas variáveis

repetência e evasão escolar. É importante frisar que estas variáveis foram analisadas indiretamente num estudo feito por RIANI (2005), que abordou a questão da eficiência e do rendimento do aluno no ensino fundamental, identificando os determinantes da probabilidade do aluno cursar a escola na idade correta e da probabilidade de progressão por série (PPS). Porém, somente nesse estudo a repetência e a evasão têm sido estudadas de forma direta, devido à disponibilidade de informações que favorecem essa investigação, permitindo uma melhor compreensão desses dois grandes problemas presentes no sistema educacional brasileiro.

Outra inovação relacionada à especificação formal da FPE refere-se à estrutura dos dados existentes. Segundo a literatura, nos dados educacionais há uma estrutura hierárquica, sendo que os alunos estão agrupados em escolas que possuem características diversas. Portanto, a estimação dos modelos na forma convencional (MQO), conforme especificado pela FPE tradicional, gera resultados viesados. Para evitar esse problema de viés, é preciso incorporar essa hierarquia nas equações.

Na tese, são estimados modelos logísticos hierárquicos longitudinais de três níveis. No primeiro nível, a variável-resposta é uma função do tempo (série) associado à ocorrência do evento e de fatores relacionados aos alunos, que mudam ao longo do tempo; no segundo, das características fixas dos alunos e; no terceiro, dos fatores relacionados às escolas. Um dos méritos desses modelos é que eles permitem verificar como as variáveis num determinado nível afetam as variáveis nos demais níveis.

A seleção das variáveis inseridas nas regressões baseou-se, sobretudo, na FPE. Tanto nas regressões cuja variável-resposta mede a probabilidade de repetência quanto da evasão assume-se que os insumos familiares, as características do aluno e os insumos escolares são importantes determinantes da ocorrência desses eventos.

Os insumos familiares incluem dois indicadores do nível socioeconômico, uma medida do *background* familiar. As características dos alunos envolvem tanto variáveis que mudam ao longo do tempo como variáveis fixas no tempo. Ou ainda, tanto variáveis relacionadas à sua trajetória escolar passada quanto relacionadas à trajetória escolar contemporânea.

Os insumos escolares abrangem os recursos físicos (infra-estrutura, sistema de segurança) e humanos (qualificação do corpo docente), variáveis relacionadas ao tamanho (total de matrículas no ano base de 1999) e ao corpo discente (total de matrículas de alunos promovidos da 3ª para a 4ª série em 1999), além da região em que a escola está localizada (Norte, Nordeste e Centro-Oeste).

Além da inclusão dos vetores de insumos familiares e escolares, dois novos e importantes vetores foram inseridos na especificação dos modelos. Um deles contém o tempo e a série cursada pelo aluno. Porque um vetor com tempo e série? Por um lado, o tempo explicita o comportamento longitudinal dos eventos de interesse, ou seja, descreve a trajetória temporal da repetência e da evasão entre 1999 e 2003. Por outro, é possível inferir que a série também corresponde a uma medida de duração, pois de certa forma está associada ao tempo de permanência do aluno no estudo. É interessante substituir o tempo pela série cursada, visando identificar a série com maior probabilidade de ocorrência do evento, além de verificar se o seu efeito sobre cada evento varia ou não de acordo com as diferentes escolas.

O outro vetor é constituído pela proficiência do aluno (centralizada na média da escola) no ano letivo anterior ao acontecimento do evento, pela proficiência média da escola (desempenho dos colegas ou pares) e pela aprovação. A intenção, no primeiro caso, é estimar a probabilidade de repetência e de evasão do aluno no ano letivo $t+1$, condicional ao seu desempenho em testes padronizados realizados no ano letivo t . No segundo, pretende-se verificar o efeito do desempenho dos colegas da escola sobre a probabilidade de um aluno específico vivenciar os eventos de interesse. Finalmente, no terceiro caso, a variável aprovação tem a função de controle nas regressões cuja variável-resposta é a evasão. É estimada a probabilidade de evasão do aluno no ano $t+1$, condicional à sua aprovação na série k no ano t .

A partir desse conjunto de vetores, a idéia é verificar até que ponto o *background* familiar, as variáveis da trajetória escolar passada e contemporânea do aluno e os recursos existentes na escola influenciam a ocorrência da primeira repetência e da evasão entre a 4ª e a 8ª série do ensino fundamental. Será analisada a interação entre as variáveis dos três níveis, identificando-se, entre elas, aquela que mais afeta o resultado educacional. A suposição é que esses fatores impactam de forma diferenciada a ocorrência de cada evento específico.

Nesta investigação, ainda que a análise realizada para a coorte não contemple o ensino fundamental na sua totalidade, pois as séries estudadas variam da 4ª a 8ª, a expectativa é que a aplicação conjunta das duas metodologias aos dados longitudinais possibilite elucidar importantes questões sobre a repetência e a evasão escolar nesse nível de ensino. Além da elucidação dessas questões, a pretensão é que os resultados obtidos contribuam efetivamente para um melhor direcionamento das políticas educacionais

(públicas ou internas às escolas) no sentido de se reduzir a ocorrência de tais eventos no sistema educacional brasileiro.

Esta tese está organizada em sete capítulos, além desta parte introdutória. No primeiro, são apresentados alguns estudos e indicadores educacionais para o país e respectivas regiões, com destaque para os eventos reprovação, repetência, afastamento por abandono e evasão. Ênfase foi dada aos eventos de fluxo escolar, repetência e evasão, por constituírem as variáveis-resposta dos modelos de regressão. O capítulo 2 contempla o embasamento teórico que subsidiou o desenvolvimento deste estudo, mais especificamente, a Função de Produção Educacional, base para a especificação dos modelos estimados. No capítulo seguinte foram descritos os métodos da análise de sobrevivência e do modelo de regressão hierárquico logístico longitudinal, empregados para responder aos questionamentos levantados na tese. Uma descrição da amostra e das bases de dados utilizadas é feita no capítulo 4. No capítulo 5 é realizada uma discussão sobre os eventos básicos e as variáveis incluídas nas técnicas de análise. São explicitadas as estratégias empíricas adotadas para adequação às metodologias propostas. No capítulo 6 é apresentada a análise descritiva, visando uma melhor compreensão da trajetória escolar da coorte, e os resultados da análise de sobrevivência. O capítulo 7 apresenta os modelos de regressão estimados, com análise dos resultados obtidos. Por último, são feitas as considerações finais.

1. INDICADORES EDUCACIONAIS NO ENSINO FUNDAMENTAL BRASILEIRO: EVIDÊNCIAS TEÓRICAS E EMPÍRICAS RECENTES

Pretende-se, neste capítulo, apresentar alguns estudos relacionados a uns importantes problemas existentes no ensino fundamental brasileiro: reprovação, repetência, abandono e evasão. A idéia é mostrar o que tem sido feito pelos pesquisadores para um melhor entendimento desses problemas tão acentuados no sistema educacional do país. Após a abordagem teórica são apresentados e discutidos alguns indicadores educacionais, com ênfase para aqueles associados ao fluxo escolar: taxa de promoção, taxa de repetência e taxa de evasão. Esses indicadores de transição ou fluxo permitem fazer algumas inferências sobre a eficiência das redes de ensino e respectiva capacidade para gerar concluintes no sistema educacional. A pretensão é que o quadro a ser exposto contribua para uma melhor compreensão dos resultados obtidos na tese e respectiva análise desses resultados. O capítulo é finalizado com a colocação de alguns questionamentos sobre a possível influência do resultado escolar passado, de variáveis relacionadas à trajetória escolar passada e contemporânea do aluno e dos fatores escolares sobre a ocorrência de dois eventos intrinsecamente associados ao fluxo escolar: a repetência e evasão.

1.2 Indicadores educacionais no ensino fundamental brasileiro: evidências teóricas

Uma ampla literatura tem apontado que a falta de acesso à escola, a repetência e a evasão são problemas sérios enfrentados pelos sistemas educacionais contemporâneos. Estes problemas afetam principalmente os alunos matriculados nas primeiras séries do ensino fundamental e que freqüentam escolas localizadas nos países em desenvolvimento (sobretudo nas áreas rurais).

Como será visto na próxima seção, o problema do acesso à escola no ensino fundamental brasileiro está quase resolvido. Contudo, ainda são bastante elevadas as taxas de evasão e repetência nesse nível de ensino. Desta forma, nos últimos anos a atenção dos pesquisadores brasileiros tem sido direcionada basicamente para tais problemas.

COSTA-RIBEIRO (1993), analisando dados dos censos escolares para a década de 80, percebeu que havia uma grande distorção entre a população na faixa etária de 7 a 14 anos e a população matriculada no ensino fundamental. Em algumas séries, existiam mais alunos matriculados do que o total de crianças na idade correspondente à série.

Posteriormente, utilizando dados das Pesquisas Nacionais por Amostras de Domicílios (PNAD's), alguns autores (KLEIN, 1995; FLETCHER, 1997) constataram que

o problema apontado por Costa-Ribeiro estava associado à repetência, que retinha as crianças na escola. Estes autores, através do modelo Profluxo (desenvolvido para o cálculo de taxas de transição para as sucessivas séries), perceberam outro importante aspecto: a evasão escolar estava fortemente associada à reprovação, no sentido de que após sucessivas reprovações as crianças abandonavam a escola.

É consenso que as taxas de repetência estão relacionadas à baixa qualidade do ensino. Alguns estudos têm mostrado que a repetência acontece, principalmente, na 1ª e na 4ª série do ensino fundamental.

Utilizando o método Probabilidade de Progressão por Série (PPS), com base nos dados do Censo Demográfico 2000, para as pessoas de 20 a 64 anos na data de referência do censo, RIOS-NETO (2004) identificou as séries historicamente mais importantes para a evolução da escolaridade da população brasileira. Ao analisar as coortes extremas nascidas entre 1943-1987 e 1983-1987, foi observado um aumento de 3,7 anos de estudo entre ambas as coortes, que foi associado basicamente ao aumento em duas PPS's: e_0 (conclusão da 1ª série do ensino fundamental) e e_4 (conclusão da 5ª série do ensino fundamental). Estas PPS's explicaram aproximadamente 76% da variação na escolaridade entre as duas coortes.

Os resultados sugerem, para a 1ª série, um aumento na cobertura educacional e uma redução da repetência escolar nesta série, entre as coortes. Os resultados indicam também uma redução na reprovação na 4ª série, que apresentou uma maior probabilidade de progressão para a 5ª série.

Ao analisar a taxa de distorção idade-série, que equivale ao total de matrículas de pessoas cursando uma série numa idade superior à considerada ideal sobre o total de matrículas nesta série, RIANI e GOLGHER (2004) verificaram que a 1ª série do ensino fundamental foi a que apresentou os maiores níveis de repetência. De acordo com os autores, em 2000 havia no Brasil cerca de 38,5% de alunos matriculados na 1ª série fora da idade adequada, sendo que o Nordeste apresentava os maiores índices para esta taxa.

Diante das altas taxas de repetência observadas no Brasil, comparadas até mesmo às taxas observadas em alguns países da África Sub-Sahariana (UNESCO, 2008), nos últimos anos tem havido uma ampliação do volume de trabalhos visando avaliar a qualidade do ensino no país. Entre estes trabalhos, são listados alguns cuja ênfase foi dada à reprovação, à repetência e à evasão escolar no ensino fundamental.

GOMES-NETO & HANUSHEK (1996) procuraram analisar os determinantes da repetência escolar no Nordeste brasileiro. Duas questões foram tratadas. Primeiro,

considerando que em geral inexistiam escolas com séries mais avançadas (até a 4ª série) nas áreas rurais da Região, foram analisadas as causas da não oferta dessas séries pelas escolas. A hipótese levantada foi que as características das escolas e dos municípios constituíam os principais fatores que afetavam a probabilidade da escola ofertar séries avançadas. A outra questão referiu-se à análise dos fatores que afetavam a repetência do aluno. Os alunos que repetiram a 2ª série duas vezes foram comparados com os demais alunos (que evadiram, que cursavam a 3ª série e os promovidos para a 4ª série).

Foram estimados modelos de regressão logística, por meio do conjunto de dados do EDURURAL¹ (dados longitudinais para os anos de 1981, 1983 e 1985). As variáveis explicativas utilizadas foram divididas em três categorias: características das escolas, condições econômicas do município e organização/fatores governamentais. Para mensurar as características das escolas foi criado um índice, chamado “hardware”, que incluía o número de salas de aulas, cadeiras para estudantes, cozinha e secretaria/escritório. As condições econômicas do município foram medidas pela proporção de famílias que vendiam uma parte de sua colheita e pela proporção que participava de um programa emergencial (relacionado à limitação da produção agrícola). A variável relacionada à organização governamental foi constituída por um índice incluindo a quantidade e a qualidade do *staff*.

Os autores ressaltaram que o banco de dados não foi desenhado de forma a responder aos dois principais problemas de interesse, isto é, as causas e os efeitos da repetência, mas que sua avaliação foi uma boa oportunidade para se direcionar esses problemas. A análise se restringiu às primeiras séries do ensino fundamental (1ª a 4ª), ofertadas nas escolas localizadas na área rural da região (estados do Ceará, Pernambuco e Piauí).

Entre os principais resultados, pode-se citar: i) a probabilidade de ofertar a 4ª série foi maior para as escolas que tinham um grande número de alunos e que possuíam melhores instalações físicas; ii) a probabilidade de ofertar disciplinas só até a 2ª série foi maior para as escolas que “funcionavam” na casa dos professores, mas o efeito estimado não foi estatisticamente significativo; iii) as variáveis associadas às condições econômicas locais e os fatores governamentais não foram estatisticamente relacionados à estrutura da série da escola; iv) o *background* familiar afetou diretamente as probabilidades de repetência: alunos cujos pais tinham mais escolaridade apresentaram menores

¹ Programa de expansão e melhoria da educação no meio rural do Nordeste.

probabilidades de repetir um ano em comparação aos alunos cujos pais tinham menos escolaridade ou eram analfabetos; v) menores testes escores resultaram em maiores probabilidades de repetência; vi) o sexo e a idade do aluno não tiveram efeito significativo sobre as probabilidades de repetência; vii) alunos de escolas cuja 2ª série era a mais elevada foram mais prováveis à retenção nesta série, sinalizando que uma proporção da repetência era relacionada à falta de outras oportunidades escolares; viii) alunos residentes em municípios de maiores índices socioeconômicos apresentaram maiores probabilidades de repetir a série e; ix) as probabilidades de repetência no Ceará foram mais elevadas em relação aos demais estados.

De uma forma geral, os dois principais determinantes da repetência foram os níveis de desempenho do aluno (baixo desempenho) e a política governamental (no sentido de não ofertar séries mais avançadas). Posteriormente, GOMES-NETO & HANUSHEK (1996) estimaram os efeitos da repetência sobre a aprendizagem do aluno, utilizando como variáveis dependentes testes de português e de matemática. Uma variável indicadora foi incluída no modelo, designando se o aluno era repetente ou não. Os autores ressaltaram alguns problemas presentes na especificação do modelo, afirmando que ele deveria ser visto apenas como uma crua aproximação dos efeitos da repetência. Por exemplo, inferiram que a repetência não é um fator exógeno, sendo afetada pelo desempenho, o que implica numa causalção em ambas as direções, fazendo com que as estimativas do efeito puro da aprendizagem da repetência sejam viesadas. Foi frisado também que a medida de repetência não indicava quantos anos o aluno repetiu, mas apenas se o aluno esteve na mesma série no ano anterior e que a estrutura do banco não permitia a estimação dentro do contexto do valor-adicionado.

Os autores verificaram que os alunos que repetiram a série tiveram um melhor desempenho nos testes que os não repetentes, levando à conclusão que os alunos aprendem com a repetência. Concluíram afirmando que a melhor política com relação à repetência é a melhoria na qualidade das escolas primárias, que resulta num aumento direto do desempenho do aluno.

O trabalho desenvolvido por SOUZA (2001) centrou-se na questão da baixa qualidade do ensino, expressa pelas altas taxas de repetência e evasão. Seu estudo foi restrito ao estado de São Paulo. Nele foram feitas referências às políticas adotadas a partir de meados da década de 80 no Estado. Entre elas, a implementação do ciclo básico, ou promoção automática.

A autora desenvolveu seu trabalho abordando o polêmico debate relacionado à incompetência do professor, estudando a perspectiva dos professores com relação a ações de uma educação continuada. Ela questionou a ênfase dada ao argumento da incompetência do professor como principal explicação para o mau desempenho do sistema educacional. Segundo o argumento, a principal estratégia a ser adotada para melhorar a qualidade do ensino é aumentar a competência dos professores através de programas de educação continuada.

Sua tese contesta o pensamento dominante, ao considerar que a educação do professor não pode ser tomada como a única causa da baixa qualidade do ensino no país. Nela, é argumentado que são necessárias condições de trabalho adequadas para o desenvolvimento do trabalho de magistério, além da reformulação de planos de carreira dos professores e salários decentes. É destacado também que devem ser considerados outros importantes fatores como a pobreza, o desemprego, o sistema de habitação e saúde, entre outros.

CERQUEIRA (2004) investigou os principais determinantes do desempenho escolar no ensino fundamental brasileiro, utilizando dados do Censo Escolar de 1999 e da Pesquisa de Informações Básicas Municipais de 1999, do IBGE. O desempenho escolar foi representado pela taxa de distorção idade-série, taxa de repetência e taxa de abandono. Foram estimados modelos de regressão, pelo método dos Mínimos Quadrados Ordinários, aplicando-se uma transformação logit às taxas mencionadas. As variáveis explicativas utilizadas procuraram caracterizar, sobretudo, os aspectos relacionados à infra-estrutura educacional dos municípios.

Eis alguns dos resultados obtidos: i) maiores taxas de distorção idade-série associadas aos municípios com um número mais elevado de alunos por turma e menores taxas associadas a melhores indicadores de infra-estrutura municipal; ii) entre a 1ª e a 4ª série o percentual de escolas com quadra e a qualificação docente tiveram um efeito redutor nas taxas de repetência, enquanto entre a 5ª e a 8ª série um maior número de alunos por turma teve um efeito positivo sobre essas taxas e; iii) entre a 1ª e a 4ª série a qualificação docente exerceu um efeito redutor sobre as taxas de abandono, sendo que entre a 5ª e a 8ª série esse efeito foi exercido pelo percentual de escolas com laboratório de ciências.

O autor concluiu seu artigo destacando a importância da infra-estrutura escolar e da qualificação docente como fatores relacionados às taxas de eficácia escolar.

MACEDO (2004) desenvolveu seu estudo buscando identificar o efeito valor

adicionado e o efeito da heterogeneidade sobre os fatores associados ao rendimento escolar dos alunos matriculados na 5ª série. Um dos resultados obtidos refere-se à perda do efeito dos fatores estruturais (sexo e raça) sobre os rendimentos do aluno na série, diante da inclusão da nota do teste aplicado na 4ª série. Com a inclusão do valor adicionado, houve perda de efeito também para alguns fatores de fluxo. Entretanto, os fatores relacionados à trajetória escolar passada (repetência) permaneceram bastante expressivos, enquanto as variáveis relacionadas à escola foram afetadas de forma mais aleatória. Com relação aos efeitos da heterogeneidade, quando foram comparados os resultados do modelo contemporâneo que incluiu todos os alunos matriculados na 5ª série em 2000 com os resultados do modelo contemporâneo que incluiu somente os alunos da 5ª série em 2000 que cursaram a 4ª série em 1999, verificou-se que, de uma forma geral, a heterogeneidade reduziu o impacto da maioria das variáveis na determinação do rendimento escolar. Entre os principais resultados, pode-se citar: i) alunos negros tiveram um menor rendimento escolar; ii) o trabalho foi prejudicial ao rendimento dos alunos; iii) um maior número de repetências afetou mais negativamente o rendimento e; iv) quanto melhores as instalações físicas das escolas, melhor o rendimento dos alunos.

É importante destacar que, a partir da implementação do Sistema de Avaliação para Educação Básica (SAEB)² em 1990, cresceu substancialmente o número de estudos visando analisar o desempenho dos alunos matriculados no ensino fundamental. Grande parte destes estudos tem utilizado a metodologia dos modelos de regressão hierárquicos, em geral considerados mais apropriados para a análise dos fatores determinantes do desempenho escolar, dada a estrutura hierárquica dos dados educacionais.

FERRÃO et al (2002) desenvolveram seu estudo baseando-se no pressuposto da existência da “escola eficaz”. A partir dos dados do SAEB 1999, foram investigados alguns dos principais fatores associados à melhoria do desempenho escolar dos alunos da 4ª série do ensino fundamental, sendo estimados modelos de regressão hierárquicos em que a variável-resposta foi a proficiência dos alunos.

Principais resultados: i) em todas as regiões brasileiras, os alunos declarados de raça/cor preta apresentaram um desempenho inferior em relação aos demais alunos; ii) o

² Importante frisar também que em 2007 foi criado, pelo Inep, o Índice de Desenvolvimento da Educação Básica (IDEB). Nele, são reunidos dois conceitos importantes associados à qualidade da educação: o fluxo escolar e as médias de desempenho nas avaliações. O cálculo do indicador é feito a partir dos dados sobre aprovação escolar (Censo Escolar), do desempenho nas avaliações para os estados e para o país (SAEB) e, no caso dos municípios, dados da Prova Brasil. Dado o período recente da sua implementação, ainda são poucos os estudos baseados no Indicador.

desempenho dos alunos com distorção idade-série foi inferior ao dos alunos com idade adequada à série; iii) com relação ao fato do aluno fazer lição de casa, os sinais dos coeficientes estimados tiveram a direção esperada; iv) não foram encontradas evidências de que alunos em turno integral tenham melhor desempenho em relação aos alunos em turno parcial; v) à exceção da região Sul, as escolas com ensino organizado apenas em ciclos tiveram, em média, piores resultados em relação às escolas com o ensino organizado por séries; vi) o desempenho do aluno foi impactado positivamente pela experiência do professor, aferido pelo número de anos na profissão na região Norte (nas demais regiões, o coeficiente associado à variável não foi estatisticamente significativo) e; vii) nas regiões Norte e Sudeste observou-se que os alunos cujos professores tinham menos escolaridade tiveram os seus resultados escolares reduzidos em relação aos resultados dos alunos cujos professores tinham licenciatura.

Posteriormente, FERRÃO et al. (2002?) estimaram modelos de regressão hierárquicos utilizando o mesmo banco de dados, com o objetivo de investigar o impacto de políticas de não-repetência no desempenho escolar dos alunos matriculados na 4ª série do ensino fundamental na região Sudeste e verificar se os alunos com defasagem idade-série pertencentes a escolas com promoção automática tinham desempenho diferenciado em relação aos demais alunos. A variável-resposta foi a proficiência, sendo as principais covariáveis a defasagem escolar e o regime de organização do ensino.

Entre os principais resultados (alguns similares ao do estudo realizado pelos autores, para o Brasil), observou-se que: i) o desempenho dos alunos negros foi menor do que o desempenho dos alunos de outras cores/raças; ii) as estimativas dos parâmetros associados à defasagem idade-série mostraram que para cada ano adicional de defasagem tinha-se uma redução na proficiência; iii) as estimativas pontuais relacionadas à adoção da promoção automática foram desfavoráveis tanto na rede pública quanto na particular; iv) não foram constatadas evidências de que alunos com defasagem escolar tenham o desempenho acadêmico reduzido em relação aos colegas, pelo fato de estudarem em escolas com política de não-repetência e; v) não foram encontradas evidências de que os alunos mais pobres das escolas com promoção automática tenham menor proficiência. Os autores concluíram, com relação ao regime de organização do ensino, que a promoção automática pode contribuir para a correção da distorção idade-série sem a perda da qualidade da educação.

RIOS-NETO, CÉSAR e RIANI (2002) analisaram os determinantes do desempenho escolar no Brasil, utilizando modelos de regressão logísticos hierárquicos,

introduzindo como variável-resposta a progressão escolar (uma novidade em relação aos métodos tradicionalmente utilizados). Os autores concatenaram os microdados de crianças numa determinada idade com respectiva família (nível 1) com os macrodados de um determinado ano e localização (nível 2), visando captar os efeitos macros da política educacional. Foram analisadas as probabilidades de conclusão da 1ª e da 4ª série do ensino fundamental, com base nos dados das PNAD's para as décadas de 80 e 90.

Entre os principais resultados, constatou-se uma associação positiva e significativa entre a educação materna e a probabilidade de progressão em ambas as séries. Evidenciou-se também um efeito positivo e significativo da educação média do professor sobre o intercepto. Os autores mencionam como principal conclusão o efeito substituição entre a escolaridade média dos professores e a escolaridade materna, efeito importante na probabilidade de progressão na 1ª série.

Mais recentemente, RIANI (2005) investigou os determinantes do resultado educacional no país, focando sua análise nos níveis de ensino fundamental e médio, utilizando os dados do Censo Demográfico e do Censo Escolar de 2000. Foi adotada a metodologia hierárquica e hierárquica espacial, considerando como resultado educacional duas importantes dimensões da educação: o acesso ao sistema de ensino e a eficiência e rendimento na escola. A primeira dimensão foi mensurada pela probabilidade de freqüentar a escola para os alunos de 7 a 14 anos e de 15 a 17 anos, e a segunda pela probabilidade dos alunos (das mesmas faixas etárias) estarem matriculados numa série adequada à idade.

Nos modelos estimados a autora considerou no primeiro nível o aluno e no segundo, os municípios. A seleção das variáveis relevantes foi baseada na especificação da FPE. Os insumos relacionados ao primeiro nível ou ao *background* familiar do aluno foram: educação materna, categoria de ocupação do chefe, chefia feminina e presença de familiares conviventes. Foram incluídas também as variáveis idade, sexo, cor e localização da residência. No segundo nível, foram incluídas variáveis relacionadas ao perfil da rede escolar do município. Estas variáveis foram assim agrupadas: 1) qualidade dos recursos humanos (média de alunos por turma, média de horas/aula por dia e porcentagem de professores com curso superior); 2) qualidade da infra-estrutura escolar, restrição de oferta (razão professor/população em idade escolar); 3) nucleação (distribuição das matrículas segundo o tamanho dos estabelecimentos) e; 4) distribuição das matrículas segundo a dependência administrativa e a organização do ensino em ciclos. Neste nível, incluiu-se também uma variável relacionada ao dividendo demográfico (tamanho relativo da coorte em idade escolar).

Os principais resultados decorrentes da estimação dos modelos hierárquicos foram: i) uma menor pressão demográfica tem um importante impacto nas duas dimensões consideradas; ii) os fatores de *background* familiar foram positivamente associados ao resultado educacional; iii) o fator restrição de oferta foi o mais significativo para o maior acesso escolar dos jovens de 15 a 17 anos; iv) os fatores relacionados à qualidade dos serviços educacionais foram os que apresentaram o maior impacto sobre a probabilidade dos alunos de 7 a 14 anos freqüentarem a escola na idade correta, enquanto que para os alunos de 15 a 17 anos o principal fator foi a infra-estrutura/restrrição de oferta e; v) as variáveis porcentagem de professores com curso superior, média de alunos por turma e o fator de infra-estrutura tiveram maior impacto na probabilidade do aluno passar de zero para um ano estudo, de quatro para cinco anos de estudo e de oito para nove anos de estudo. Para esta última progressão, a razão professor/população também foi significativa.

Com relação à modelagem hierárquica espacial, os modelos foram especificados de forma a determinar o impacto nas variáveis dependentes em função de mudanças nos fatores contextuais dentro do município (efeito direto) e de forma a permitir a avaliação do impacto nas variáveis dependentes em função de mudanças nas variáveis contextuais no município e nos municípios vizinhos (efeito indireto).

Os novos resultados podem ser assim listados: i) na análise do dividendo demográfico, ao se separar os efeitos direto e indireto, há uma queda substancial no primeiro efeito, sendo que a restrição de oferta passa a ser o fator contextual mais importante para o acesso escolar; ii) mudanças significativas nas probabilidades dos alunos freqüentarem a escola decorrem de variações no tamanho relativo da coorte em idade escolar nos municípios vizinhos; iii) a inclusão do espaço nos modelos reduz os impactos diretos nos indicadores de eficiência e rendimento e; iv) variações nos fatores contextuais nos municípios vizinhos têm impactos substanciais sobre os indicadores de determinados municípios.

Outra importante contribuição do estudo foi a análise do *trade off* entre educação materna e os fatores associados à rede escolar dos municípios. A autora constatou que os fatores relacionados à qualidade escolar têm um efeito substituição com a educação materna, sobretudo para a probabilidade de freqüentar a escola para os alunos de 7 a 14 anos e para a probabilidade de progressão na 1ª série. Assim, ela infere que a melhoria dos fatores relacionados ao perfil escolar do município pode contribuir para a redução das desigualdades educacionais entre os alunos de diferentes origens socioeconômicas.

Nessa seção, os estudos realizados apontam para a necessidade de melhorar a qualidade dos fatores escolares, uma vez que essa melhoria tem impactos significativos sobre os resultados educacionais. Além disso, ressaltam também a importância do *background* familiar nesse processo, entre outros aspectos. A seção seguinte tem por objetivo mostrar a atual situação do ensino fundamental no Brasil, com base em alguns indicadores educacionais.

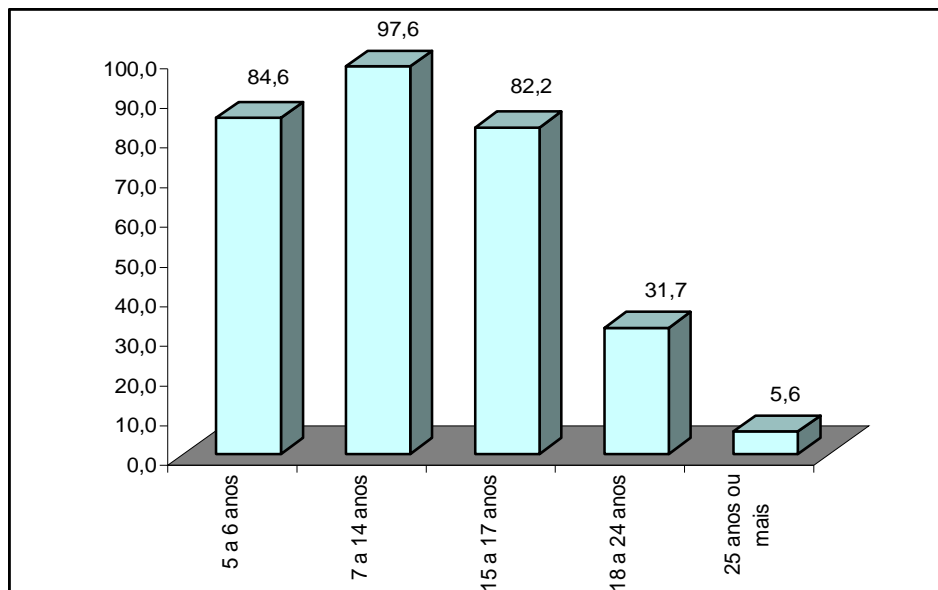
1.3 Indicadores educacionais no ensino fundamental brasileiro: evidências empíricas

No Brasil, até a década de 60, o acesso ao sistema de ensino era restrito a relativamente poucos alunos. Esse quadro começa a ser mudado nas décadas de 60 e 70, quando houve um aumento expressivo da demanda por educação, no país. Esse aumento foi atribuído, em grande medida, à migração da população das áreas rurais e dos estados pobres do Nordeste para as áreas urbanas e mais desenvolvidas da região Sudeste (SOUZA, 2001).

Na década de 80 o acesso à escola ainda constituía uma questão prioritária na agenda governamental, sendo que somente no ano 2000 foi observada uma tendência à universalização nesse acesso, para o ensino fundamental. Os dados do Censo Demográfico de 2000 mostraram uma cobertura de quase 95% para as crianças de 7 a 14 anos, vis-à-vis o índice de cobertura observado no início da década de 90 (da ordem de aproximadamente 80%, conforme Censo de 1991). Dados mais recentes (2006) mostram que a taxa de escolarização³ para os estudantes desse grupo etário equivaleu a quase 98% (GRÁF.1).

³ A taxa de escolarização corresponde à percentagem de estudantes de um determinado grupo etário em relação à população do mesmo grupo etário.

GRÁFICO 1
Taxas de escolarização das pessoas de 5 anos ou mais de idade,
segundo a faixa etária: Brasil, 2006

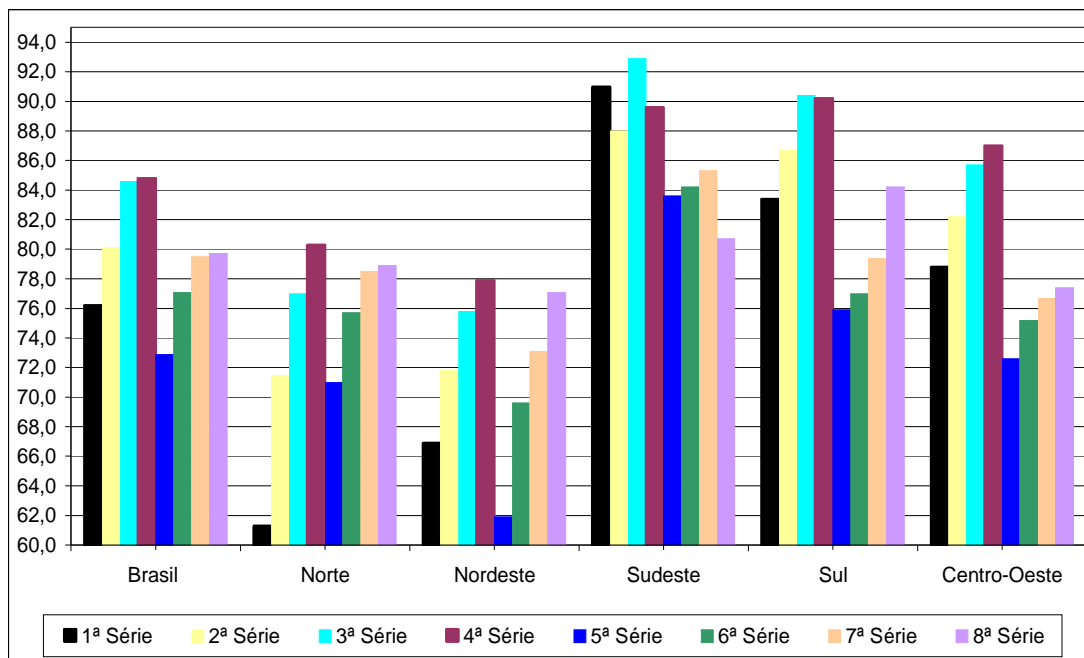


Fonte: Elaboração própria a partir dos dados do IBGE: Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento. Pesquisa Nacional por Amostra de Domicílios 2006.

Uma vez que o país está prestes a atingir a universalização no ensino fundamental, atenção deve ser direcionada aos problemas da repetência e da evasão. Na seqüência, são apresentadas as taxas de aprovação, taxas de reprovação e taxas de abandono para as regiões brasileiras, pois essas taxas têm uma estreita relação com os problemas em questão. Como a tese contempla as séries 4^a a 8^a e as regiões Norte, Nordeste e Centro-Oeste, ênfase será dada à tais séries e à tais regiões.

A taxa de aprovação corresponde à proporção de alunos que são aprovados na série k , no ano t , em relação à matrícula total desse ano. A taxa de aprovação no ensino fundamental registrada no Brasil, em 2005, equivaleu a 79,5% (GRÁF. 2). Nota-se que no país e em quase todas as regiões essa taxa foi crescente nas séries iniciais, sofrendo uma queda brusca na 5^a série. Comportamento diferenciado é observado na região Sudeste, que apresenta taxas oscilantes entre as diversas séries.

GRÁFICO 2
 Taxas de aprovação do ensino fundamental, por série,
 segundo a região geográfica: Brasil, 2005

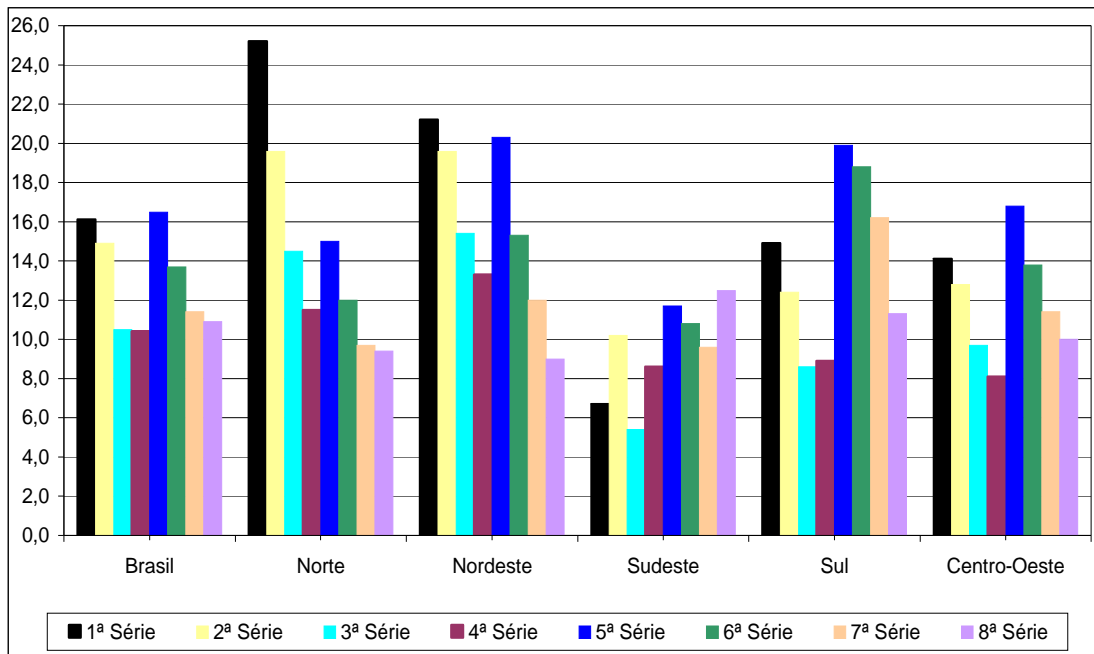


Fonte: Elaboração própria a partir dos dados do MEC/INEP.

No gráfico é possível observar que para quase todas as regiões brasileiras, as taxas de aprovação são maiores na 4ª série, sendo bem mais elevadas nas regiões mais desenvolvidas (em torno de 90% para o Sul e Sudeste). Em contrapartida, na 5ª série são registradas as menores taxas. A exceção é verificada para o Norte e o Sudeste, cujo indicador tem menores valores para a 1ª e a 8ª série respectivamente. Considerando-se as séries 4ª a 8ª e as três regiões sob estudo, nota-se que as taxas de aprovação são menores na região Nordeste.

A taxa de reprovação corresponde à proporção de alunos que são reprovados em relação à matrícula total, na série k, no ano t. No Brasil, o valor desse indicador para o ensino fundamental equivaleu a 13% em 2005. O GRÁF. 3 mostra que no país a maior taxa de reprovação foi registrada na 5ª série, ou seja, uma série que marca a transição entre os métodos de organização curricular. Chama a atenção o fato das regiões Nordeste e Sul apresentarem praticamente o mesmo valor para esse indicador nessa série (aproximadamente 20%). Nota-se ainda que, entre a 4ª e a 8ª série, a 6ª série também se sobressai por apresentar altas taxas de reprovação.

GRÁFICO 3
Taxas de reprovação do ensino fundamental, por série,
segundo a região geográfica: Brasil, 2005

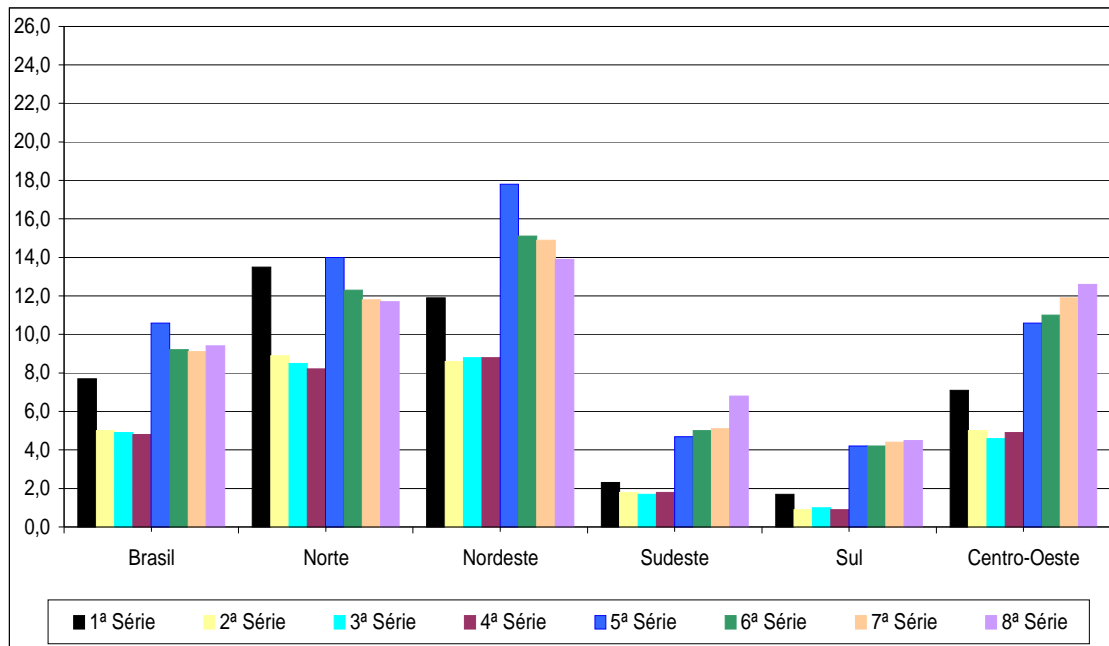


Fonte: Elaboração própria a partir dos dados do MEC/INEP.

Importante mencionar que os resultados apresentados podem estar associados à forma de organização do ensino adotada pelas escolas das diferentes regiões (sistema seriado, sistema seriado com progressão continuada, sistema de ciclos, outras formas). O Sudeste destaca-se entre as demais regiões brasileiras por possuir o maior percentual de escolas e de alunos matriculados no sistema de ciclos e no sistema seriado com progressão continuada.

As taxas de abandono são mostradas no GRÁF. 4. Essas taxas são calculadas considerando a proporção de alunos afastados por abandono no ano t e na série k , em relação à matrícula total no ano t . Em 2005, a taxa nacional média de abandono, no ensino fundamental, equivaleu a 7,5%. No país, a maior taxa foi registrada para a 5ª série (10,6%) e a menor para a 4ª série (4,8%). Uma análise por região revela que as maiores taxas de abandono, para as escolas das regiões Norte e Nordeste, foram registradas na 5ª série, enquanto para as escolas das demais regiões o abandono foi mais significativo na 8ª série. De uma forma geral, observa-se que as regiões Sul e Sudeste apresentaram valores para esse indicador bem inferiores aos valores registrados para as outras regiões. A partir da 4ª série constata-se que o Nordeste se destaca por apresentar as maiores taxas de abandono do país.

GRÁFICO 4
Taxas de abandono do ensino fundamental, por série,
segundo a região geográfica: Brasil, 2005



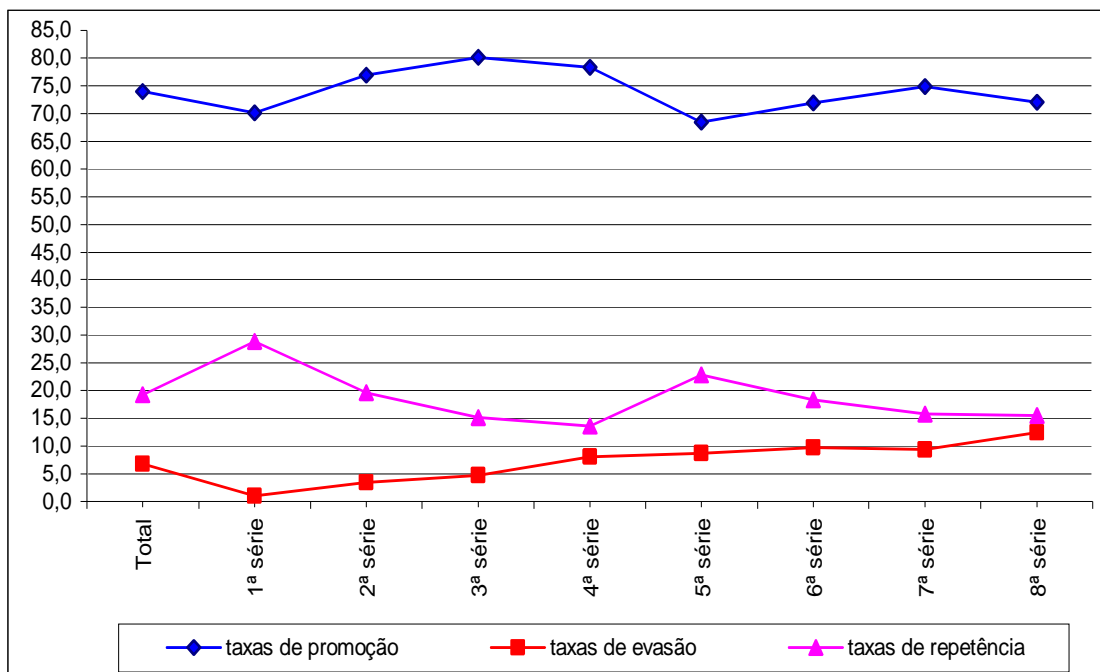
Fonte: Elaboração própria a partir dos dados do MEC/INEP.

Na seqüência são apresentados alguns indicadores relacionados à trajetória escolar dos alunos no sistema educacional brasileiro, nível fundamental: taxa de promoção, taxa de repetência e taxa de evasão⁴. Esses indicadores consistem num importante instrumento para verificar a progressão do aluno no sistema escolar, num determinado nível de ensino seriado, considerando a situação de promoção, repetência e evasão.

Em 2003, a taxa de promoção para o ensino fundamental equivaleu a 74%. De acordo com o GRÁF. 5, o valor mais elevado para esse indicador foi registrado na 3ª série (80,1%) e o menor valor na 5ª série (68,4%). A taxa nacional de repetência equivaleu a 19,2%. A curva referente a essa taxa tem seus picos na 1ª (28,9%) e na 5ª série (22,9%), atingindo o menor valor na 4ª série (13,6%). A visualização gráfica revela também que as taxas de evasão aumentam gradativamente ao longo das séries, sendo mais elevadas na última série do ensino fundamental. Os valores mínimo e máximo para esse indicador são 1% e 12,5% respectivamente, sendo a taxa total igual a 6,8%.

⁴ A taxa de promoção é calculada considerando a proporção de alunos da matrícula total na série k, num determinado ano t, que tenham se matriculado na série k+1 no ano t+1. O cálculo da taxa de repetência considera a proporção de alunos da matrícula total na série k, no ano t, matriculados na série k no ano t+1. A taxa de evasão corresponde à proporção de alunos da matrícula total na série k, no ano t, não matriculados no ano t+1.

GRÁFICO 5
Taxas de transição por série, ensino fundamental: Brasil, 2003.



Fonte: Elaboração própria a partir dos dados do MEC/INEP.

Embora alguns estudos apontem que em geral nos países em desenvolvimento cujas taxas de repetência são elevadas têm-se altas taxas de evasão (entre eles, Relatório UNESCO 2008), as evidências empíricas mostradas não apontam para uma correlação positiva entre ambas as taxas, ainda que sejam altas em relação às observadas noutros países.

Para se ter uma idéia sobre a magnitude da taxa de repetência no Brasil, serão apresentados alguns dados sobre esse indicador noutros países. Relatório da UNESCO (2008?), baseado em dados educacionais para o ano de 2005, aponta taxas de repetência no ensino fundamental de aproximadamente 30% para a República Centro Africana e Burundi. Valores elevados também foram registrados para o Brasil (21,2%), Nepal (20,6%) e Cabo Verde (15,4%). Por outro lado, entre os países desenvolvidos, a maior taxa foi registrada em Portugal (10,2%). Na Espanha e na Alemanha esse indicador equivaleu a 2,3% e 1,4%, respectivamente, sendo inferior a 1% em países como Finlândia, Grécia, Irlanda e Itália. O Relatório destaca também que em diversas regiões esse indicador atinge seu valor máximo na 1ª série, citando o Brasil como um dos países que apresenta uma taxa bastante elevada de repetentes nessa série (27%).

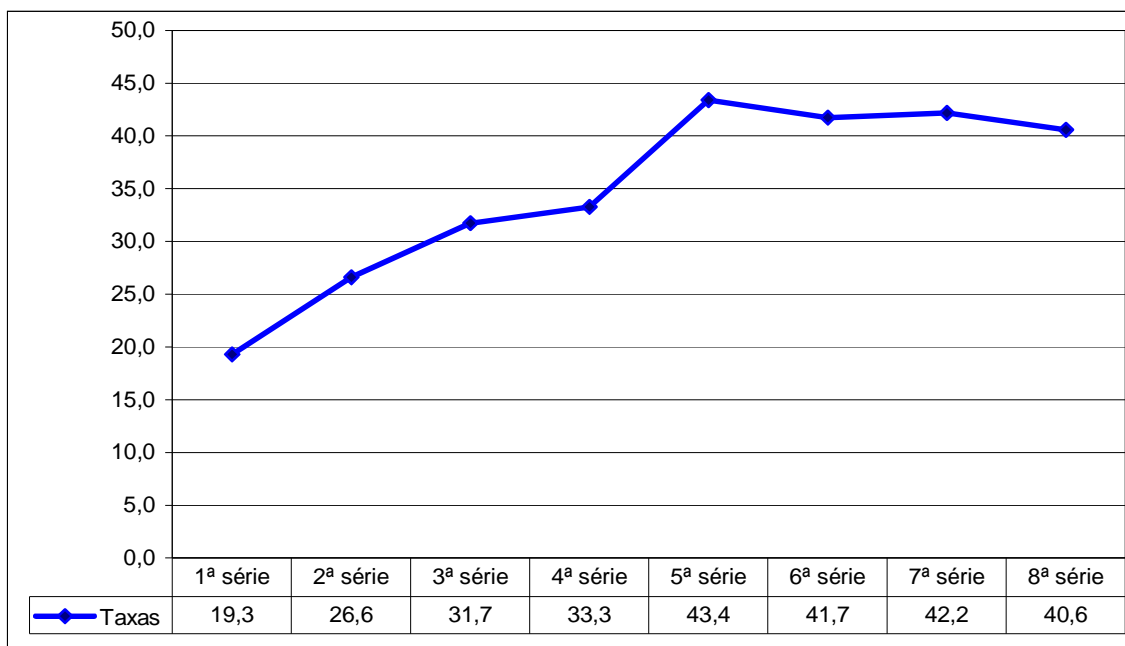
Evidencia-se que o país tem uma taxa de repetência para o ensino fundamental cujo valor está entre os maiores do mundo. Entretanto, é preciso ter cautela na análise

comparativa desse indicador, pois a repetência está relacionada também às políticas de promoção adotadas nos diversos sistemas educacionais dos diferentes países.

Com relação à evasão, o Relatório destaca que suas causas são múltiplas e complexas, abrangendo situações específicas dos diversos países, fatores particulares do aluno e o nível das redes de ensino. Entre os problemas das redes de ensino, são citados a falta de recursos e de segurança nas escolas, o excesso de alunos nas salas de aula e a falta de qualificação dos professores. Foi relatado que nos países em desenvolvimento, mesmo as escolas bem equipadas são incapazes de evitar a evasão, se o aluno estiver submetido a uma situação de pobreza ou miséria. Mas foi frisado que diante dos problemas enfrentados pelo aluno (pessoais, familiares, financeiros, de trabalho) as escolas podem evitar sua saída do sistema caso lhe seja dado o apoio necessário para lidar com as dificuldades externas à sala de aula.

Sabe-se que a repetência e o abandono no ensino fundamental resultam em altas taxas de distorção idade-série. Portanto, é importante mostrar o comportamento desse indicador nas diversas séries desse nível de ensino, no Brasil.

GRÁFICO 6
Taxas de distorção idade-série, no ensino fundamental: Brasil, 2003



Fonte: Elaboração própria a partir dos dados do MEC/INEP.

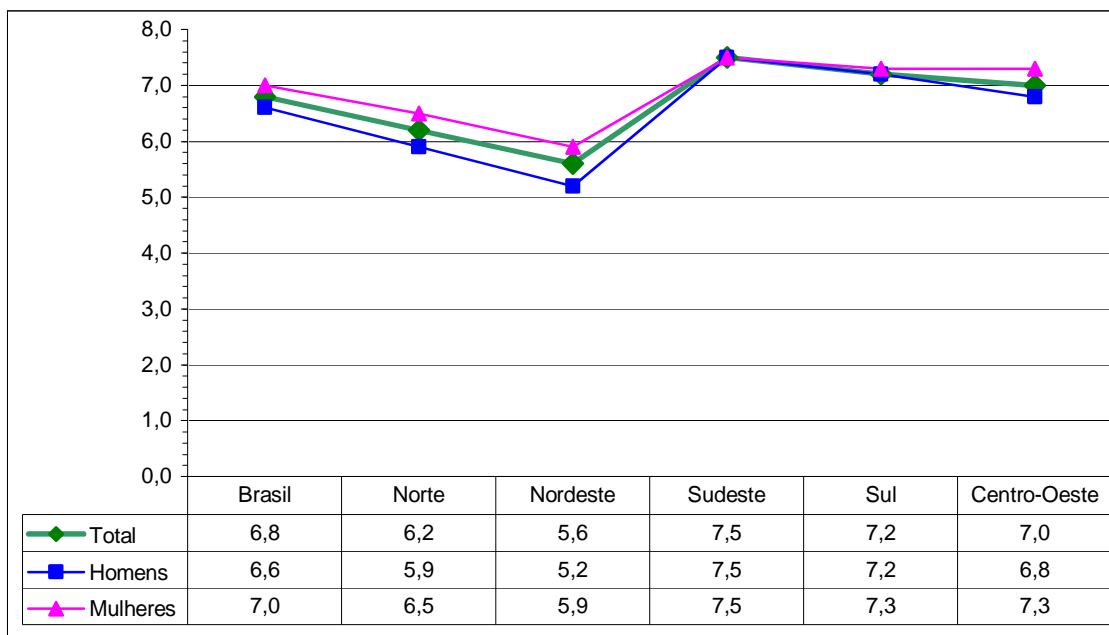
Na totalidade das séries cursadas, evidencia-se que são bastante elevados os valores das taxas de distorção idade-série. Verifica-se que na 1ª e na 5ª série é que se concentram, respectivamente, o menor e o maior valor para essa taxa. Esse resultado sugere que a defasagem idade-série está relacionada basicamente à retenção do aluno ao longo das séries (e não à sua entrada tardia no sistema). Nas séries iniciais essa taxa é crescente, declinando sensivelmente após a 5ª série. Na transição da 4ª para a 5ª série é que houve a maior variação percentual nesse indicador. Esses dados são preocupantes, uma vez que estudos existentes têm apontado que alunos com atraso escolar têm um desempenho inferior aos alunos com idade adequada à série.

Um último indicador é apresentado: o número médio de anos de estudo da população brasileira com 10 anos ou mais de idade. No seu cálculo, cada série concluída com aprovação corresponde a um ano de estudo. É interessante mostrar o comportamento desse indicador (GRÁF. 7), uma vez que a sua estimativa inclui o grau de atendimento do sistema de ensino⁵ e as taxas de rendimento escolar: aprovação, reprovação e evasão. Quanto maiores os níveis de atendimento escolar e as taxas de aprovação, mais alta a escolaridade média. Em contrapartida, altas taxas de reprovação e evasão resultam em menor escolaridade média.

⁵ O grau de atendimento escolar é medido pela proporção da população que frequenta a escola. Se esse cálculo é feito por faixa etária, tem-se a taxa de atendimento escolar.

GRÁFICO 7

Número médio de anos de estudo das pessoas de 10 anos ou mais de idade, segundo regiões geográficas e sexo: Brasil, 2005-2006



Fonte: Elaboração própria a partir dos dados do IBGE: Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento. Pesquisa Nacional por Amostra de Domicílios 2005-2006.

Nota: No cálculo estão incluídas as pessoas com idade ignorada.

Em 2005 o número médio de anos de estudo da população brasileira correspondeu a 6,8 anos, sendo que as mulheres tiveram uma escolaridade média superior à dos homens. Uma análise por região revela que o Sudeste apresentou o mais alto valor para esse indicador (inclusive não houve variação nesse valor, considerando o sexo), enquanto o menor valor foi registrado para o Nordeste. É possível argumentar que a reprovação e a evasão contribuem, negativamente, para o baixo número de anos de estudo dos brasileiros.

As evidências empíricas apresentadas mostraram a ampliação do acesso à escola no ensino fundamental brasileiro. Entretanto, apesar desse avanço no sistema educacional do país, os indicadores revelaram a continuidade da dicotomia repetência x evasão. Estes problemas, que constituem um dos principais fatores determinantes da distorção idade-série, assumem dimensões ainda maiores nas regiões Nordeste, Norte e Centro-Oeste. Neste sentido, é relevante a realização de um estudo que contemple a trajetória escolar dos alunos do ensino fundamental, matriculados principalmente nas escolas destas regiões.

A seção seguinte enfatiza os eventos repetência e evasão, por constituírem o foco dessa tese.

1.3 Repetência e evasão no ensino fundamental brasileiro: em busca de uma melhor compreensão sobre estes eventos associados ao fluxo escolar

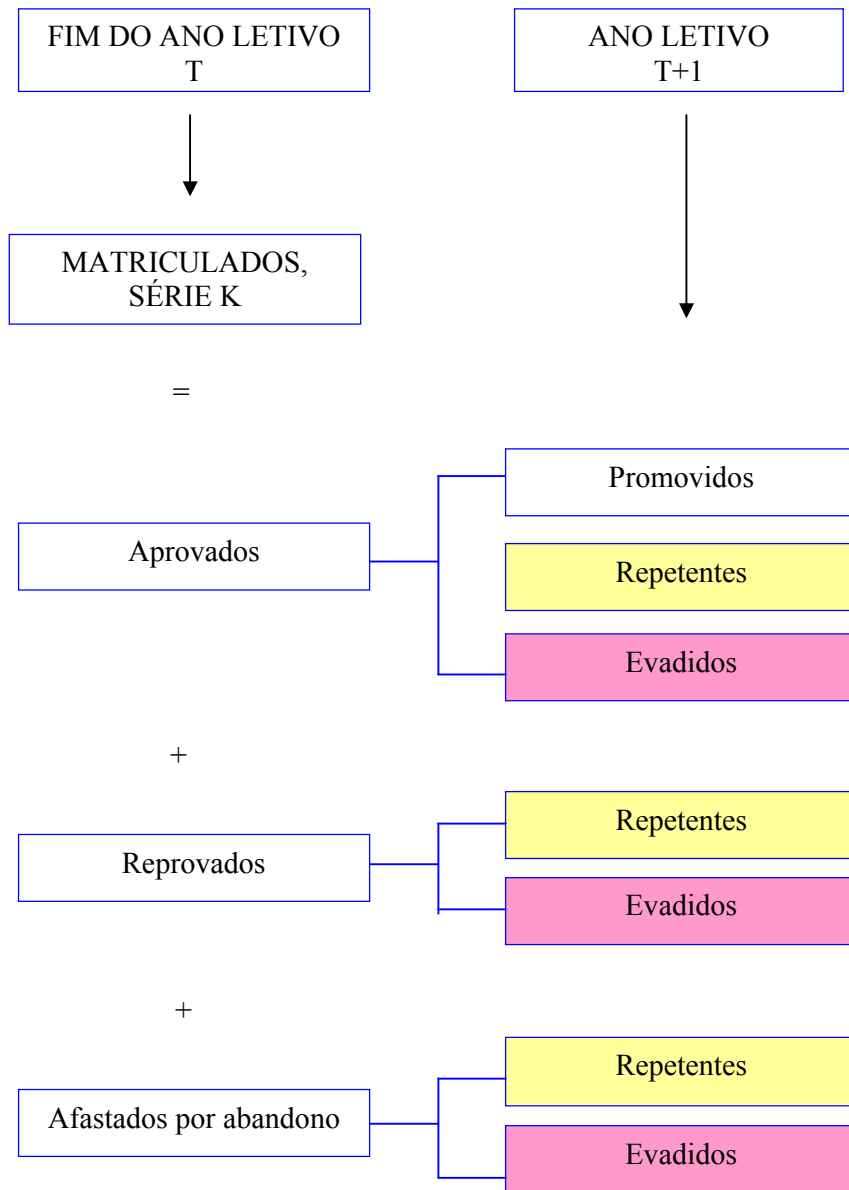
Entre os diversos indicadores vistos nesse capítulo, ênfase é dada, na presente tese, aos indicadores de fluxo escolar, notadamente aqueles que constituem sérios problemas no sistema educacional. Mais precisamente, não são contempladas as taxas de repetência e de evasão, e sim a ocorrência dos eventos repetência e evasão entre os alunos da coorte acompanhada.

Focar esses eventos é particularmente importante considerando-se as estatísticas apresentadas. Vale frisar que estudos têm considerado a repetência como um dos indicadores relacionados à qualidade da educação e ao desempenho dos alunos. Estudos têm mostrado também que, dentre os alunos matriculados na 1ª série do ensino fundamental, parte significativa não conclui esse nível de ensino. Uma explicação para esse fato é a evasão escolar. Portanto, é crucial enfatizar tais problemas visando melhor compreendê-los.

Como o próprio nome sinaliza, os indicadores de fluxo estão associados ao fluxo escolar, ou seja, à trajetória escolar do aluno. Assim, o evento escolar ocorrido num determinado ano letivo $t+1$ está associado ao resultado escolar ocorrido numa determinada série k , no ano letivo anterior, t . Seja a seguinte figura:

FIGURA 1

Situação dos alunos referente eventos escolares de fluxo



Fonte: Elaboração própria, com base em KLEIN (1995).

O esquema é baseado no fluxo escolar proposto por KLEIN (1995), sob o pressuposto de um sistema fechado. Ao fim do ano letivo t os alunos matriculados nesse sistema, numa determinada série k, são considerados aprovados, reprovados ou afastados por abandono. É considerado afastado por abandono o aluno que, ao fim do ano letivo, não foi considerado aprovado nem reprovado.

No esquema, há um destaque para os eventos indicativos de fluxo que constituem o objeto de interesse desse estudo: a repetência e a evasão. O aluno é considerado repetente caso tenha se matriculado, no ano letivo $t+1$, na mesma série k cursada no ano t . Verifica-se que no ano $t+1$ a repetência numa determinada série k pode ocorrer em função de três situações ocorridas no ano anterior: aprovação, reprovação e afastamento por abandono, nesta série. De forma similar, a evasão no ano letivo $t+1$ pode ocorrer diante dessas mesmas situações, verificadas no ano t . A diferença é que nesse caso o aluno não efetivou sua matrícula no sistema em nenhuma série do nível de ensino.

Na tese, a principal pretensão é identificar, através da estimação de modelos hierárquicos longitudinais, os principais fatores explicativos da ocorrência da primeira repetência e da evasão entre a 4^a e a 8^a série, na coorte sob estudo. No processo de estimação dos modelos, são considerados fatores intra-alunos, inter-alunos e escolares. Pretende-se verificar até que ponto esses fatores contribuem para um melhor entendimento dos problemas da repetência e da evasão escolar no ensino fundamental. A suposição inicial é que o resultado escolar anterior do aluno é um importante determinante da ocorrência desses eventos.

Como visto na FIG. 1, a repetência e a evasão no ano $t+1$ estão associadas a um determinado resultado escolar ocorrido no ano t . Neste estudo, o resultado escolar é mensurado pela proficiência do aluno em testes de português e de matemática (valor médio). Obviamente, se o aluno abandonou os estudos ao longo do ano letivo t , ele não respondeu as questões dos testes (realizados em novembro de cada ano).

Com relação ao papel dos diversos fatores sobre a ocorrência da repetência e da evasão, os seguintes questionamentos podem ser levantados:

- i) Uma maior proficiência na série k no ano letivo t contribui significativamente para a redução da ocorrência da evasão num determinado ano letivo $t+1$? O efeito é similar para a repetência?
- ii) Em que medida a probabilidade de repetência e de evasão de um aluno específico é afetada pela proficiência dos seus colegas de escola?
- iii) Na identificação dos determinantes da ocorrência dos eventos de interesse, são consideradas variáveis associadas à trajetória escolar passada e contemporânea do aluno. Em que medida tais variáveis

contribuem para aumentar ou reduzir a probabilidade de ocorrência desses eventos?

- iv) Entre as variáveis relacionadas ao *background* familiar e à escola, quais são mais importantes para reduzir a probabilidade de repetência numa série? O resultado é válido também para o caso da evasão?
- v) Dados do MEC/INEP (2003) revelaram que entre a 4ª e a 8ª série, a 5ª se sobressai por apresentar as maiores taxas de repetência do país, enquanto as taxas de evasão aumentam gradualmente ao longo das séries. Para a coorte em estudo, a probabilidade de repetência é mais acentuada também na 5ª série? E a probabilidade de evasão por série segue o mesmo comportamento observado para o país?
- vi) Qual o efeito do contexto socioeconômico regional sobre a ocorrência da repetência e da evasão no ensino fundamental? Há diferenciação nesse efeito para cada evento específico analisado?

Uma importante contribuição da tese está relacionada às respostas às questões levantadas. No Brasil esses aspectos dos dados educacionais ainda não foram estudados, tendo-se como variáveis-resposta a evasão e a repetência, devido à inexistência de uma base longitudinal apropriada.

Ressalta-se que a proficiência do aluno é geralmente incluída nas regressões, porém, na forma de variável explicada. E, por falta de dados na ficha escolar, a variável indicadora de duração, série cursada, não é incorporada na especificação dos modelos. Acredita-se que a introdução dessas variáveis nas regressões gere resultados que permitam importantes ganhos aos estudos existentes sobre os determinantes do desempenho escolar.

Uma contribuição adicional refere-se aos resultados da análise de sobrevivência. Este método, que tem um caráter mais descritivo, contempla além dos eventos destacados no esquema, aqueles não assinalados e que são estreitamente relacionados à repetência e à evasão, ou seja, a reprovação e o afastamento por abandono. A pretensão é comparar as curvas de sobrevivência do aluno a cada evento de interesse visando verificar, entre outros aspectos, possíveis diferenciais nessas curvas relacionadas às características das escolas/regiões em que essas escolas estão inseridas.

A expectativa é que os resultados obtidos através dos dois métodos de análise preencham importantes lacunas associadas à repetência e à evasão no ensino fundamental brasileiro. No próximo capítulo é apresentado o marco teórico que subsidiou o desenvolvimento do estudo.

2. ARCABOUÇO TEÓRICO: A FUNÇÃO DE PRODUÇÃO EDUCACIONAL

No processo de identificação dos fatores associados à primeira repetência e à evasão entre a 4ª e a 8ª série do ensino fundamental foram estimados modelos de regressão hierárquicos, especificados com base na Função de Produção Educacional. Na sequência é feita uma breve descrição da FPE e das principais dificuldades relacionadas à sua especificação.

2.1 A Função de Produção Educacional: questões relacionadas à especificação

A FPE, bastante utilizada nos estudos sobre o resultado escolar, tem o seguinte formato geral⁶:

$$O_{it} = f(F_t^{(t)}, S_t^{(t)}, O_t^{(t)}, A_t) + v_{it} \quad (1)$$

Sendo:

O_{it} o desempenho do aluno i no tempo t ;

$F_t^{(t)}$ o vetor dos insumos da família, acumulativos no tempo t ;

$S_t^{(t)}$ o vetor dos insumos acumulativos das escolas;

$O_t^{(t)}$ o vetor de outros insumos relevantes;

A_t a habilidade natural do aluno e ;

v_{it} o termo de erro aleatório, que capta os efeitos aleatórios que influenciam o desempenho do aluno no tempo t .

Nos estudos baseados na FPE, uma questão bastante discutida refere-se às variáveis que devem ser consideradas. No caso dos insumos escolares, HANUSHEK (2002) destaca as principais medidas utilizadas: i) recursos reais das salas de aula (nível educacional do professor, experiência do professor e tamanho de turma ou razão professor-aluno); ii) recursos financeiros agregados (gastos por aluno e salário do professor) e; iii) medidas de outros recursos escolares (características específicas do professor, insumos administrativos e instalações).

⁶ Questões relacionadas à especificação da FPE podem ser acompanhadas em LAZEAR (1999) e HANUSHEK (2002), entre outros.

Quanto aos insumos familiares, na maioria dos estudos (BARBOSA & FERNANDES, 2001; HANUSHEK & LUQUE, 2002; RIOS-NETO, CÉSAR & RIANI, 2002; HANUSHEK, LAVY & HITOMI, 2006) têm sido utilizadas as variáveis escolaridade dos pais e nível socioeconômico da família.

No vetor de outros insumos relevantes geralmente são considerados fatores de turma ou de pares e fatores comunitários (HANUSHEK, 1998; HANUSHEK et al, 2001; KRUEGER, 2003; RIANI, 2005).

O efeito da habilidade natural do aluno, em geral, é captado pelo termo de erro aleatório, devido à dificuldade associada à sua mensuração.

Desde a proposição de COLEMAN (1966), diversos estudos têm sido feitos baseados na abordagem da FPE, ampliando o debate sobre o verdadeiro papel das escolas e do *background* familiar sobre o desempenho escolar (HEYNEMAN & LOXLEY, 1983; HANUSHEK, 1997; KRUEGER, 1998; KRUEGER, 2002; HANUSHEK, 2002). Especificamente com relação às variáveis relacionadas às escolas (insumos escolares), os resultados têm sido divergentes. É consensual entre os autores que problemas de identificação, forma funcional, nível de agregação e seleção das variáveis-controle relevantes resultam em vieses nos parâmetros estimados, o que pode explicar parte das divergências verificadas. Noutras palavras, grande parte da incerteza existente na literatura sobre a importância dos efeitos dos recursos escolares sobre o desempenho é devida à especificação incorreta da Função de Produção Educacional.

Segundo HANUSHEK (2002), o maior obstáculo na fase de especificação das regressões está relacionado à natureza acumulativa do processo de produção educacional. A história de vida da família (*background* familiar), a influência da composição dos pares ao longo das séries estudadas e os recursos existentes das escolas contribuem para o desempenho do estudante num determinado ano. No entanto, são raros os bancos de dados que possuem informações retrospectivas para essas variáveis.

Nesse sentido, na ausência de uma base de dados apropriada (longitudinal), estudiosos têm utilizado as seguintes metodologias, entre outras, visando explicar o resultado escolar do aluno:

a) Modelos de especificação contemporânea: segundo a proposição desses modelos, o rendimento do aluno é explicado pelos insumos familiares e educacionais correntes, mais um termo de erro aleatório, que capta o efeito dos insumos passados e da capacidade cognitiva não observada do aluno. Os pressupostos desses modelos são: i) somente os insumos contemporâneos são importantes na determinação do desempenho corrente; ii) os

efeitos desses insumos não mudam ao longo do tempo, de tal forma que os insumos correntes captam os efeitos de toda a trajetória passada referente ao aprendizado do aluno e; iii) ausência de correlação entre os insumos contemporâneos e a capacidade cognitiva não observada do aluno.

b) Modelos de valor-adicionado: o que diferencia esses modelos da especificação contemporânea é a inclusão na equação do rendimento do aluno na série anterior (score do teste anterior). O pressuposto básico do modelo é que o score do teste anterior é uma medida suficiente para captar os efeitos das informações passadas e da capacidade cognitiva não observada do aluno.

Entre esses modelos, o mais comumente utilizado é o de valor-adicionado, pois seus pressupostos são mais coerentes. Porém, existem críticas relacionadas à sua aplicação na análise do rendimento escolar. Por exemplo, seja a variável tamanho de classe ou de turma. KRUEGER (2002) ressalta que muitos estudos que buscaram estimar o efeito dessa variável sobre o desempenho utilizaram a abordagem do valor-adicionado, tendo como variável-resposta os ganhos nos testes scores e como principal variável explicativa o tamanho da classe. Como o modelo identifica apenas o efeito acumulativo do tempo de permanência numa pequena classe, sendo que o efeito inicial é perdido - pela diferenciação -, o autor infere que pode haver uma perda dos principais efeitos de se estudar em classes menores.

Um problema adicional decorrente do uso da especificação do valor-adicionado para a estimação da importância do tamanho de turma é citado por TODD & WOLPIN (2003). Os autores argumentam que mesmo que não existam variáveis omitidas que sejam correlacionadas com o tamanho de turma (pressuposto improvável), ainda assim existirá um viés nos resultados se os testes scores forem correlacionados serialmente.

Além da ausência de dados acumulativos para os insumos, que implica nos problemas mencionados, existe a dificuldade de mensuração de algumas variáveis relevantes, entre elas, o efeito dos pares, ou tamanho de classe. Uma discussão sobre essa questão pode ser acompanhada em HANUSHEK (2002) e HANUSHEK & LUQUE (2002). Igualmente difícil de mensurar é a habilidade natural do aluno, devido à existência de uma correlação desse fator com as características familiares (HANUSHEK, 2002).

Como verificado, o maior problema relacionado à FPE está associado à especificação precisa dos diversos insumos. Além disso, o processo educacional é acumulativo e raramente os bancos de dados existentes permitem contemplar esse aspecto. Neste estudo, as bases longitudinais utilizadas possibilitam incorporar na FPE algumas

informações do aluno no período anterior ao início do seu acompanhamento na escola, juntamente com as informações contemporâneas. Portanto, na tese, parte dos problemas citados nessa seção pode ser contornada mediante uma re-especificação da FPE, a partir dos dados existentes. Essa re-especificação é discutida na seqüência.

2.2 Estrutura empírica da FPE: uma adaptação aos dados / variáveis existentes

Na seção anterior foi apresentada a FPE e os problemas relacionados à sua especificação. Nesta seção, são abordados os ajustes feitos à FPE de forma que ela se adapte aos dados educacionais existentes e retrate, da melhor forma possível, a realidade atinente aos fenômenos que se deseja estudar, visando responder aos questionamentos colocados no capítulo anterior.

Na equação (1) a variável-resposta é o desempenho do aluno. Na maioria dos estudos esse desempenho é medido pela escolaridade (número de anos completos de estudo) e, principalmente, pelas notas de testes padronizados (proficiência). Nesse estudo, a variável-resposta comumente utilizada - desempenho - é substituída pelas variáveis repetência e evasão. Ou seja, tem-se a substituição de uma variável-resposta medida na forma contínua por variáveis-resposta medidas na forma categórica. Essa especificidade das variáveis-resposta requer que, em vez da estimação usual dos parâmetros do modelo pelo método dos Mínimos Quadrados Ordinários, esta estimação seja feita através do método da Quase Verossimilhança.

Na quase totalidade dos estudos dos determinantes do desempenho escolar baseados na estrutura da FPE foram feitas adaptações na equação devido a inexistência de bases de dados longitudinais. Ou foram estimados modelos de valor-adicionado ou modelos considerando-se apenas dois pontos no tempo, entre outras especificações. A natureza dos dados utilizados nessa tese permite a incorporação, no lado direito da equação (1), tanto de variáveis relacionadas à trajetória passada quanto de variáveis contemporâneas do aluno. Ou ainda, permite incorporar tanto variáveis consideradas fixas (ex.: variáveis demográficas) como aquelas que variam no tempo (ex.: proficiência). Esse aspecto dos dados contorna grande parte da discutida problemática relacionada à especificação dos insumos presentes na FPE.

Na modelagem utilizada, dois novos e relevantes vetores são acrescentados. O primeiro refere-se ao tempo e à série cursada associada à ocorrência de ambos os eventos, enquanto o segundo inclui a proficiência do aluno (centralizada na média da escola) e a proficiência média da escola (proficiência dos pares ou colegas), ambas mensuradas no ano

anterior ao ano de ocorrência da repetência e da evasão, além da informação sobre aprovação. Nos estudos existentes no país focando os determinantes do resultado escolar, essas questões ainda não foram intensivamente trabalhadas, em virtude da inexistência de uma base de dados longitudinais que contemple essas informações.

Outra inovação presente na tese refere-se à estrutura dos dados utilizados. Nos dados educacionais há uma estrutura hierárquica: os alunos estão agrupados em escolas que possuem características diversas. Portanto, a princípio não seria correto impor a restrição de que tanto o intercepto quanto os coeficientes relacionados às covariáveis sejam necessariamente constantes entre os grupos (escolas). Nesse sentido, a estimação dos modelos na forma convencional, conforme especificado pela FPE tradicional, gera resultados viesados. É preciso incorporar no modelo a estrutura hierárquica dos dados. Devem ser estimados, portanto, modelos de regressão hierárquicos logísticos, onde a variável resposta (binária) é medida no menor nível de agregação e as covariáveis são medidas nos demais níveis existentes. Mais especificamente, dada a natureza longitudinal dos dados, são estimados modelos de regressão hierárquicos logísticos longitudinais.

Com base nessas colocações, a equação (1) é assim re-especificada, visando responder às questões anteriormente levantadas:

$$R_{it+1} = f(F_t^{(t)}, E_t^{(t)}, P_{t-1}, T_t) + v_{it} \quad (2)$$

Sendo:

R_{it+1} a repetência da série pelo aluno i no tempo $t+1$;

$F_t^{(t)}$ o vetor dos insumos da família e do aluno, acumulativos no tempo t ;

$E_t^{(t)}$ o vetor dos insumos acumulativos das escolas;

P_t o vetor com a proficiência média do aluno e dos colegas da escola no ano letivo t ;

T_t o vetor com o tempo (e/ou série cursada pelo aluno) associado à ocorrência do evento e ;

v_{it} o termo de erro aleatório, que capta os efeitos aleatórios que influenciam a ocorrência da repetência no tempo t .

O mesmo conjunto de insumos é incluído nos modelos cuja variável-resposta mede a probabilidade de evasão (E_{it+1}), sendo que a situação de aprovação é incluída no terceiro vetor (em substituição à proficiência média).

Importante mencionar que no Brasil, apesar do crescimento substancial do número de estudos sobre os determinantes do resultado educacional que utilizam regressões

hierárquicas (FERRÃO et al, 2002; RIOS-NETO, CESAR & RIANI, 2002; RIANI, 2005), são raros os modelos hierárquicos longitudinais estimados (entre eles, PEREIRA, 2006). Portanto, esse conjunto de adaptações aplicado à estrutura da FPE traz importantes avanços no estudo do resultado escolar, em relação à especificação dos modelos tradicionais.

Espera-se que, diante desses ajustes e dessas inovações, os resultados dos modelos estimados sejam os mais condizentes possíveis com a realidade vivenciada nos sistemas educacionais.

No próximo capítulo é feita uma breve descrição das bases de dados aplicadas às metodologias utilizadas na tese. É descrito também o processo de determinação da amostra.

3. DADOS: DETERMINAÇÃO DA AMOSTRA E BASES UTILIZADAS

No estudo da trajetória escolar da coorte de alunos das escolas amostradas, três bases de dados foram utilizadas: i) “Avaliação de desempenho: fatores associados”; ii) “Ficha Histórico Escolar” ou “Ficha B” e; iii) Censo Escolar de 1999. Nesse capítulo, inicialmente é feita uma descrição sucinta da determinação da amostra e, posteriormente, são feitos alguns comentários sobre os bancos de dados.

3.1 Descrição da amostra

A base “Avaliação de desempenho: fatores associados” foi desenhada visando atender à proposta de mensuração do impacto das intervenções do Plano de Desenvolvimento da Escola - PDE - sobre os resultados escolares dos sistemas educacionais beneficiários do programa Fundescola⁷. Fruto de uma parceria estabelecida entre o INEP e o CEDEPLAR, a pesquisa teve uma duração de cinco anos, sendo realizada em seis rodadas: a primeira coleta de dados ocorreu em abril de 1999, a segunda em novembro de 1999 e, a partir de então, as demais ocorreram também a cada mês de novembro, finalizando em 2003.

No ano de 1998⁸, definiu-se que a pesquisa seria realizada em dois estados de cada região atendida pelo Fundescola: Pará e Rondônia (Norte), Pernambuco e Sergipe (Nordeste) e Mato Grosso do Sul e Goiás (Centro-Oeste).

A unidade de análise da pesquisa foi o aluno. Entretanto, a unidade amostral foi a escola. A seleção do total de escolas foi baseada no número de matrículas efetivadas na 4ª série. Na definição do total de alunos que seriam avaliados, foram consideradas eventuais perdas desses alunos por reprovação, abandono ou transferência ao longo das séries e o volume de recursos necessários à realização da pesquisa, nas suas diversas fases. Baseando-se nesses dois parâmetros, estimou-se um total desejado de alunos na 4ª série equivalente a 12.000.

⁷ O Fundescola (Fundo de Fortalecimento da Escola) é um programa do Ministério da Educação, desenvolvido em parceria com as secretarias estaduais e municipais de educação. Na sua elaboração, foi estabelecido um conjunto de ações visando melhorar a qualidade das escolas do ensino fundamental, nas regiões Norte, Nordeste e Centro-Oeste. Objetivo: assegurar a permanência das crianças nas escolas públicas, nesse nível de ensino.

⁸ Nesse ano, o programa ainda não tinha sido implementado no Nordeste. Entretanto, optou-se pela inclusão dessa região na pesquisa, sendo que os resultados de 1999 seriam considerados como a *baseline* anterior à entrada do PDE.

Através da estimativa (baseada nos dados do Censo Escolar) da média de alunos matriculados na 4ª série e seguintes, no ensino fundamental, definiu-se um número ideal de escolas participantes igual a 160. Na realidade, o total de escolas variou entre 1999 e 2003, em função de fatores administrativos (escolas que deixaram de ofertar determinadas séries) e encerramento de convênio com a prefeitura (escolas deixando de ser municipais e passando a ser particulares), entre outros. Ressalta-se que na amostra foram incluídas tanto escolas que participavam do PDE quanto escolas que não participavam⁹.

Importante mencionar que, no processo de estratificação do universo, analisou-se o perfil censitário das 400 escolas participantes do PDE. Foram considerados fatores como o tamanho e as condições de funcionamento da escola, a dependência administrativa e algumas variáveis relacionadas ao desempenho do aluno. Dentre as 400 escolas, foram sorteadas 120 (Grupo PDE). Foi selecionado um número equivalente de escolas não participantes do Programa (Grupo Controle). De acordo com o Censo Escolar de 1998, essas 240 unidades escolares abrangiam cerca de 18.000 alunos matriculados nas 4ª séries do período diurno. Ou seja, um número muito superior ao previsto inicialmente (12.000 alunos). Assim, o total dessas unidades foi reduzido para 158 (52 com PDE e 106 sem PDE).

As escolas amostradas, situadas nas áreas urbanas das microrregiões dos seis estados brasileiros citados, pertenciam às dependências administrativas estaduais e/ou municipais, ofertando todas as séries do ensino fundamental (no período diurno), tendo pelo menos 200 alunos matriculados nestas séries.

Em 2001, várias escolas tornaram-se participantes do PDE. Nesse sentido, o objetivo principal da pesquisa deixou de ser a avaliação dos efeitos do Programa sobre os resultados escolares, passando a ser a identificação e análise dos fatores determinantes do desempenho escolar dos alunos sob acompanhamento.

No ano de 2003 foi realizada uma pesquisa de campo com o objetivo de reconstituir a trajetória escolar dos alunos que em 1999 participaram da pesquisa “Avaliação de desempenho: fatores associados”, respondendo às provas e aos questionários aplicados. Essa base de dados foi denominada “Ficha Histórico Escolar”, ou “Ficha B”, como será tratada a partir daqui. A construção dessa base foi importante uma vez que houve uma perda bastante expressiva de alunos participantes da pesquisa anterior, ao longo do período de acompanhamento, em função de fatores como transferência e evasão, entre outros.

⁹ Outras informações sobre o processo de amostragem podem ser acompanhadas em ANDRADE (1999).

A idéia da construção desse instrumento fundamentou-se na expectativa de realização de estudos mais aprofundados sobre o desempenho escolar do aluno, através do uso conjunto das informações de ambas as bases. É essa a intenção desta tese ao utilizar, pela primeira vez no país, o cruzamento da diversidade de dados coletados nas duas pesquisas realizadas. Uma rápida caracterização dessas bases é feita na seqüência.

3.2 Bases de dados

Foi utilizado, neste estudo, o banco de dados “Avaliação de desempenho: fatores associados”, resultado de uma parceria firmada entre o CEDEPLAR e o INEP. Trata-se do primeiro banco de dados longitudinais sobre o rendimento escolar no Brasil, sendo ainda pouco utilizado no meio acadêmico. Este banco de dados contém o resultado de testes (compostos por itens das provas do SAEB) de português e matemática, aplicados aos alunos. Além disso, contém também respostas de questionários socioeconômicos aplicados a alunos, professores e diretores, bem como características das escolas amostradas. Na tese foram utilizadas apenas parte das informações constantes na base, c

Um banco complementar, também utilizado, foi o “Ficha B”, estruturado de forma a reconstituir o histórico escolar dos alunos que fizeram parte da pesquisa “Avaliação de Desempenho: Fatores Associados”, iniciada no ano de 1999. É a primeira vez que as informações constantes neste banco de dados são utilizadas. Nele, ao fim de cada ano letivo consta o resultado escolar (aprovação, reprovação, transferência, afastamento por abandono, evasão e falecimento) de cada aluno acompanhado ao longo do período de 1999 a 2003. Esses resultados são mais bem detalhados adiante.

Em ambas as bases, foram coletadas informações anuais para o período de 1999 a 2003, para os alunos matriculados nas escolas do ensino fundamental componentes da amostra. O Quadro 1 mostra o formato da coleta de dados relacionado à primeira base citada. A descrição do quadro é feita na seqüência.

QUADRO 1
Esquema da coleta de dados da pesquisa “Avaliação de Desempenho:
Fatores Associados”

1999	2000	2001	2002	2003
A4 (1)				
	A5+B5 (2)			
		A6+B6+C6 (3)		
			A7+B7+C7+D7 (4)	
				A8+B8+C8+ D8+E8 (5)

Fonte: Relatórios do PDE.

Descrição do Quadro:

- (1) Alunos das escolas amostradas que em 1999 cursavam a 4ª série (A4);
- (2) Alunos da coorte A4 promovidos para a 5ª série (A5) + demais alunos da classe, transferidos de outras escolas e/ou repetentes da 5ª série (B5);
- (3) Alunos promovidos para a 6ª série (A6+B6) + demais alunos da classe transferidos e/ou repetentes (C6);
- (4) Alunos promovidos para a 7ª série (A7+B7+C7) + demais alunos da classe transferidos e/ou repetentes (D7);
- (5) Alunos promovidos para a 8ª série (A8+B8+C8+D8) + demais alunos da classe transferidos e/ou repetentes (E8).

O Quadro 2 retrata o processo de coleta de dados para o banco “Ficha B”.

QUADRO 2
Esquema da coleta de dados da pesquisa “Ficha B”

	Ano1	Ano2	Ano3	Ano4	Ano5	Ano6	Ano7	Ano8	Ano9	Ano10	Ano11	Ano12	Ano13
Ano (1)													
Série (2)													
Result.Final (3)													

Fonte: Relatórios do PDE.

No cabeçalho desta Ficha constam as seguintes informações: ID da escola, nome da escola, ID do aluno, nome e data de nascimento do aluno, turno e turma.

Descrição do Quadro:

(1) Na Ficha de cada aluno foi registrado o ano para o qual havia informação sobre a série cursada e o resultado obtido ao término do ano letivo. O ano 1 variou de 1988 a 1999, sinalizando que na base de dados havia alunos cujo primeiro ano para o qual havia registro relacionado à série cursada e resultado final variou de 1988 a 1999. Ou seja, para parte dos alunos foi possível reconstituir informação retrospectiva, no “histórico escolar”, para os anos anteriores ao ano base 1999. Entretanto, para outros alunos o ano base 1999 foi o primeiro ano com informações sobre série e resultado¹⁰.

(2) Na Ficha do aluno foi registrada a série cursada no ano (ano1 a ano13). Esta série variou de 0 a 9, equivalendo a pré-escola e 1º ano do ensino médio, respectivamente. Contudo, existem pouquíssimos casos relacionados a essas séries “extremas”;

(3) Ao término de cada ano letivo, para cada série foi anotada a ocorrência dos seguintes eventos: aprovação (A), reprovação (R), transferência¹¹ (T) ou evasão (E) escolar para os alunos sob acompanhamento. Conforme a situação de cada aluno, alguns destes eventos podem ter sido registrados de forma simultânea. Assim, numa série, num determinado ano, ele pode ter sido aprovado e transferido (AT), aprovado e evadido (AE), reprovado e transferido (RT), reprovado e evadido (RE), evadido e transferido (ET). Se o aluno faleceu ao longo do período, isso também constou na Ficha, para o ano de ocorrência do evento (F, AF, RF, EF, TF). Portanto, é possível verificar a ocorrência de diversas combinações dos eventos possíveis, ao longo da trajetória escolar.

A partir das informações constantes na ficha do aluno foi possível identificar a ocorrência de afastamento por abandono numa determinada série e num ano específico. Conforme será visto no capítulo 5, possivelmente houve algumas inconsistências nesse processo de identificação (devido ao confundimento entre abandono e evasão).

Outras informações coletadas referem-se ao tempo de permanência do aluno na escola 1, escola 2 e escola 3 (supondo-se transferências), a série de saída da escola, o total de repetências, o nome da escola para a qual o aluno foi transferido e, se ele retornou, o

¹⁰ Na ausência de casos de reprovação, afastamento por abandono, evasão e falecimento, e supondo registros completos para todos os alunos, na Ficha de cada aluno constariam informações que variariam do ano1 ao ano8 (1ª a 8ª série, respectivamente). Contudo, têm-se informações até o ano13, pois parcela significativa dos alunos experimentou tais eventos ao longo da sua trajetória escolar, estendendo assim a ficha de registro.

¹¹ Considerou-se transferência o fato do aluno ter sido transferido para outra escola, bem como ter sido transferido para o supletivo ou para a EJA (Educação de Jovens e Adultos).

nome da escola de retorno. A inclusão das questões tempo de permanência e total de repetências teve por objetivo evitar inconsistências relacionadas aos dados.

A junção das bases “Avaliação de desempenho: fatores associados” e “Ficha B”, quando aplicadas às metodologias de análise, resulta numa contribuição bastante valiosa ao entendimento da problemática da repetência e da evasão no sistema educacional, considerando-se suas características específicas e sua complementaridade.

Entretanto, apesar das diversas variáveis contidas nessas bases, que permitem responder a uma série de questões presentes na tese, um terceiro banco de dados foi utilizado: o Censo Escolar de 1999. Desse banco, foram trabalhadas informações relacionadas à estrutura física da escola (total de salas, presença de quadra, laboratório de ciência e informática, biblioteca, sala de tv e vídeo), aos recursos humanos existentes (total de professores segundo a formação) e ao acesso à escola (total de matrículas segundo a série e situação de aprovação/reprovação). A inclusão dessas informações nos modelos possibilitou uma melhor avaliação do papel dos fatores escolares sobre a ocorrência da repetência e da evasão para os alunos da coorte.

Para concluir essa seção, é preciso ressaltar que o desenvolvimento dessa tese está baseado no acompanhamento da coorte A4 (primeira coluna do Quadro 1), ou seja, são acompanhados até 2003 somente os alunos matriculados nas escolas selecionadas no ano inicial de 1999. Assim, novos alunos (transferidos de outras escolas, repetentes de outras séries) não foram inseridos no estudo, nos anos subsequentes. A Ficha B (Quadro 2) consistiu na reconstituição do histórico escolar da coorte A4. Informações sobre a junção de ambas as bases serão prestadas no capítulo 5.

O próximo capítulo aborda os métodos utilizados, tendo em vista a resposta aos questionamentos feitos na parte inicial desta tese.

4. ARCABOUÇO METODOLÓGICO: ANÁLISE DE SOBREVIVÊNCIA E MODELO HIERÁRQUICO LOGÍSTICO LONGITUDINAL

Para o desenvolvimento deste estudo são utilizadas duas metodologias, a análise de sobrevivência (método não-paramétrico) e o modelo hierárquico logístico longitudinal (método paramétrico). O primeiro método visa a estimação e análise das funções de sobrevivência do aluno por UF's, no período de 1999 a 2003, segundo os eventos reprovação, repetência, afastamento por abandono e evasão; e a situação de defasagem idade-série na 4ª série do ensino fundamental. O segundo método tem por objetivo a identificação e análise dos determinantes da probabilidade de ocorrência da repetência e da evasão. Neste caso, verifica-se que os eventos básicos são as variáveis indicadoras de fluxo na trajetória escolar do aluno. Na fase de identificação desses determinantes, considera-se o efeito de variáveis fixas e de variáveis que variam com o tempo, tanto relacionadas ao aluno quanto à respectiva família e escola. Destaque é dado, entre as covariáveis, ao resultado escolar obtido pelo aluno no ano que antecede a ocorrência dos eventos de interesse.

No Brasil, essas metodologias ainda não foram utilizadas na área da educação, sendo as variáveis básicas a repetência e a evasão (além da reprovação e do afastamento por abandono), devido às especificidades dos raros bancos de dados longitudinais existentes. Deve-se ressaltar que o trabalho realizado por GOMES-NETO & HANUSHEK (1996) baseou-se em dados longitudinais, para o período de 1981 a 1985. Contudo, o banco de dados não foi construído de forma a identificar as causas e efeitos da repetência (evento analisado). Além disso, foram as escolas os agentes acompanhados ao longo do tempo, e não os alunos. Com isso, há uma perda importante de informações: por exemplo, perdem-se informações sobre os alunos que estavam na escola em 1981 e não estavam em 1983. Portanto, a aplicação das técnicas propostas, aplicadas aos eventos citados, constitui fato inédito no país.

4.1 Análise de sobrevivência

Estudos têm apontado a importância da metodologia da análise de sobrevivência nas pesquisas do ciclo de vida dos indivíduos (BLOSSFELD, HAMERLE & MAYER, 1989; MAYER & TUMA, 1990; TRUSSELL, HANKINSON & TILTON, 1992). Entretanto, a utilização desta técnica é bastante limitada pela escassez de bases de dados longitudinais.

Uma das bases de dados utilizada na tese - a “Ficha B” - contém informações prospectivas da coorte de alunos sob estudo. Esta coorte foi acompanhada por um período de 5 anos (desde 1999 a 2003). São alunos matriculados na 4ª série do ensino fundamental, no ano de 1999, e que deveriam concluir a 8ª série em 2003, pertencentes à escolas situadas nas áreas urbanas das microrregiões dos seis estados focalizados pelo PDE: Pernambuco, Sergipe, Pará, Rondônia, Goiás e Mato Grosso do Sul.

Na ficha do aluno foi registrado, ao fim de cada ano letivo, o resultado escolar ou evento ocorrido. Abaixo, são listados os resultados e/ou eventos que constam nesta ficha:

- i) aprovação;
- ii) reprovação;
- iii) transferência (para outra escola ou para a EJA/Supletivo Seriado);
- iv) afastamento por abandono e;
- v) falecimento.

Constata-se que é possível saber, ao término de cada ano letivo, se o aluno foi aprovado, reprovado, transferido, afastado por abandono ou se faleceu. No ano seguinte, os integrantes da coorte que permaneceram na escola podem vivenciar as seguintes situações: promoção, repetência, transferência, evasão e falecimento. Essas situações são vistas de forma mais detalhada no capítulo 6.

Os eventos considerados nesta metodologia de análise são aqueles enfocados no primeiro capítulo da tese, ou seja, a reprovação, a repetência, o afastamento por abandono e a evasão. Mais especificamente, a idéia é estimar e analisar as curvas de sobrevivência a cada evento, entre a 4ª e a 8ª série do ensino fundamental. Como alguns desses eventos se repetiram ao longo da trajetória escolar do aluno (reprovação, repetência e abandono), será considerado apenas o tempo até a sua ocorrência, pela primeira vez, no período sob estudo.

Dois aspectos relacionados aos dados devem ser mencionados:

- i) ao fim do período de acompanhamento parte dos alunos não experimentou os eventos de interesse e;
- ii) entre 1999 e 2003 houve uma redução significativa da coorte. Houve diversos casos de transferência e evasão, além de alguns casos de óbitos. Ou seja, o acompanhamento do aluno foi interrompido.

Tanto na situação (i) quanto na situação (ii) os dados são censurados. A censura corresponde à observação parcial da resposta. Mas, apesar de incompletas, as observações censuradas dão informações sobre o tempo de sobrevivência dos alunos. É a presença de dados censurados que requer o uso do método da análise de sobrevivência.

Uma vez que a análise de sobrevivência está relacionada ao tempo de ocorrência do evento, quando é atribuído um valor a esse tempo, na realidade é escolhida uma escala de tempo e uma origem. Neste estudo a escala é anual (anos 1, 2, 3, 4, 5), referindo-se aos anos de 1999 a 2003. Por outro lado, a origem foi determinada a partir da efetivação da matrícula dos alunos nas escolas da amostra, na 4ª série do ensino fundamental, no ano de 1999.

A variável-resposta é o tempo (T) desde a matrícula do aluno na 4ª série, em 1999, até a ocorrência dos eventos básicos, ou seja, até a ocorrência da primeira reprovação, da primeira repetência, do primeiro afastamento por abandono e da evasão, entre 1999 e 2003. Verifica-se que a ocorrência dos eventos está concentrada em poucos pontos no tempo (T=1 a T=5). Neste caso, a variável-resposta é medida num tempo discreto.

O tempo de ocorrência (T) é uma variável aleatória que tem uma distribuição de probabilidades. Na análise de sobrevivência esta variável é especificada pela função de sobrevivência¹². Antes de apresentar esta função, será especificada a função de densidade acumulada:

$$F(t) = \Pr (T \leq t). \quad (3)$$

Essa função dá a probabilidade de uma variável T ser menor ou igual a um determinado valor t, sendo t qualquer número não negativo. A função de sobrevivência é estreitamente relacionada à função de densidade acumulada, sendo assim definida:

$$S(t) = \Pr (T \geq t) = 1 - F(t). \quad (4)$$

Se o objetivo do estudo é determinar a probabilidade do aluno permanecer no estado de aprovado desde o momento em que é acompanhado até o fim do período sob análise, a função de sobrevivência dá a probabilidade de sobrevivência após o tempo t. Nesse estudo, as funções de sobrevivência são estimadas utilizando-se o método Kaplan-Meier. Trata-se de um estimador não-paramétrico, assim definido:

$$\hat{S}(t) = \prod_{j:t_j \leq t} \left(\frac{1 - r_j}{n_j} \right) \quad (5)$$

Sendo:

t o tempo de ocorrência do evento;

n_j os alunos sob o risco de um determinado evento (ainda não experimentaram o evento e nem foram censurados até o tempo t_j);

r_j o número de eventos ocorridos no tempo t_j .

¹² A variável T é especificada também pela taxa de risco.

Após a estimação das funções de sobrevivência é utilizado o teste de Log-rank (também conhecido como o teste Mantel-Haenszel) para testar a hipótese nula de que essas funções são iguais para as seis UF's.

Formalização do teste: tendo-se duas funções de sobrevivência $S_1(t)$ e $S_2(t)$, equivalentes a dois grupos de tratamento, o teste é utilizado para testar se $S_1(t) = S_2(t)$ para todo t . A estatística de teste pode ser escrita como uma função dos desvios dos números observados em relação aos números esperados dos eventos. Para o grupo 1, a estatística de teste pode ser assim expressa:

$$\sum_{j=1}^r (d_{1j} - e_{1j}) \quad (6)$$

Sendo:

- r o número de vezes de ocorrência do evento,
- d_{ij} o número de eventos que ocorreram no grupo 1 no tempo j ,
- e_{ij} o número esperado de eventos no grupo 1 no tempo j .

O número esperado de eventos é dado por $\frac{n_{1j}d_j}{n_j}$

Sendo:

- n_j o número total de casos que estão sob risco antes do tempo j ,
- n_{1j} o número de casos sob risco antes do tempo j no grupo 1 e,
- d_j o número total de casos no tempo j em ambos os grupos.

Ressalta-se que a generalização do teste de Log-Rank para se testar a igualdade de mais de duas funções de sobrevivência é facilmente realizada. Para o caso de duas funções de sobrevivência, em cada tempo de falha t_j os dados são dispostos em forma de uma tabela de contingência 2×2 com d_{ij} falhas e $n_{ij} - d_{ij}$ sobreviventes na coluna i . Para o caso de mais de duas funções de sobrevivência, ter-se-á uma tabela de contingência de $2 \times r$ com d_{ij} falhas e $n_{ij} - d_{ij}$ sobreviventes na coluna i .

Este teste foi utilizado para comparar as curvas de sobrevivência aos eventos de interesse. A pretensão é identificar possíveis diferenciais por escolas/UF's, relacionados ao comportamento dos alunos frente à tais eventos.

Uma consideração adicional deve ser feita: no caso da trajetória escolar, e mais especificamente deste estudo, parte-se de uma situação inicial (efetivação da matrícula na 4ª série em 1999), sendo possível a ocorrência de diversos eventos ao término do ano

letivo: aprovação, reprovação, afastamento por abandono, transferência e morte. No ano letivo seguinte o aluno pode evadir do sistema (não realizar sua matrícula). Além disso, a partir dos resultados observados num determinado ano o aluno está sujeito às seguintes situações no ano seguinte: promoção e repetência (além das demais situações já expostas). Portanto, existem diversos riscos competindo entre si. Contudo, como o interesse está centrado apenas na ocorrência de um tipo de evento por vez (ex.: reprovação), consideram-se os eventos competindo como censurados.

4.2 O modelo de regressão hierárquico logístico longitudinal

Os alunos da coorte sob estudo foram acompanhados durante cinco anos letivos, entre 1999 e 2003. Ao longo desse período, esses alunos estiveram expostos a diversos resultados e/ou eventos escolares. Entre esses eventos, como dito anteriormente, o interesse está centrado na repetência e na evasão. Pretende-se, com a estimação dos modelos, identificar e analisar os principais determinantes desses problemas, ocorridos entre a 4ª e a 8ª série do ensino fundamental.

Por ser um estudo de trajetória, as mudanças ocorridas ao longo do tempo, associadas ao aluno e à escola, devem ser incorporadas ao modelo. A esse respeito, BRYK E RAUDENBUSH (2002) ressaltam a existência de três problemas metodológicos relacionados ao tratamento de dados longitudinais: i) os estudos sobre mudanças (na trajetória) dos indivíduos raramente identificam um modelo explícito de crescimento do indivíduo; ii) estes estudos geralmente usam instrumentos que são desenvolvidos para tratar os indivíduos em pontos fixos no tempo, desconsiderando-se as diferenças nas taxas de mudanças entre os indivíduos, fazendo com que o pressuposto de uma variância constante no tempo tenha sérias implicações nos estudos de mudança e dos determinantes dessas mudanças e; iii) a maioria desses estudos coleta dados em apenas dois pontos no tempo, resultando na inadequação dos estudos do crescimento do indivíduo.

Os pontos ressaltados pelos autores mostram a importância de uma especificação correta dos modelos de trajetória de indivíduos. Um dos bancos de dados utilizados nesta tese - Ficha B - tem informações dos resultados dos alunos ao fim de cada ano letivo. O outro banco de dados - Fatores associados: avaliação de desempenho - tem informações sobre os alunos, famílias e escolas, entre outras. Variáveis complementares, relacionadas à escola, também foram extraídas do Censo Escolar de 1999. Existem variáveis consideradas fixas, como o sexo e a raça/cor do aluno e variáveis que variam no tempo, como a situação

de trabalho e a proficiência do aluno. O modelo precisa ser especificado de forma a incorporar essas questões.

4.2.1 Especificação do modelo incondicional

Os modelos de regressão hierárquicos partem do pressuposto que há uma estrutura hierárquica nos dados. Cada um dos níveis na estrutura hierárquica é representado por um sub-modelo, que expressa o relacionamento entre as variáveis dentro de um determinado nível, além de especificar como as variáveis num nível influenciam os relacionamentos que ocorrem noutros níveis.

Nessa tese, os modelos hierárquicos estimados são logísticos, sendo que as variáveis-resposta referem-se à probabilidade de primeira repetência e de evasão entre a 4ª e a 8ª série do ensino fundamental. Para modelar essas probabilidades, foi utilizado o valor “1” para denotar a ocorrência do evento (repetência, evasão) e o valor “0” caso contrário.

Para verificar a dimensão da variação entre as escolas na probabilidade de repetência (e de evasão) num determinado ano letivo t , inicialmente os modelos de nível 1 e 2 são estimados sem a inclusão de covariáveis.

O modelo de nível 1 é assim especificado:

$$\eta_{ij} = \beta_{0j} \quad (7)$$

No nível 2 a equação correspondente é:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (8)$$

A substituição de (8) em (7) resulta em:

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

Sendo:

$i = 1, 2, \dots, I$ unidades de nível 1 (alunos); $j = 1, 2, \dots, J$ unidades de nível 2 (escolas);

η_{ij} = o log da chance de sucesso (ocorrência do evento) para o aluno i , na escola j ;

γ_{00} = o parâmetro da parte fixa do modelo, que corresponde ao *log-odds* médio de repetência (ou evasão) entre as escolas;

u_{0j} = o efeito aleatório relacionado ao nível 2 (das escolas).

Através da estimação do modelo incondicional é possível estimar a variabilidade relacionada com cada um dos níveis. No modelo logístico, cujos erros aleatórios seguem

uma distribuição binomial, a variância do nível 1 não é constante. Em geral, considera-se o valor $\pi^2/3 = 3.29^{13}$. Nesse caso, o coeficiente de correlação intra-níveis, é assim calculado:

$$\frac{u_{oj}}{u_{oj+(\pi^2/3)}} \quad (10)$$

Parte dessa variabilidade pode ser explicada através da inclusão de covariáveis nos diferentes níveis.

No caso específico dessa tese as equações anteriores serão estendidas para incorporar três níveis que refletem a estrutura de dependência existente nos dados longitudinais utilizados. O nível intra-alunos (nível 1) refere-se às mudanças relacionadas ao aluno ocorridas durante o período. Nesse nível, a dependência entre as observações ocorre em função do mesmo aluno ser medido “n” vezes ao longo do estudo. O nível inter-alunos (nível 2) possibilita verificar quais mudanças individuais ao longo do tempo diferem entre os alunos em função de suas características fixas. Nessa estrutura hierárquica, os fatores intra-alunos estão aninhados nos fatores inter-alunos. Como os alunos estão aninhados dentro das escolas, tem-se um terceiro nível relacionado aos fatores escolares (WILLETT, 1997). Nesse sentido, novas especificações são apresentadas com a inclusão das variáveis nesses níveis.

4.2.2 Especificação do modelo condicional

4.2.2.1 Variável-resposta: primeira repetência

No nível 1, a trajetória temporal do evento pode ser acompanhada pela inclusão da variável Tempo (T), associada à ocorrência da primeira repetência entre a 4ª e a 8ª série. Com essa inclusão é possível fazer um importante diagnóstico do padrão temporal da repetência escolar. Tem-se a seguinte equação inicial:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}T_{ij} \quad (11)$$

Nesse nível, além da inclusão do tempo associado à ocorrência do evento, são incluídas outras variáveis relacionadas aos alunos que explicam a variação em η_{ij} . São as variáveis que variam ao longo do tempo (ex.: situação de trabalho, proficiência). Denotando-se esse vetor de covariáveis por X, tem-se a equação:

¹³ Ver: Snijders e Boske (1999).

$$\eta_{ij} = \beta_{0j} + \beta_{1j}T_{ij} + \beta_{2j}X_{ij} \quad (12)$$

Onde os coeficientes β_{0j} , β_{1j} e β_{2j} neste modelo de nível 1 tornam-se as variáveis-resposta nas equações para o nível 2.

Entretanto, parte-se do pressuposto de que o intercepto, o parâmetro da trajetória temporal e o parâmetro relacionado às covariáveis que variam ao longo do tempo variam entre os alunos em função das características desses alunos (características que não mudam no tempo, como o sexo e a raça/cor). Chamando-se o vetor com essas características fixas de Z , o modelo de nível 2 é assim re-especificado:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}T_{ij} + \beta_{2j}X_{ij} + \beta_{3j}Z_{ij} \quad (13)$$

Contudo, estes alunos estão aninhados dentro de escolas. Portanto, é preciso incluir um terceiro nível no modelo para captar a variabilidade existente entre as escolas. Cada coeficiente “ β ” (β_{0j} , β_{1j} , β_{2j} e β_{3j}) presente nos níveis 1 e 2 se torna uma variável-resposta no modelo de nível 3. Nesse nível, pode-se (ou não) incluir um termo aleatório associado à cada β .

Nas equações de nível 3, listadas abaixo, inicialmente não é considerada a presença de covariáveis (modelo incondicional), além de se considerar que os coeficientes associados a T_{ij} , X_{ij} e Z_{ij} têm um efeito aleatório neste nível.

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (14a)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (14b)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (14c)$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \quad (14d)$$

Após a estimação do modelo incondicional é incluído um vetor de covariáveis (aqui denotado por W). Tem-se um novo conjunto de equações, onde cada coeficiente β é definido como tendo um efeito fixo ou aleatório. De uma forma geral, cada coeficiente de nível 1 e nível 2 pode ser modelado no nível 3 de três formas: como efeito fixo, variando aleatoriamente e variando de forma não aleatória. Na especificação do modelo final, atenção deve ser dada à escolha adequada destas formas. A inclusão do vetor W no nível 3 resulta nas equações a seguir:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \quad (15a)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j} \quad (15b)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}W_j + u_{2j} \quad (15c)$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}W_j + u_{3j} \quad (15d)$$

Sendo:

β_{0j} o intercepto e β_{1j} , β_{2j} e β_{3j} os coeficientes das variáveis explicativas de nível 1 e 2, considerados como tendo efeitos aleatórios e;

u_{0j} , u_{1j} , u_{2j} e u_{3j} os efeitos aleatórios de nível 3.

Substituindo-se as equações 15a a 15d em (12) tem-se o modelo final:

$$\eta_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}T_{ij} + \gamma_{20}X_{ij} + \gamma_{30}Z_{ij} + \gamma_{11}T_{ij}W_j + \gamma_{21}X_{ij}W_j + \gamma_{31}Z_{ij}W_j + u_{0j} + u_{1j}T_{ij} + u_{2j}X_{ij} + u_{3j}Z_{ij} \quad (16)$$

As estimativas dos parâmetros da equação 15 permitem que o termo η_{ij} seja transformado no logaritmo da chance de sucesso para o aluno i , de forma que:

$$\eta_{ij} = \log\left(\frac{\varphi_{ij}}{1 - \varphi_{ij}}\right) \quad (17)$$

Neste caso, o valor de φ_{ij} está compreendido entre 0 e 1, para qualquer valor de η_{ij} .

Um aspecto interessante é que o logaritmo da chance de sucesso pode ser transformado no valor previsto da probabilidade aplicando-se o exponencial a η_{ij} . Tem-se a seguinte expressão:

$$\varphi_{ij} = \frac{1}{1 + \exp(-\eta_{ij})} \quad (18)$$

4.2.2.2 Variável-resposta: evasão

Para fins comparativos, nos modelos cuja variável-resposta é a evasão é interessante incluir as mesmas variáveis e seguir basicamente o mesmo procedimento descrito no tópico referente à repetência. Assim, de forma análoga ao caso anterior, a trajetória temporal da evasão pode ser descrita pela equação de nível 1 a seguir:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}T_{ij} \quad (19)$$

Como os estudos existentes têm apontado que em geral a evasão acontece após uma ou sucessivas reprovações na série é conveniente incluir na regressão a variável aprovação,

para fins de controle. É importante analisar também o efeito da aprovação ao longo do tempo sobre a ocorrência do evento. Uma forma de verificar esse efeito é mediante a inclusão de um termo de interação entre as variáveis aprovação e tempo. A equação resultante tem o seguinte formato:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}T_{ij} + \beta_{2j}A_{ij} + \beta_{3j}A.T_{ij} \quad (20)$$

Nesse nível são incluídas também as variáveis relacionadas aos alunos que explicam a variação em η_{ij} . Como no caso da repetência, o vetor das variáveis que variam ao longo do tempo foi denominado X. Com a inclusão desse vetor, tem-se a presente equação:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}T_{ij} + \beta_{2j}A_{ij} + \beta_{3j}A.T_{ij} + \beta_{4j}X_{ij} \quad (21)$$

Os coeficientes β_{0j} , β_{1j} , β_{2j} , β_{3j} e β_{4j} neste modelo de nível 1 tornam-se as variáveis-resposta nas equações para o nível 2.

Como no caso anterior, o pressuposto é que o intercepto, o parâmetro da trajetória temporal e o parâmetro relacionado ao vetor de variáveis que variam ao longo do tempo variam entre os alunos em função das suas características fixas. Denotando-se o vetor com as características fixas de Z, o modelo de nível 2 é assim re-especificado:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}T_{ij} + \beta_{2j}A_{ij} + \beta_{3j}A.T_{ij} + \beta_{4j}X_{ij} + \beta_{5j}Z_{ij} \quad (22)$$

Ao incluir os fatores associados à escola cada coeficiente “ β ” da equação 22 torna-se uma variável-resposta no modelo de nível 3. As equações desse nível são apresentadas na seqüência, sem a inclusão de variáveis (modelo incondicional), sendo que os coeficientes associados a T_{ij} , A_{ij} , $A.T_{ij}$, X_{ij} e Z_{ij} estão sendo considerados como tendo um efeito aleatório neste nível.

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (23a)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (23b)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (23c)$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \quad (23d)$$

$$\beta_{4j} = \gamma_{40} + u_{4j} \quad (23e)$$

$$\beta_{5j} = \gamma_{50} + u_{5j} \quad (23f)$$

Após a estimação do modelo incondicional, a inclusão do vetor com as variáveis de nível 3 permite verificar o efeito dos fatores escolares sobre a evasão. As equações com esse vetor, denominado W, são apresentadas abaixo:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \quad (24a)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j} \quad (24b)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}W_j + u_{2j} \quad (24c)$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}W_j + u_{3j} \quad (24d)$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}W_j + u_{4j} \quad (24e)$$

$$\beta_{5j} = \gamma_{50} + \gamma_{51}W_j + u_{5j} \quad (24f)$$

A substituição dessas equações na equação 22 resulta no modelo final:

$$\eta_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}T_{ij} + \gamma_{20}A_{ij} + \gamma_{30}A.T_{ij} + \gamma_{40}X_{ij} + \gamma_{50}Z_{ij} + \gamma_{11}T_{ij}W_j + \gamma_{21}A_{ij}W_j + \gamma_{31}A.T_{ij}W_j + \gamma_{41}X_{ij}W_j + \gamma_{51}Z_{ij}W_j + u_{0j} + u_{1j}T_{ij} + u_{2j}A_{ij} + u_{3j}A.T_{ij} + u_{4j}X_{ij} + u_{5j}Z_{ij} \quad (25)$$

De forma análoga ao caso anterior, o logaritmo da chance de sucesso para o aluno i e o valor previsto da probabilidade são, respectivamente:

$$\eta_{ij} = \log\left(\frac{\varphi_{ij}}{1-\varphi_{ij}}\right) \quad \text{e} \quad \varphi_{ij} = \frac{1}{1+\exp\{-\eta_{ij}\}} \quad (26)$$

A estimação dos modelos de regressão foi feita através do software MLWIN, versão 1.1. No capítulo seguinte são abordadas as estratégias empíricas aplicadas aos eventos de interesse e às diversas variáveis, para adequação às técnicas de análise.

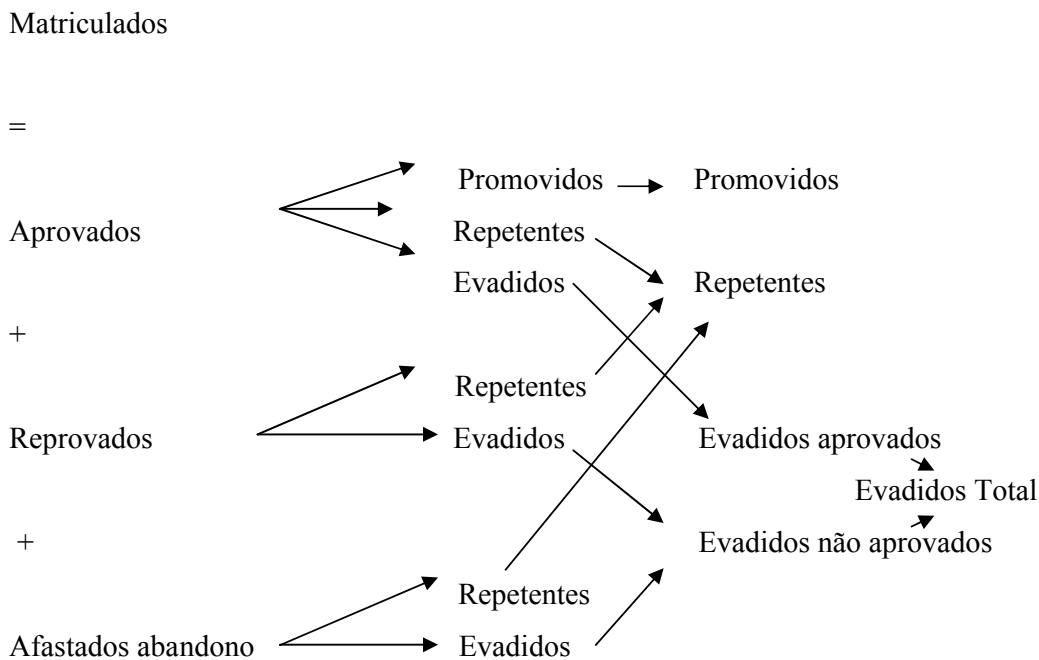
5. EVENTOS BÁSICOS E COVARIÁVEIS: ESTRATÉGIAS EMPÍRICAS PARA ADEQUAÇÃO ÀS TÉCNICAS DE ANÁLISE

Este capítulo tem por objetivo trazer maiores esclarecimentos sobre os eventos básicos considerados na análise de sobrevivência e na análise hierárquica, ocorridos entre os alunos da coorte estudada. Uma vez que um fluxo de alunos é acompanhado por um período de cinco anos e, considerando-se que a ocorrência de um evento num determinado ano algumas vezes está relacionada à ocorrência de outro num ano anterior, é preciso definir precisamente tais eventos. Ademais, devido à algumas particularidades relacionadas ao processo de coleta de dados, na época da realização da pesquisa de campo foram necessárias algumas estratégias para o tratamento estatístico dos dados.

5.1 Eventos de interesse: estratégias empíricas aplicadas aos dados da “Ficha B”

Para uma melhor definição e entendimento dos eventos de interesse, será utilizado o modelo de fluxo proposto por KLEIN (1995). Segundo o autor, num sistema fechado¹⁴, no fim do ano letivo um aluno matriculado numa escola é considerado aprovado, reprovado ou afastado por abandono. Ele apresenta o seguinte esquema:

FIGURA 2
Modelo de fluxo escolar



Fonte: KLEIN (1995)

¹⁴ Sistema em que um aluno admitido numa escola após a data de ocorrência da matrícula inicial é um aluno transferido do próprio sistema, ou seja, já foi contado na matrícula inicial do sistema.

Pode ser verificado que os alunos matriculados nesse sistema num determinado ano letivo t correspondem ao somatório dos alunos aprovados, reprovados e afastados por abandono. No ano $t+1$ as seguintes situações podem ocorrer:

- i) um aluno que foi aprovado na série k no ano t pode ser promovido à série $k+1$ (aprovado promovido), pode repetir a série k (repetente aprovado) ou pode evadir do sistema (evadido aprovado);
- ii) um aluno que foi reprovado na série k no ano t pode repetir a série k (repetente reprovado) ou evadir do sistema (evadido reprovado) e;
- iii) um aluno que foi afastado por abandono na série k no ano t pode repetir a série ou evadir do sistema.

No caso da tese proposta poder-se-ia supor um sistema fechado, uma vez que uma coorte de alunos, matriculada na 4ª série em 1999, é acompanhada até o ano de 2003 quando deveria concluir o ensino fundamental, não havendo entrada de novos alunos no sistema nos anos subseqüentes.

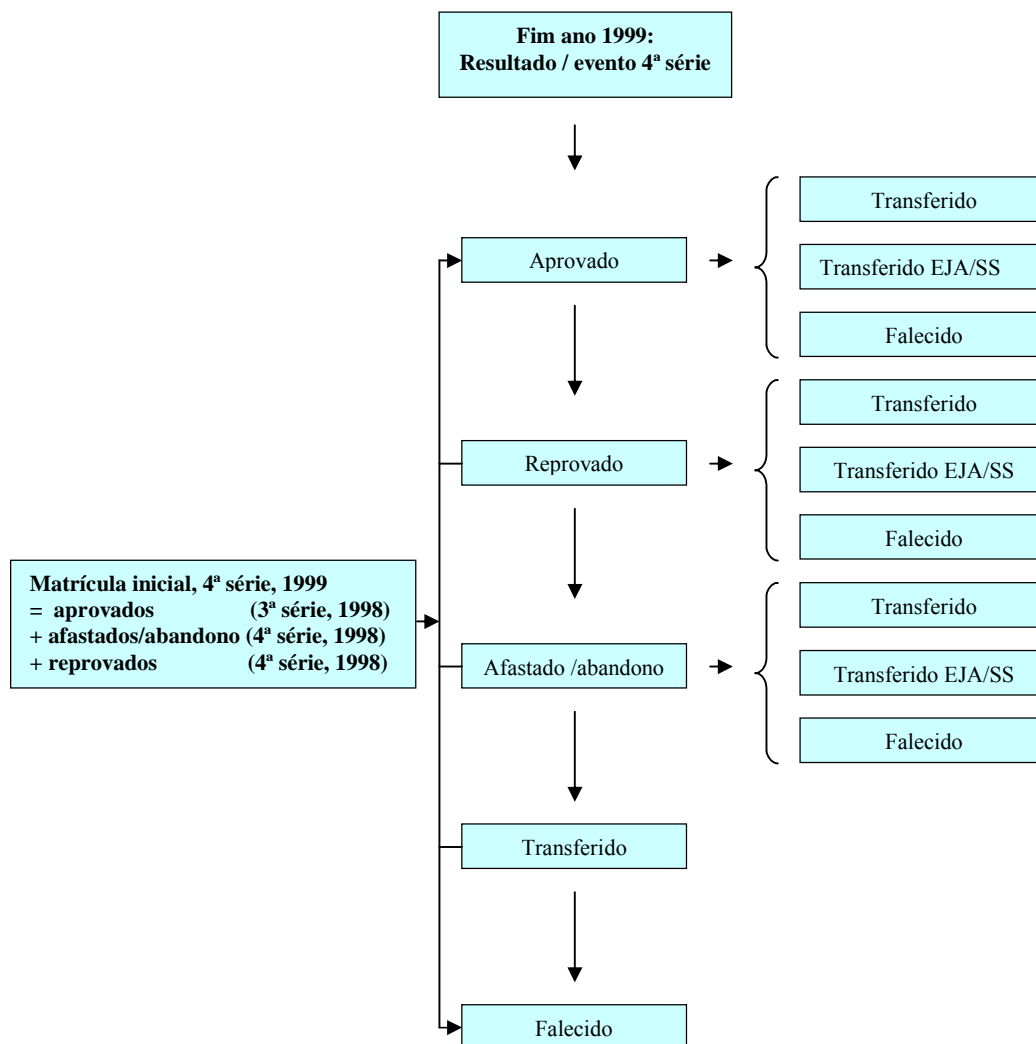
Na FIG. 3 são retratadas as informações constantes na “Ficha B”, relacionadas ao resultado obtido pelo aluno ao término de cada ano letivo. Enquanto no modelo do KLEIN (1995) são apresentados os eventos possíveis na trajetória escolar do aluno no ano $t+1$, associados ao resultado ocorrido no ano t , no próximo modelo são apresentados os resultados (ou eventos) registrados na ficha escolar do aluno ao fim de um determinado ano letivo t . O esquema retrata as especificidades dos dados da “Ficha B”, ou seja, inclui os resultados / eventos observados e registrados na ficha dos alunos das escolas amostradas ao fim do ano letivo.

O modelo se aplica a todos os anos letivos para o período de 1999 a 2003. Entretanto, a explicação referente ao total de matriculados será feita para o ano base de 1999, que marca o início do período de acompanhamento da coorte. O total de matriculados na 4ª série em 1999 corresponde ao somatório: i) dos alunos aprovados na 3ª série no ano de 1998, promovidos para a 4ª série; ii) dos alunos matriculados na 4ª série em 1998, que abandonaram os estudos ao longo do ano, retornando em 1999 e; iii) dos alunos reprovados na 4ª série em 1998, sendo repetentes na série em 1999.

Pelo esquema verifica-se que ao término do ano letivo t o aluno matriculado pode ser considerado aprovado, reprovado, afastado por abandono, transferido ou falecido. Na “Ficha B”, no campo “Resultado final” para a série cursada k , esses eventos podem ter sido registrados conforme a primeira parte do esquema ou podem ter sido registrados junto com um evento complementar (segunda parte do esquema). Conforme visualizado, o aluno

aprovado ou reprovado num determinado ano t pode, ao fim deste ano, solicitar transferência para outra escola ou para a EJA / Supletivo Seriado ou pode vir a falecer. Nesse caso, essas informações constam na ficha no ano t, juntamente com a informação de aprovação ou reprovação. O aluno afastado por abandono é aquele que se matriculou no ano t, mas ao fim do ano não foi considerado aprovado ou reprovado, em função de ter abandonado os estudos durante o ano. No fim do ano letivo t, junto com o evento abandono pode constar a informação de transferência ou falecimento. Finalmente, na ficha do aluno, ao término do ano t podem constar apenas os registros “transferido” e “falecido”.

FIGURA 3
Resultados (eventos) registrados na “Ficha B”



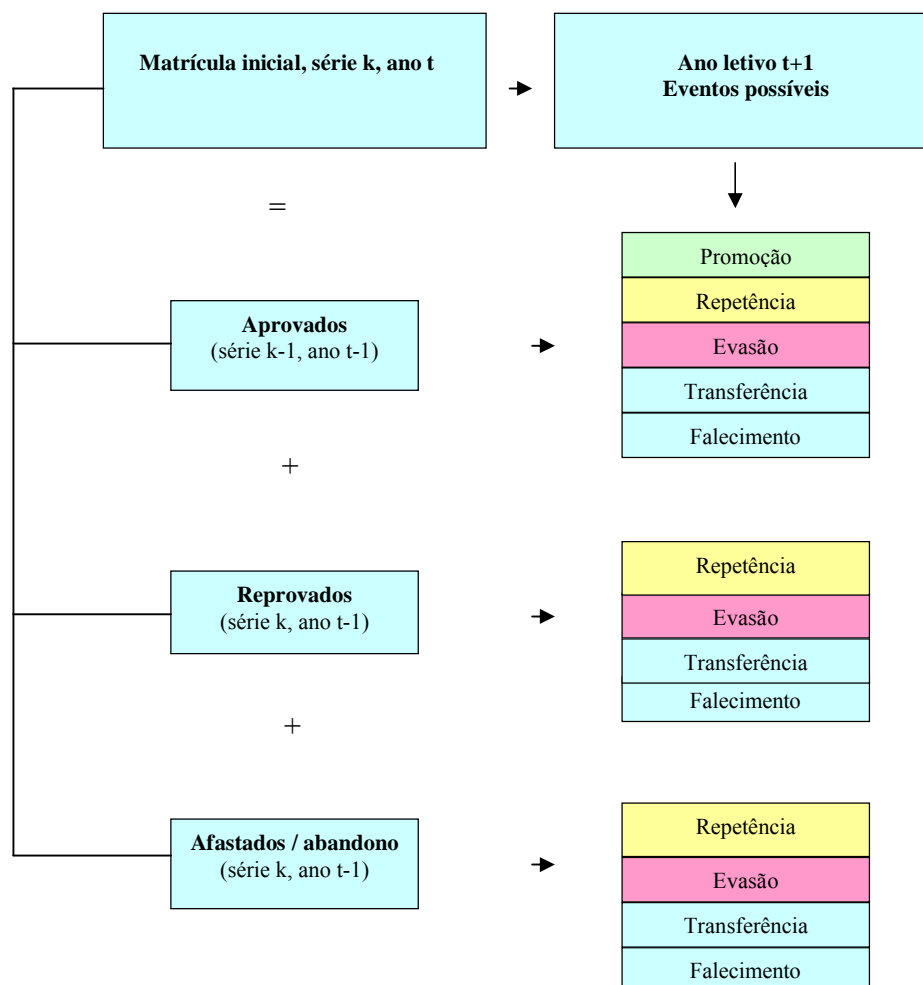
Fonte: Elaboração própria a partir das informações da Ficha B, anos 1999 a 2003.
Nota: EJA: Educação de Jovens e Adultos; SS: Supletivo Seriado.

Importante ressaltar que, ainda que o aluno tenha sido transferido para outra escola que também faz parte da amostra, por questões metodológicas optou-se por não dar seqüência ao seu acompanhamento no sistema. Para os alunos transferidos para a EJA ou para o Supletivo Seriado foram aproveitadas as informações (resultados finais) constantes na Ficha B anteriores à transferência. Para o ano específico de 1999 não foi caracterizada nenhuma evasão (mas sim abandono).

A FIG. 3 mostrou os resultados registrados na “Ficha B” ao término do ano de 1999. Na figura seguinte será mostrado o fluxo escolar do aluno, considerando-se dois anos consecutivos, t e $t+1$, tal como visto no modelo proposto por KLEIN (1995). A diferença em relação àquele modelo é a inclusão dos eventos “transferência” e “falecimento” no ano $t+1$.

De forma similar ao esquema anterior, na FIG. 4 o total de matrículas no ano t corresponde ao somatório: i) dos alunos aprovados na série $k-1$ no ano $t-1$, promovidos para a série k ; ii) dos alunos reprovados na série k em $t-1$, sendo repetentes na série em t ; iii) dos alunos matriculados na série k no ano $t-1$, que abandonaram os estudos durante o ano, retornando no ano t . Essas informações podem ser vistas na primeira parte da figura.

Figura 4
Fluxo escolar, segundo registros da “Ficha B”



Fonte: Elaboração própria a partir das informações da Ficha B, anos 1999 a 2003.

O diagrama exclui os alunos transferidos no ano t-1 e os que evadiram no próprio ano t (uma vez que não efetivaram sua matrícula no sistema escolar).

Observa-se que para os alunos aprovados na série k no ano t podem ocorrer as seguintes situações em t+1: i) promoção à série k+1; ii) repetência da série k; iii) evasão do sistema; iv) transferência (para outras escolas ou para a EJA / supletivo) e; v) falecimento. Os alunos reprovados ou afastados por abandono têm um fluxo similar, com exceção do evento “promoção”. Na figura estão destacados os eventos que indicam transição escolar. Noutras palavras, a ocorrência desses eventos em t+1 está relacionada à ocorrência de outro evento no ano anterior, t.

Entre os eventos apresentados, dois merecem uma particular atenção: o afastamento por abandono e a evasão. Para uma melhor compreensão desses eventos é importante frisar a diferença conceitual existente.

Considera-se afastamento por abandono a situação em que o aluno tenha se matriculado numa série k , num determinado ano letivo t , mas tenha abandonado os estudos ao longo deste ano. Se esse aluno efetivou sua matrícula na escola no(s) ano(s) subsequente(s), ele “repetirá” a série k (afastado por abandono repetente). Caso ele não retorne à escola no(s) ano(s) subsequente(s) é considerado afastado por abandono evadido. Ou seja, o abandono pode ser acompanhado pela evasão do sistema. Em contrapartida, se um aluno foi aprovado ou reprovado num ano, não se matriculando na escola nos demais anos do acompanhamento da coorte é caracterizada a evasão. Nesse caso, o aluno é considerado aprovado ou reprovado evadido. Assim, o que diferencia ambos os termos é a efetivação da matrícula no sistema escolar num determinado ano. Na tese, são utilizados os conceitos explicitados, propostos pelo KLEIN (1995) e adotados pelo INEP.

Algumas considerações devem ser feitas com relação ao registro de informações relacionadas à evasão e ao afastamento por abandono na “Ficha B”. No processo de coleta de dados não foi considerada a diferença conceitual entre ambos os termos, registrando-se apenas a informação “E” (equivalendo, no dicionário da pesquisa, à Evasão). Portanto, no tratamento estatístico dos dados, essa distinção foi feita com base nos conceitos de evasão e abandono mencionados.

Às situações seguintes, foi dado o tratamento especificado:

- i) como todos os alunos da coorte se matricularam em 1999, o registro “E” na “Ficha B”, naquele ano, foi designado como abandono. Caso o aluno não tenha retornado ao sistema em 2000 (e anos posteriores, no período sob estudo), o caso foi tratado como evasão no ano 2000 (evasão após afastamento por abandono em 1999);
- ii) caso num ano específico t (2000, 2001 ou 2002) tenha sido registrado “E” na ficha do aluno e ele não tenha voltado ao sistema escolar no ano seguinte (e demais anos do período sob análise), o caso foi tratado como evasão no ano t ;
- iii) se num determinado ano letivo t constou a informação “E” e no ano $t+1$ constou o resultado aprovado (A) ou reprovado (R) na ficha do aluno, a informação “E” foi considerada afastamento por abandono no ano t ;
- iv) na “Ficha B” constou também a informação ET, assim discriminada no dicionário da pesquisa: Evadiu (abandonou) e depois pediu transferência. Nesse

- caso, a informação foi considerada como afastamento por abandono seguido por transferência no ano t ;
- v) em 2003, se o registro que constou na ficha do aluno foi “E” o caso foi considerado como evasão. Talvez o aluno tenha até voltado ao sistema no ano seguinte, mas como esse ano designa o término do estudo da coorte, foi dado esse tratamento à informação; ou talvez ele tenha feito sua matrícula na escola nesse ano, abandonando os estudos ao longo do ano, mas na ausência dessa informação, foi considerado como evadido;
 - vi) houve muitos casos de alunos que foram considerados aprovados ou reprovados ao fim de um dado ano letivo t , sem constar nenhuma informação na sua ficha no(s) ano(s) seguinte(s). Nesses casos, considerou-se que o aluno evadiu no ano t para o qual já não constava nenhuma informação na sua ficha escolar.

É possível que algumas dessas questões tenham conduzido a alguns lançamentos incorretos nas planilhas de dados (casos de abandono considerados como evasão e subestimação do tempo até a ocorrência da evasão), mas acredita-se que a proporção desses casos tenha sido mínima.

5.2 Tempo de sobrevivência aos eventos: estratégias empíricas aplicadas

Na análise de sobrevivência os eventos básicos são a reprovação, a repetência, o afastamento por abandono e a evasão escolar. Em geral, há uma estreita relação entre os dois primeiros e os dois últimos eventos. Daí, a expectativa de obtenção de resultados bem similares diante do uso da técnica de análise. Entretanto, a questão a ser levantada nessa seção é de ordem metodológica. Essa questão é discutida com base na FIG. 4 (pg.58).

O fluxo escolar retratado na FIG. 4 mostra que os resultados ocorridos no ano letivo t (ano da matrícula inicial) independem do resultado ocorrido no ano $t+1$. Situação oposta é verificada para alguns dos eventos registrados no ano $t+1$, pois esses têm sua ocorrência associada ao evento ocorrido no ano t . Na figura, entre os eventos registrados no tempo $t+1$, associados a um evento ocorrido no tempo t , são destacados apenas aqueles que são objetos de estudo dessa tese: a repetência e a evasão. Observa-se que a repetência da série k no ano $t+1$ está relacionada à reprovação ou afastamento por abandono na série k , no ano t ; sendo possível também a repetência apesar da aprovação na série $k-1$ no ano t . Raciocínio semelhante pode ser aplicado à situação de evasão: os alunos que evadiram no

ano $t+1$ estiveram matriculados na escola no ano t , tendo sido considerados, naquele ano, aprovados, reprovados ou afastados por abandono.

Essa especificidade relacionada à associação (ou não) entre os eventos nos anos letivos t e $t+1$ requer um tratamento especial no ato da determinação do tempo (T) de sua ocorrência. Portanto:

- i) sendo os eventos básicos a reprovação e o afastamento por abandono (ocorridos no ano t independentemente do que ocorreu noutro ano), considerou-se a seguinte escala de tempo anual: $T = 1, 2, 3, 4$ e 5 , designando os anos de 1999 a 2003;
- ii) no caso da repetência e da evasão, dada a associação existente entre estes eventos ocorridos no ano letivo $t+1$ e um outro evento ocorrido no ano t , adotou-se a seguinte escala de tempo: $T=1$ para o biênio 1999-2000; $T=2$ para o biênio 2000-2001, $T=3$ para o biênio 2001-2002 e $T=4$ para o biênio 2002-2003.

Assim, na determinação de T , para os eventos em que é considerado apenas o resultado escolar ocorrido independentemente num ano letivo específico, foram considerados 05 pontos no tempo. Para os eventos que sinalizam uma transição entre um estado e outro, ou entre um evento e outro, entre dois anos consecutivos, foram considerados 04 pontos no tempo.

Após a estimação das curvas de sobrevivência, são estimados os modelos de regressão hierárquicos logísticos longitudinais. A seguir, são apresentadas as covariáveis incluídas nos modelos e algumas estratégias utilizadas relacionadas à tais variáveis.

5.3 Variáveis dos modelos de regressão: estratégias empíricas aplicadas

As variáveis-resposta foram extraídas da base Ficha B. Quanto às covariáveis, parte foi extraída da Ficha B, parte da base Fatores Associados (arquivo de alunos e de escolas) e parte do Censo Escolar de 1999. Portanto, inicialmente foi feito uma junção referente aos arquivos das três bases.

Na base Ficha B, os alunos da coorte estão distribuídos entre 155 escolas. Uma vez que as variáveis relacionadas ao estabelecimento escolar são incluídas no nível 3 dos modelos, foi preciso checar a consistência entre as escolas constantes na Ficha B e as

constantes no arquivo “escolas” da base Fatores Associados¹⁵. Nesse arquivo, o total de escolas varia de ano para ano. Por exemplo, em 1999 constam 148, enquanto em 2000 esse total equivale a 155. Dessa forma, verifica-se que para o ano inicial há um déficit na base, em relação à Ficha B, equivalente a sete escolas. Para não perder as informações dos alunos nelas matriculados, foram utilizadas as informações escolares do ano 2000 para o ano-base. A suposição é que não houve mudanças significativas nas informações entre ambos os anos.

Na especificação dos modelos, nos níveis 1 e 2 constam as variáveis de aluno e no nível 3 as variáveis da escola. Essas últimas variáveis são consideradas, nos modelos, como “fixas” no tempo, ou seja, são fixadas no ano-base de 1999. A justificativa é feita a seguir:

- i) Para as variáveis relacionadas à estrutura básica escolar (presença de quadra, existência de laboratório de informática, de sala de tv/vídeo e biblioteca) é razoável supor que não houve mudanças significativas quanto a sua existência ao longo do período, mantendo-se as informações para o ano-base (dados do Censo Escolar);
- ii) Para as variáveis relacionadas ao sistema de segurança na escola (presença de guarda/vigia, sistema de proteção contra incêndio e controle de entrada de estranhos), não foram verificadas mudanças significativas entre 1999 e 2003 (dados da Ficha B: teste de médias);
- iii) No ano-base constam informações, no Censo Escolar, sobre o total de matrículas efetivadas na 4ª série nas escolas da amostra. Constam também informações sobre o total de matrículas de alunos promovidos da 3ª para a 4ª série, naquele ano. Como uma coorte é acompanhada por cinco anos e a entrada de novos alunos a cada ano letivo não é considerada no estudo, o ideal é considerar as informações existentes somente para o ano-base.
- iv) De forma similar ao tratamento dado às demais variáveis de escola, considerou-se o total de salas e o total de professores de nível superior existentes na escola registrados no ano-base de 1999 (Censo Escolar). A suposição é que esse total não variou de forma significativa ao longo do período de acompanhamento.

¹⁵ As informações sobre as variáveis de escola estão contidas nesse arquivo.

Com relação aos itens (i) e (ii) é importante mencionar que foram construídos indicadores a partir do conjunto de variáveis citadas¹⁶. Outra justificativa para considerar tais indicadores como fixos refere-se à maior facilidade na interpretação dos resultados.

Entre as variáveis relacionadas aos alunos, também foram construídos indicadores, associados ao nível socioeconômico familiar. Como no caso das variáveis associadas à escola, para a construção destes indicadores utilizou-se o método *Homals*, indicado para o caso de variáveis dicotômicas. Trata-se de um método estatístico de análise de homogeneidade cujo objetivo é agrupar as diversas informações sobre cada variável em poucas dimensões. Mais precisamente, agrupar as diferentes respostas para os diferentes itens em poucas dimensões. No estudo, optou-se por manter apenas duas dimensões para cada indicador, em função do seu (alto) poder explicativo e da maior facilidade na interpretação dos resultados.

Na seqüência, são apresentadas as variáveis incluídas nas regressões e uma breve descrição dessas variáveis.

5.3.1 Variáveis incluídas na modelagem econométrica

Nessa seção são apresentadas as variáveis-resposta e respectivas covariáveis que foram incluídas nos modelos de regressão estimados.

Variáveis-resposta

As duas variáveis-resposta inseridas nas regressões foram mensuradas ao nível do aluno. Uma, medindo a probabilidade do aluno repetir uma série pela primeira vez entre a 4ª e a 8ª série do ensino fundamental no período de 1999 e 2003, para as escolas das seis UF's. A outra, medindo a probabilidade do aluno evadir no mesmo período. Na modelagem dessas probabilidades utilizou-se o valor “1” caso o evento (repetência, evasão) tenha ocorrido e “0” caso contrário.

Covariáveis

As covariáveis de nível 1, 2 e 3 que constam nos modelos estimados foram selecionadas a partir das evidências empíricas levantadas no capítulo 2 e da revisão teórica feita no capítulo 3. No nível 1 foram incluídas informações contemporâneas da trajetória

¹⁶ Maiores detalhes sobre esses indicadores, especificamente sobre sua composição e valores, ver anexos.

escolar do aluno, que variam ao longo do tempo, além do tempo, visando captar o comportamento longitudinal dos eventos básicos. No nível 2 foram inseridas as variáveis do aluno que são fixas no tempo (variáveis relacionadas à trajetória escolar passada, características demográficas, como medida de controle nas regressões e informações do *background* familiar). Uma vez que os alunos estão aninhados dentro de escolas, foi incluído o nível 3 nos modelos, representado por variáveis relacionadas às escolas. A descrição dessas covariáveis é feita a seguir.

Covariáveis do nível 1

No nível 1, além das variáveis tempo e série cursada, foram incluídas outras variáveis que variam no tempo, listadas a seguir.

- **Série/Tempo:** dada a natureza longitudinal dos dados é necessário considerar, no modelo, o tempo associado à ocorrência do evento. Como os alunos foram acompanhados desde sua matrícula na 4ª série em 1999 até 2003, ano que deveriam completar o ensino fundamental, conseqüentemente a série varia da 4ª a 8ª. Mas, dado o período de estudo, enquanto para a evasão a série mais elevada considerada na análise é a 8ª, para a repetência esta série é a 7ª. No caso da variável Tempo, pelo fato dos eventos analisados corresponderem a medidas de fluxo ou transição, seus valores variam de 1 a 4. Tanto no caso da repetência quanto da evasão foi dado o seguinte tratamento à variável: biênio 1999/2000, $t=1$; biênio 2000/2001, $t=2$; biênio 2001/2002, $t=3$ e biênio 2002/2003, $t=4$. Na construção da base de dados, observou-se se o evento de interesse ocorreu no segundo ano de cada biênio, sendo que as variáveis explicativas associadas à ocorrência (ou não) do evento nesse ano têm seus valores determinados para o primeiro ano de cada biênio. Assim, tais valores referem-se aos anos de 1999 a 2002.

- **Informações contemporâneas:** segundo a literatura existente e segundo a especificação da FPE, o desempenho educacional corrente é afetado tanto pelas informações passadas quanto pelas informações contemporâneas do aluno. Nesse estudo, essa dimensão dos dados educacionais foi captada pelas variáveis trabalho e proficiência do aluno.

a) Trabalho: na pesquisa “Avaliação de Desempenho: fatores associados” questionou-se ao aluno se ele trabalhou ou não num determinado ano letivo. Se a resposta foi positiva,

definiu-se uma variável indicadora igual a “1” e, se negativa, igual a “0”, sendo a categoria omitida “não trabalhou”. A literatura corrente tem apontado um efeito negativo do trabalho sobre o resultado escolar (CÉSAR & SOARES, 2001; MACEDO, 2004; PEREIRA, 2006).

b) Proficiência no ano letivo anterior, centralizada na média da escola: foi calculada a proficiência média do aluno (centralizada na média da escola), nos exames de português e matemática, para os anos de 1999 a 2002¹⁷. Variável contínua, referente ao ano t+1, associada à ocorrência do evento no ano t. A idéia é verificar como o resultado escolar corrente é afetado pelo resultado escolar passado do aluno. Esse estudo fez uso dos cálculos da proficiência utilizados em PEREIRA (2006). No seu estudo, a autora estimou a proficiência do aluno através do modelo da Teoria da Resposta ao Item. Ela se baseia em HAMBLETON (1993), que infere que a habilidade ou proficiência de um indivíduo deve ser assumida como uma habilidade cognitiva, sendo, portanto, uma característica latente que deve ser estimada através de modelos estatísticos de variáveis latentes.

c) Aprovação: variável indicadora com valor igual a 1 se o aluno foi aprovado na série cursada antes da ocorrência do evento e valor igual a 0, caso contrário (se foi reprovado ou afastado por abandono). Incluída como medida de controle nos modelos cuja variável-resposta é a evasão.

Diante da inclusão destas variáveis nos modelos, a expectativa é que alunos que tenham trabalhado em algum ano letivo e que tenham uma menor proficiência (nos testes de português e matemática) apresentem maiores probabilidades de repetência e evasão entre a 4ª e a 8ª série do ensino fundamental. Espera-se, também, que se o aluno foi aprovado numa determinada série menor a probabilidade dele evadir da escola no ano seguinte.

Covariáveis do nível 2

Nesse nível do modelo foram incluídas as variáveis consideradas invariáveis no tempo, como as características demográficas do aluno, aspectos da sua trajetória escolar

¹⁷ Para cada biênio observou-se se o evento ocorreu no ano t+1 e a proficiência obtida no ano t. Assim, o último ano com informações para a proficiência é 2002, pois foi observado se o aluno repetiu ou evadiu em 2003 e respectiva proficiência em 2002. Procedimento similar foi adotado para as demais variáveis.

passada (anterior à 4ª série) e variáveis de *background* familiar. A seleção das variáveis baseou-se em diversos estudos (conforme capítulos 2 e 3) que mostraram diferenciais no resultado escolar em função da influência desses fatores.

- **Características demográficas:** as variáveis demográficas sexo e cor foram incluídas para fins de controle nos modelos de regressão.

a) Sexo: variável indicadora, assumindo os valores “1” e “0”, para mulheres e homens respectivamente, e cuja categoria omitida é o sexo masculino.

b) Cor: foram criadas duas categorias, branca/amarela e outras (mulato, negro e indígena). A categoria omitida que assume o valor “0” é outras.

Em conformidade com os estudos existentes, esperam-se menores probabilidades de ocorrência dos eventos básicos para as meninas e para os alunos da cor branca/amarela.

- **Trajatória passada:** na Função de Produção Educacional é mostrado como as informações relacionadas ao aluno, num tempo $t-1$, afetam o seu desempenho no tempo t . A idéia é verificar até que ponto os fatos da trajetória escolar passada do aluno influenciam a sua trajetória escolar presente. Foi selecionada, para a análise, a variável repetência em séries anteriores à 4ª. Estudos anteriores que também utilizaram a base “Avaliação de desempenho: fatores associados” mostraram um efeito negativo dessa variável sobre o resultado escolar (MACEDO, 2004; LUZ, 2005; PEREIRA, 2006).

Repetência: o aluno foi interrogado se ele já havia repetido de ano (antes da 4ª série) e, se sim, quantas vezes (uma, duas, três, mais de três). Construiu-se uma variável indicadora que assumiu o valor “1” se o aluno respondeu sim e “0” se respondeu não.

- **Variável de *background* familiar:** entre as variáveis associadas ao *background* familiar, uma bastante utilizada na literatura que influencia o desempenho educacional do aluno é o nível socioeconômico da família (NSE). É esperada uma associação negativa e significativa entre esta variável e as variáveis-resposta utilizadas nessa tese. Ou seja, quanto maior o NSE da família, menor a probabilidade de repetência e evasão do aluno. Nas bases de dados constavam informações, também, para a escolaridade dos pais. Mas, devido à alta correlação existente entre ambas as variáveis (comprovada pela literatura existente), optou-se por incluir nos modelos apenas os indicadores do NSE.

Nível socioeconômico - NSE: Foram construídos dois indicadores a partir da posse dos seguintes bens duráveis: rádio, televisão a cores, vídeo-cassete, geladeira, freezer, máquina de lavar, aspirador de pó e automóvel. Além disso, considerou-se a existência (ou não) de empregada doméstica no domicílio. Esses itens foram selecionados a partir do critério de classificação econômica proposto pela ABEP (Associação Brasileira de Empresas de Pesquisa). Ressalta-se que a ABEP também inclui a existência (ou não) de banheiro no domicílio. Devido à pouca variabilidade nas respostas, esse item não foi considerado na construção dos indicadores. A interpretação dos indicadores é feita a seguir (a composição deles pode ser vista nos anexos).

- 1) Ind. NSE_1: Esse indicador se destaca por mostrar uma relação positiva entre a posse de bens duráveis/empregada e a primeira dimensão.
- 2) Ind. NSE_2: Esse indicador é caracterizado por apresentar uma relação forte e negativa entre não posse dos bens básicos (rádio, tv e geladeira) e a segunda dimensão.

Os dois indicadores foram considerados como fixos (no ano-base de 1999), pois o teste para médias apontou que não houve variação significativa no nível socioeconômico familiar entre 1999 e 2003.

Covariáveis do nível 3

As variáveis selecionadas nesse nível medem basicamente os recursos físicos e humanos das escolas (infra-estrutura, sistema de segurança, escolaridade do professor). São consideradas também variáveis associadas ao tamanho (total de matrículas na 4ª série) e ao corpo discente da escola (total de matrículas de alunos promovidos da 3ª para a 4ª série, proficiência média da escola). A região em que a escola está inserida também foi incluída nesse nível.

FERRÃO et al (2002) constataram uma associação positiva entre infra-estrutura e segurança na escola e desempenho acadêmico. Variáveis que medem a qualidade física da escola e recursos humanos e financeiros também foram significativamente relacionadas ao desempenho noutros estudos (HANUSHEK, GOMES-NETO e HARBISON, 1996; ALBERNAZ et al, 2002). No estudo realizado por MACEDO (2004), o sistema de segurança não teve efeito significativo sobre o desempenho escolar. Entretanto, melhores instalações físicas das escolas mostraram-se associadas à melhores rendimentos.

a) Estrutura básica: foram construídos dois indicadores de estrutura básica da escola, a partir das variáveis do Censo Escolar que indicam a existência ou não dos seguintes itens: quadra, laboratório de informática, sala de tv/vídeo e biblioteca.

1) Ind. Estrutura_1: Indicador caracterizado por mostrar uma relação negativa entre a existência dos itens na escola e a primeira dimensão.

2) Ind. Estrutura_2: A existência de laboratório de informática na escola é o componente de maior poder explicativo nesse indicador.

b) Sistema de segurança: o questionário da Ficha B contempla questões relacionadas à segurança na escola (se havia ou não policial/vigia em turno integral, controle de entrada de estranhos e sistema de proteção contra incêndio). Dois indicadores foram construídos a partir das respostas (afirmativas ou negativas) fornecidas.

1) Ind.Segurança_1: Esse indicador se caracteriza por apresentar uma relação negativa mais forte entre a existência dos itens listados e a primeira dimensão.

2) Ind.Segurança_2: O componente de maior poder explicativo nesse indicador é a existência de sistema de proteção contra incêndio na escola.

c) Escolaridade dos professores: no Censo Escolar de 1999 constam informações sobre o total de professores da escola segundo o nível de escolaridade (fundamental, médio, superior). Optou-se por incluir o total de professores com nível superior, pela suposição de um maior diferencial sobre o resultado escolar.

d) Total de matrículas na 4ª série: No Censo Escolar há informações sobre o total de matrículas efetivadas em cada série, por nível de ensino. A inclusão da variável total de matrículas no ano-base é importante por estar associada ao tamanho do estabelecimento escolar.

e) Total de matrículas de alunos promovidos da 3ª para a 4ª série em 1999: essa informação, constante no Censo Escolar, permite verificar o efeito de se ter um maior número de alunos promovidos num determinado ano letivo sobre o resultado escolar no ano letivo subsequente.

f) Proficiência média da escola: Com base na proficiência calculada para os alunos das escolas da amostra (PEREIRA, 2006), calculou-se a proficiência média da escola para os anos de 1999 a 2002. Variável contínua, referente ao ano t , associada à ocorrência do evento no ano $t+1$. Espera-se que quanto maior a proficiência média da escola, ou seja, quanto maior o desempenho médio dos colegas da escola, menor a probabilidade estimada de ocorrência dos eventos para um determinado aluno.

g) Total de salas: Variável contínua, extraída do Censo Escolar de 1999. Essa variável também pode refletir o tamanho do estabelecimento escolar.

h) Região: foram construídas variáveis indicadoras para as regiões Norte (escolas do Pará e de Rondônia), Nordeste (escolas de Pernambuco e de Sergipe) e Centro-Oeste (escolas de Goiás e do Mato Grosso do Sul), sendo a categoria omitida a região Nordeste. A idéia é identificar possíveis diferenciais regionais sobre a probabilidade de ocorrência da repetência e da evasão escolar.

Nos dois próximos capítulos são apresentados e discutidos os resultados obtidos a partir das metodologias utilizadas.

6. ANÁLISE DESCRITIVA E ANÁLISE DE SOBREVIVÊNCIA: ACOMPANHANDO A TRAJETÓRIA DA COORTE ESCOLAR

Este capítulo tem dois objetivos: i) apresentar alguns dados que descrevem alguns aspectos da coorte de alunos sob estudo e; ii) apresentar as curvas de sobrevivência aos eventos reprovação, repetência, afastamento por abandono e evasão. Inicialmente, são mostradas algumas informações mais gerais relacionadas à trajetória escolar da coorte, entre 1999 e 2003. Posteriormente, são mostradas informações mais específicas relacionadas aos eventos citados acima, segundo as UF's. Finalmente, as curvas de sobrevivência a esses eventos são apresentadas e analisadas.

6.1 Trajetória da coorte: aspectos descritivos

Os alunos das escolas da amostra foram acompanhados desde sua matrícula na 4ª série em 1999 até 2003, ano que deveriam concluir o ensino fundamental. Para uma melhor compreensão dos resultados obtidos nesse estudo de natureza longitudinal, é interessante mostrar alguns dados relacionados ao estoque e fluxo desses alunos, para a totalidade do período analisado.

Inicialmente, é apresentada uma tabela com o total original de escolas e alunos registrados na Ficha B.

TABELA 1
Total de alunos segundo o número de escolas e de turmas e UF's, Ficha B

UF	Escolas	Turmas	Alunos
Rondônia	14	47	1.111
Pará	31	88	2.444
Pernambuco	34	95	2.446
Sergipe	16	43	984
Mato Grosso do Sul	35	80	2.023
Goiás	32	75	1.912
Total	162	428	10.920

Fonte: Relatório CEDEPLAR/INEP (2005?).

Segundo a tabela, a coorte era composta por 10.920 alunos distribuídos entre 162 escolas. Entretanto, no processo de limpeza dos dados, foram constatados 291 casos de alunos cuja ficha não constavam informações referentes a 4ª série, e sim para a 5ª série e seguintes. Nesse total, estavam incluídos também casos de alunos que cursaram a 4ª série em 1998. Esses casos foram excluídos da análise. Outra situação constatada refere-se à duplicidade no registro de alunos: se ele estudou em duas escolas, havia na base duas

linhas de informações para esse aluno. Nesse caso, se numa escola constavam informações do aluno para as séries iniciais e, na outra escola, para as séries seguintes, manteve-se o registro apenas para a primeira escola, acrescentando-se a informação de transferência na sua ficha escolar.

Subtraindo-se esses casos do total inicialmente registrado, a coorte de alunos tem a magnitude especificada na tabela seguinte. Nesta tabela, consta o total de alunos matriculados por ano e série, além dos casos de evasão e transferência. A idéia é mostrar o fluxo de alunos entre 1999 e 2003, explicitando os fatores que resultaram na redução da coorte no período.

TABELA 2
Fluxo de alunos segundo matrículas, evasão e transferências: Ficha B, 1999 a 2003

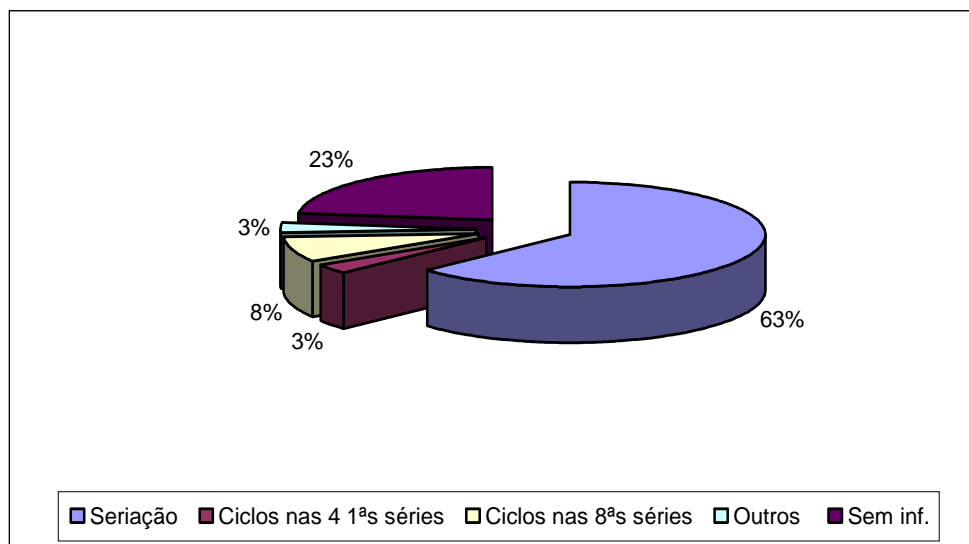
SÉRIE/ANO	1999	2000	2001	2002	2003
4ª série	10.562	736	72	8	0
5ª série	0	8.196	1.274	268	45
6ª série	0	0	5.874	1.074	245
7ª série	0	0	0	4.452	805
8ª série	0	0	0	0	3.393
Matrículas	10.562	8.932	7.220	5.802	4.488
Evadidos (não matriculados)	0	400	416	276	313
Transferidos	1.230	1.296	1.142	999	1.114
Evadidos + transferidos	1.230	1.696	1.558	1.275	1.427

Fonte: Elaboração própria com base nos dados da Ficha B – CEDEPLAR/INEP 1999/2003.

Em 1999 a coorte inicial era composta por 10.562 alunos. Ao término do ano letivo foram registradas 1.230 transferências (para outras escolas e/ou para a EJA). No ano de 2000 houve 400 casos de evasão. Subtraindo-se da quantidade inicial de alunos as transferências e os casos de evasão citados, registrou-se 8.932 matrículas (na 4ª e 5ª série). Em 2001 foram observados 416 casos de evasão que, somados às transferências ocorridas no ano anterior (1.296), resultou num total de 7.220 matriculados (na 4ª, 5ª e 6ª série). No ano de 2002 foram efetivadas 5.802 matrículas distribuídas entre a 4ª e a 7ª série (nesse ano houve 276 evasões e, no ano anterior, 1.142 transferências). Finalmente, em 2003 foram observadas 313 evasões que, somadas ao número de transferências registradas no ano anterior (999), totalizou 4.488 matrículas (na 5ª, 6ª, 7ª e 8ª série). Esses dados evidenciam a redução expressiva da coorte no período, devido ao grande número de casos de transferência e evasão. Verifica-se que a maior parte das transferências aconteceu nas séries iniciais (4ª e 5ª), sendo que a ocorrência da evasão foi mais significativa em 2001.

Antes de prosseguir, será mostrado o sistema de seqüenciação adotado nas escolas da amostra (GRÁF. 8). Isso é particularmente importante, pois foi grande o número de transferências de alunos e parte dessas transferências está relacionada ao sistema de ciclos.

GRÁFICO 8
Sistema de seqüenciação, Ensino Fundamental, por escolas da Ficha B



Fonte: Elaboração própria a partir da base Avaliação de Desempenho: Fatores Associados.

Observa-se que 63% das escolas adotavam o sistema de seriação, enquanto em aproximadamente 11% prevalecia o sistema de ciclos (8% na 8ª série e 3% nas quatro primeiras séries). Foi significativo o percentual de casos sem informação sobre o regime adotado (23%). No tratamento das informações, optou-se por interromper o acompanhamento da trajetória do aluno para as escolas com o sistema de ciclos nas séries finais do ensino fundamental. Neste caso, foram aproveitadas as informações anteriores à passagem para o regime de ciclos. Assim, no ano anterior à transferência para a EJA ou supletivo, constou na ficha escolar o resultado final obtido e a informação transferência para a EJA ou transferência para o supletivo.

A TAB. 2 revelou também que, ao longo do período, uma substancial parte da coorte ficou retida em séries anteriores à série que deveria estar sendo cursada num determinado ano letivo. Esse aspecto pode ser melhor visualizado nos gráficos seguintes.

GRÁFICO 9
Alunos matriculados (%)
por série, em 2000

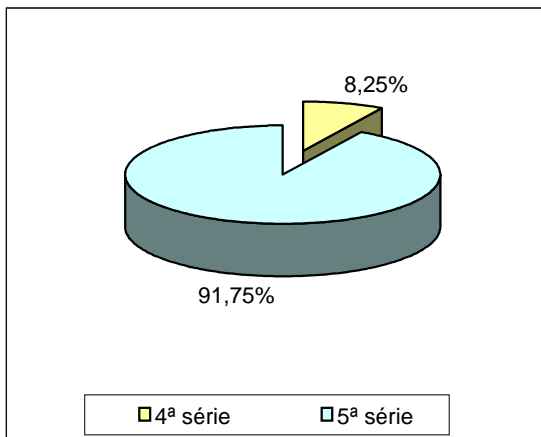
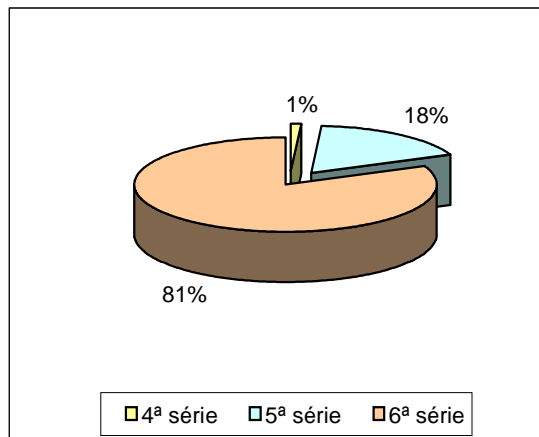


GRÁFICO 10
Alunos matriculados (%)
por série, em 2001



Nota: Total de alunos matriculados/ano: 1999=10.562; 2000=8.934; 2001=7.221.
Fonte: Elaboração própria com base na Ficha B - CEDEPLAR/INEP 1999/2003.

GRÁFICO 11
Alunos matriculados (%)
por série, em 2002

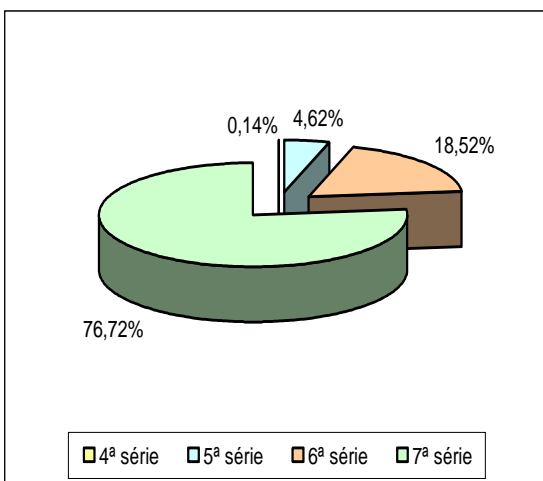
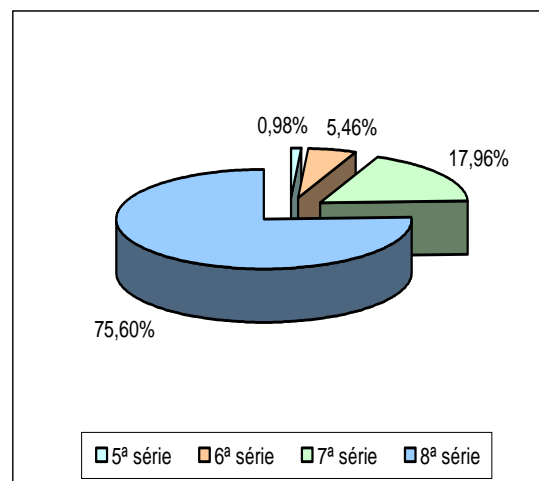


GRÁFICO 12
Alunos matriculados (%)
por série, em 2003



Nota: Total de alunos matriculados/ano: 2002=5.803; 2003=4.488.
Fonte: Elaboração própria com base na Ficha B - CEDEPLAR/INEP 1999/2003.

O GRÁF. 9 mostra que, dentre os alunos matriculados nas escolas da amostra no ano de 2000, 8,25% efetuaram sua matrícula na 4ª série (devido à reprovação ou afastamento por abandono, em 1999). No GRÁF. 10 observa-se que apenas 81% dos matriculados em 2001 cursavam a série adequada, ou seja, a 6ª série. Um expressivo percentual de alunos, da ordem de 18%, ficou retido na 5ª série. Os gráficos 11 e 12 revelam a grande porcentagem de alunos matriculados que cursavam séries anteriores à adequada, nos anos

de 2002 e 2003. Nesses anos, os alunos deveriam estar cursando a 7ª e a 8ª série respectivamente. Contudo, aproximadamente ¼ deles estavam retidos noutras séries.

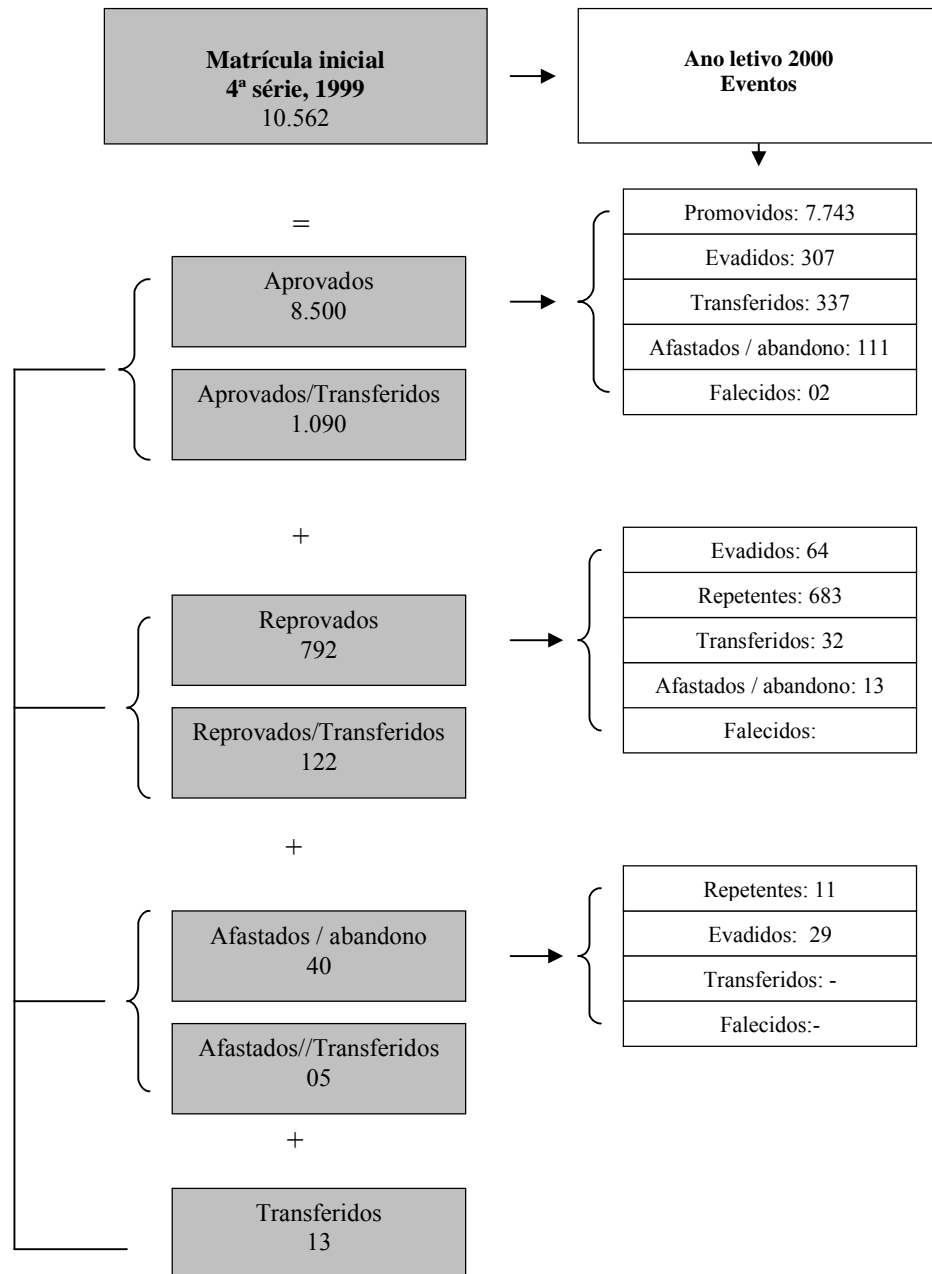
Na seqüência são apresentadas as figuras 5 a 9. Nelas constam os fluxos bienais para os alunos da coorte, sendo que no primeiro ano de cada biênio consta o total de alunos matriculados e o resultado obtido ao fim do ano letivo e, no segundo ano, o evento de fluxo correspondente. No lado esquerdo das figuras o somatório de alunos por evento totaliza o número de matriculados por ano.

Foi visto anteriormente (TAB. 2) que a grande redução da coorte entre 1999 e 2003 deveu-se basicamente às transferências e à evasão. Nas figuras, os casos de transferência podem ser acompanhados no primeiro ano de cada biênio (transferido, aprovado e transferido, reprovado e transferido, afastado por abandono e transferido), pois os alunos efetivaram sua matrícula e solicitaram a transferência ao longo ou ao término do ano letivo t. Os casos de evasão, ao contrário, não são visualizados no ano t inicial, pois a matrícula não é efetivada nesse ano específico (e nem nos anos subseqüentes, pois o aluno não retorna ao sistema). Assim, tais casos são registrados apenas no segundo ano de cada biênio, caracterizando o fluxo.

Ainda sobre as transferências: nota-se que a maioria dos casos ocorreu após aprovação numa determinada série, sendo esse volume mais significativo na 4ª e na 8ª série. Ou seja, as séries que marcam o início (1999) e o término (2003) do acompanhamento da coorte. Parte dessas transferências foi para a EJA/Supletivo seriado e parte para outras escolas (da amostra ou não). Verifica-se ainda que no período um total significativo de alunos solicitou transferência ao longo do ano letivo (transferidos). Nesse caso, não há informação na ficha escolar se o aluno começou ou não a cursar a série.

Figura 5

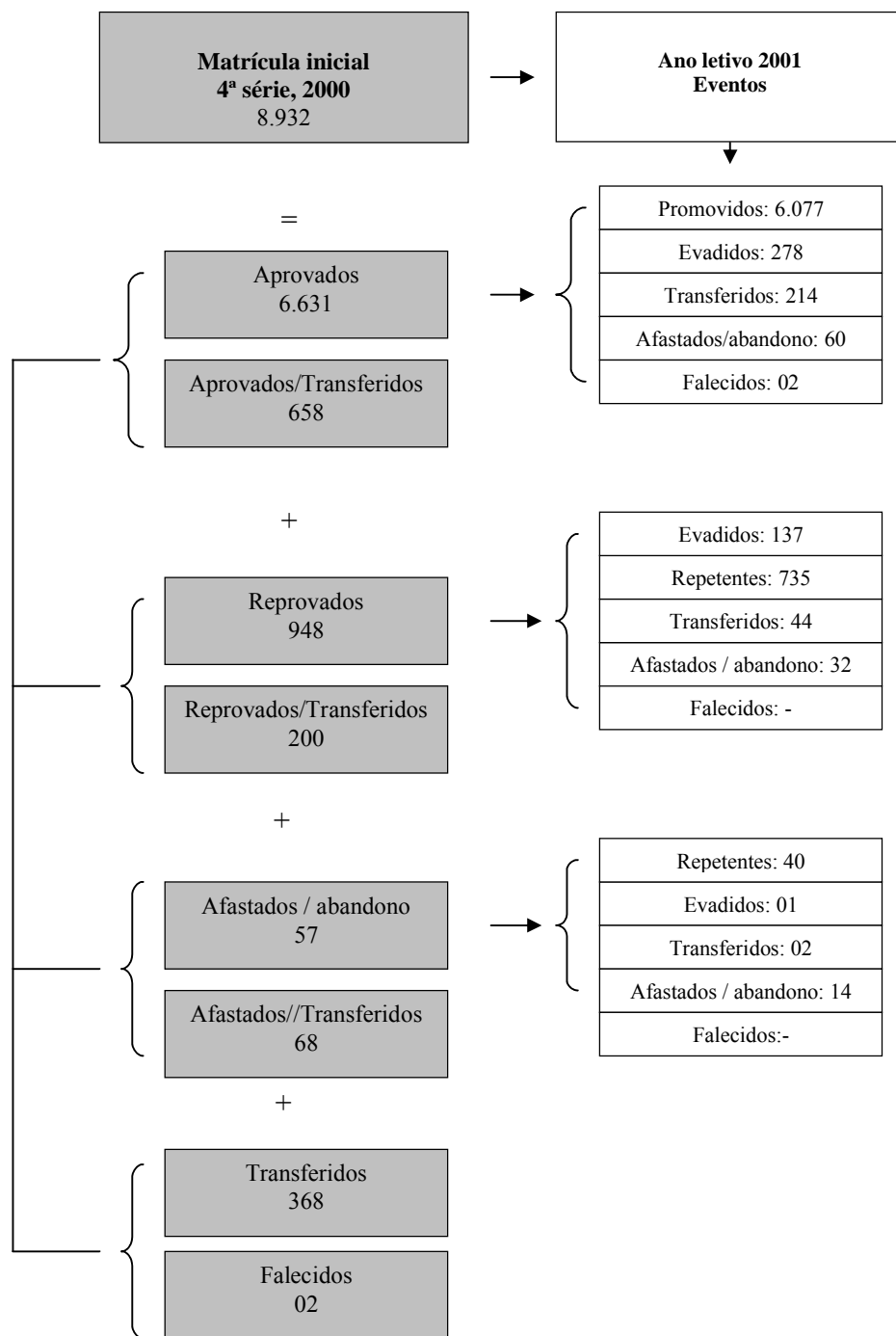
Fluxo escolar, segundo registros da “Ficha B” 1999-2000



Fonte: Elaboração própria com base na Ficha B, anos 1999 a 2003.

Figura 6

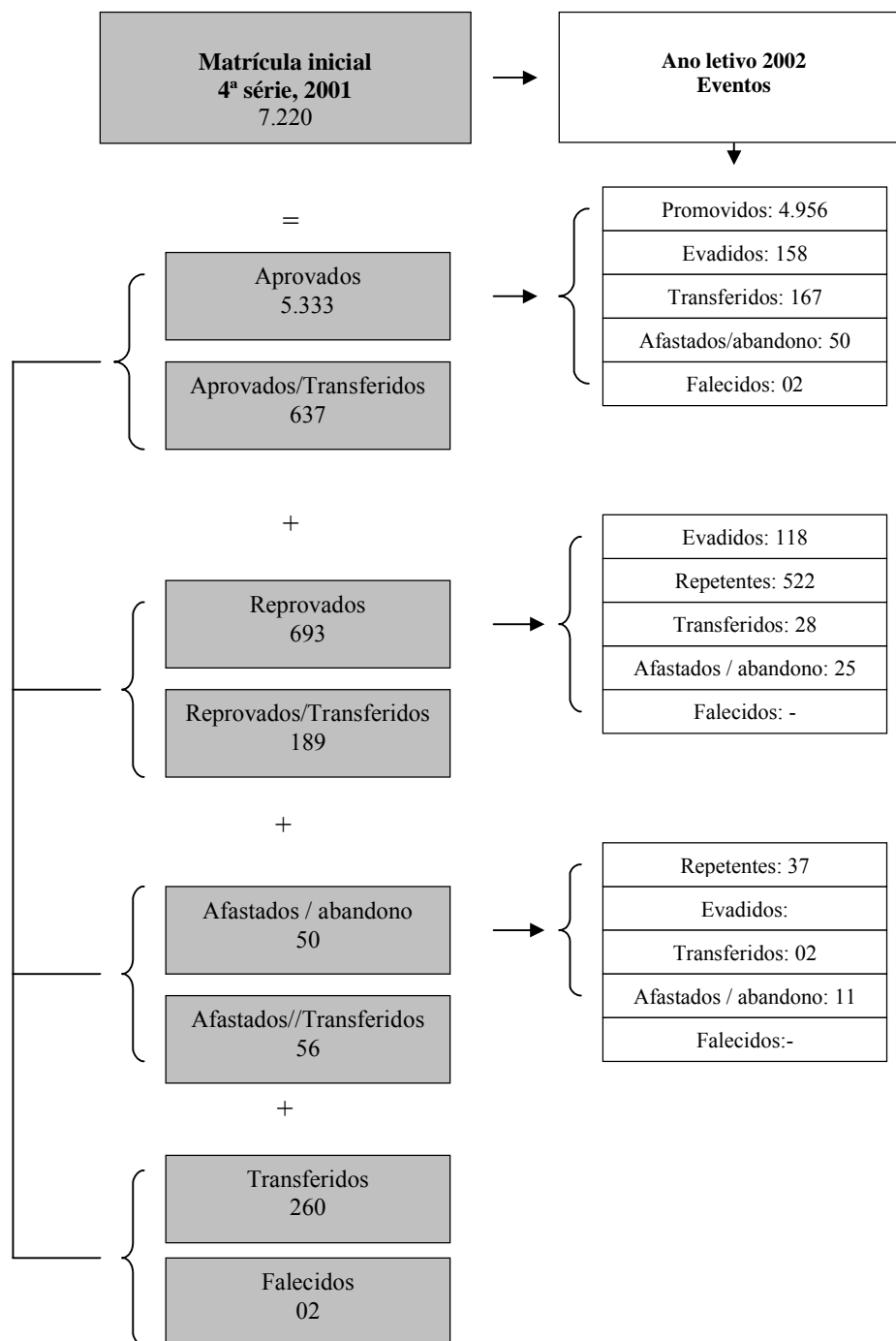
Fluxo escolar, segundo registro da “Ficha B” 2000-2001



Fonte: Elaboração própria com base na Ficha B, anos 1999 a 2003.

Figura 7

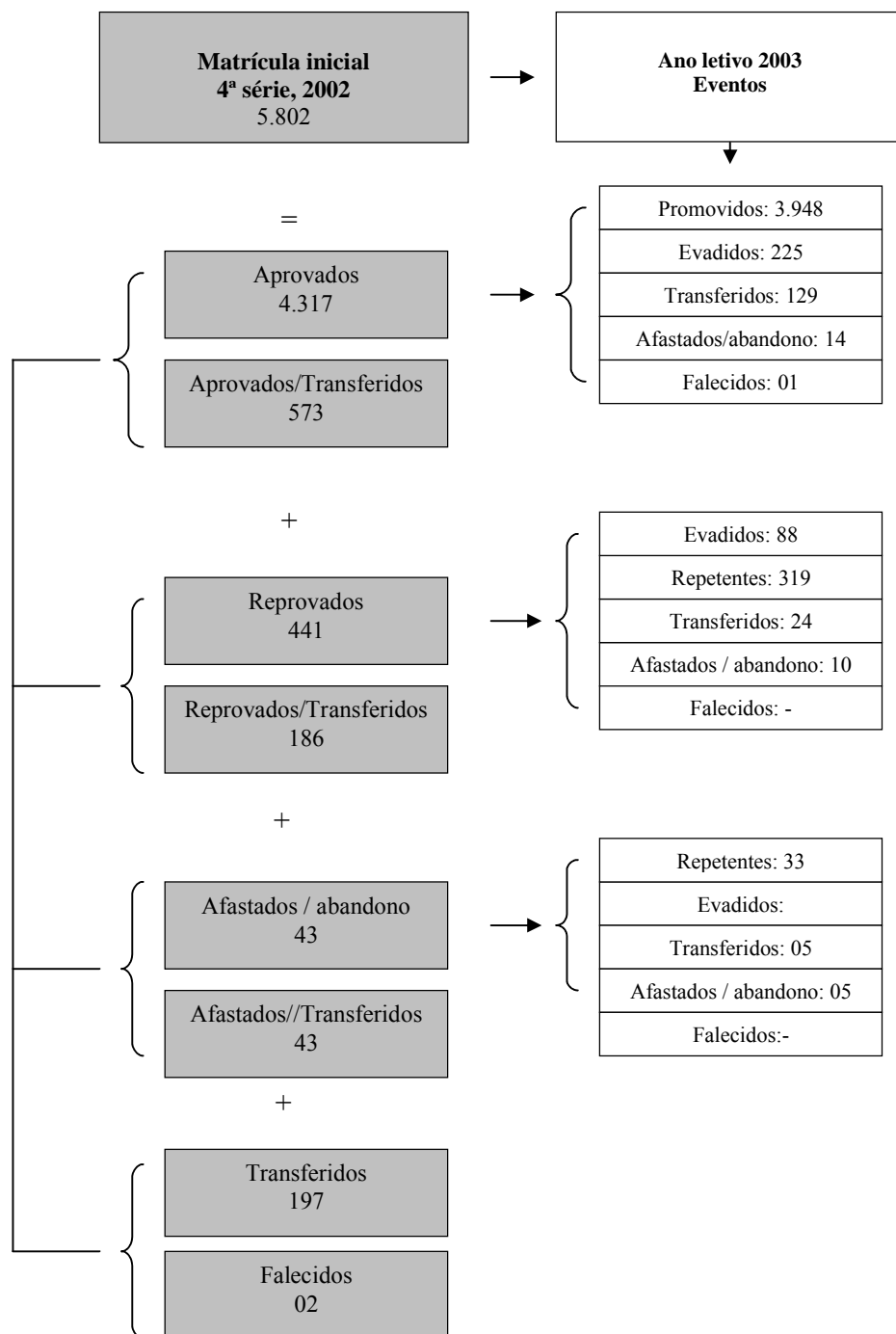
Fluxo escolar, segundo registros da “Ficha B” 2001-2002



Fonte: Elaboração própria com base na Ficha B, anos 1999

Figura 8

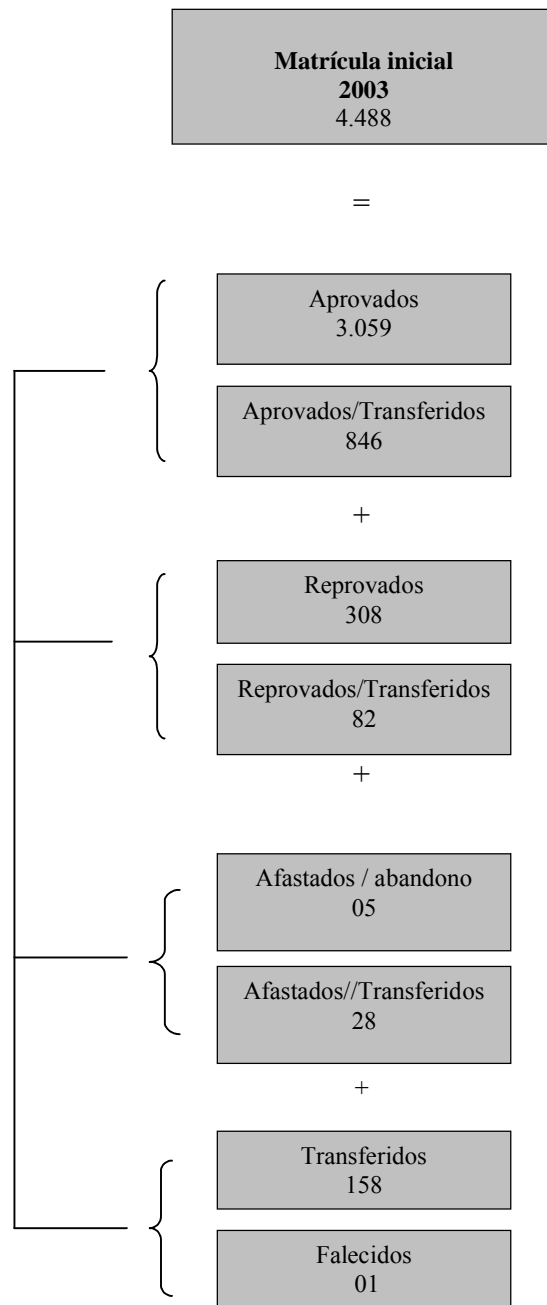
Fluxo escolar, segundo registros da “Ficha B” 2002-2003



Fonte: Elaboração própria com base na Ficha B, anos 1999

Figura 9

Matrículas / eventos segundo registros da “Ficha B” 2003



Fonte: Elaboração própria com base na Ficha B, anos 1999 a 2003.

A visualização das figuras revela que, da coorte inicial de 10.562 alunos, menos de 43%¹⁸ permaneceram nas escolas da amostra até o fim do acompanhamento do estudo. Deste total, 3.905 concluíram com aprovação o ensino fundamental (ou seja, menos de 37% da coorte). Fatores como transferência, evasão e reprovação, entre outros, explicam esse resultado.

Os dados sobre reprovação (reprovados, reprovados e transferidos) mostram que o maior volume de reprovados foi verificado em 2000. Esse evento ocorreu principalmente entre os alunos que cursavam a 5ª série. Entre 1999 e 2003 o percentual médio de repetentes no ano t+1, entre os reprovados no ano t, correspondeu a 77,85%.

Com relação aos casos de afastamento por abandono, importante lembrar que estes referem-se a alunos com matrícula efetivada num determinado ano t, que abandonaram os estudos ao longo deste ano, podendo ou não ter retornado ao sistema escolar no ano t+1 ou anos seguintes. A sequência das figuras mostra que a maioria dos afastamentos por abandono ocorreu em 2000, entre os alunos matriculados na 5ª série. Conforme explicado no capítulo anterior, existe a possibilidade de subestimação dos valores apresentados, devido à não distinção dos conceitos de abandono e evasão, no processo de coleta e registro de dados da pesquisa.

Na última parte das figuras, foram registrados os casos de falecimento. Pode ser verificado que 7 alunos (0,07% da coorte) faleceram durante o período de acompanhamento.

A seguir, é realizada uma análise descritiva dos eventos observados na trajetória escolar da coorte que constituem o objeto de estudo dessa tese. Essa análise é feita por UF's.

6.2 Eventos reprovação, repetência, afastamento por abandono e evasão, entre a 4ª e a 8ª série, segundo UF's: aspectos descritivos

Na análise de sobrevivência, como será visto, analisou-se a sobrevivência do aluno à primeira reprovação, à primeira repetência, ao primeiro afastamento por abandono e à evasão. A caracterização desses eventos é feita na sequência, visando um melhor entendimento dos resultados obtidos no emprego da metodologia mencionada.

¹⁸ Como afirmado anteriormente, se o aluno estudou numa escola da amostra e pediu transferência entre 1999 e 2003, o seu acompanhamento foi interrompido. Possivelmente, algumas das escolas de destino também pertencem à amostra. Nesse caso, esse percentual pode ser um pouco maior do que o que foi registrado.

A TAB. 3 mostra as taxas de reprovação por ano e série, para cada UF, no período analisado.

TABELA 3
Taxas de reprovação por ano, série e UF's: Ficha B, 1999 a 2003

	1999	2000	2001	2002	2003
	TAXA / SÉRIE	TAXA / SÉRIE	TAXA / SÉRIE	TAXA / SÉRIE	TAXA / SÉRIE
PE	4 ^a : 11,77 N=2.412	4 ^a : 21,14 5 ^a : 16,77 N=2.065	4 ^a : 22,22 5 ^a : 26,97 6 ^a : 13,27 N=1.650	5 ^a : 34,86 6 ^a : 18,10 7 ^a : 8,84 N=1.369	5 ^a : 40,91 6 ^a : 27,50 7 ^a : 19,77 8 ^a : 7,41 N=1.080
SE	4 ^a : 14,39 N= 931	4 ^a : 10,26 5 ^a : 19,91 N=750	4 ^a : 9,09 5 ^a : 33,52 6 ^a : 15,19 N=536	5 ^a : 27,66 6 ^a : 20,95 7 ^a : 4,33 N=383	5 ^a : 40,00 6 ^a : 21,95 7 ^a : 3,92 8 ^a : 6,63 N=268
MS	4 ^a : 3,97 N=1.967	4 ^a : 8,06 5 ^a : 13,30 N=1.671	4 ^a : 33,33 5 ^a : 26,56 6 ^a : 10,12 N=1.351	5 ^a : 26,67 6 ^a : 23,56 7 ^a : 11,27 N=1.082	6 ^a : 32,35 7 ^a : 21,34 8 ^a : 4,20 N=822
GO	4 ^a : 6,10 N=1.853	4 ^a : 6,58 5 ^a : 9,22 N=1.486	5 ^a : 8,84 6 ^a : 4,38 N=1.201	6 ^a : 12,50 7 ^a : 6,39 N=928	6 ^a : 16,67 7 ^a : 12,00 8 ^a : 4,24 N=678
PA	4 ^a : 7,06 N=2.339	4 ^a : 14,52 5 ^a : 7,70 N=2.088	4 ^a : 20,00 5 ^a : 15,54 6 ^a : 5,94 N=1.770	5 ^a : 16,00 6 ^a : 19,21 7 ^a : 4,33 N=1.452	6 ^a : 12,90 7 ^a : 15,32 8 ^a : 4,52 N=1.199
RO	4 ^a : 13,11 N=1.060	4 ^a : 13,51 5 ^a : 15,11 N=872	4 ^a : 11,11 5 ^a : 33,92 6 ^a : 15,23 N=712	5 ^a : 31,91 6 ^a : 30,72 7 ^a : 7,75 N=588	5 ^a : 60,00 6 ^a : 32,61 7 ^a : 12,62 8 ^a : 5,92 N=441
TOTAL	N = 10.562	N = 8.932	N = 7.220	N = 5.802	N = 4.488

Fonte: Elaboração própria com base na Ficha B - CEDEPLAR/INEP 1999/2003.

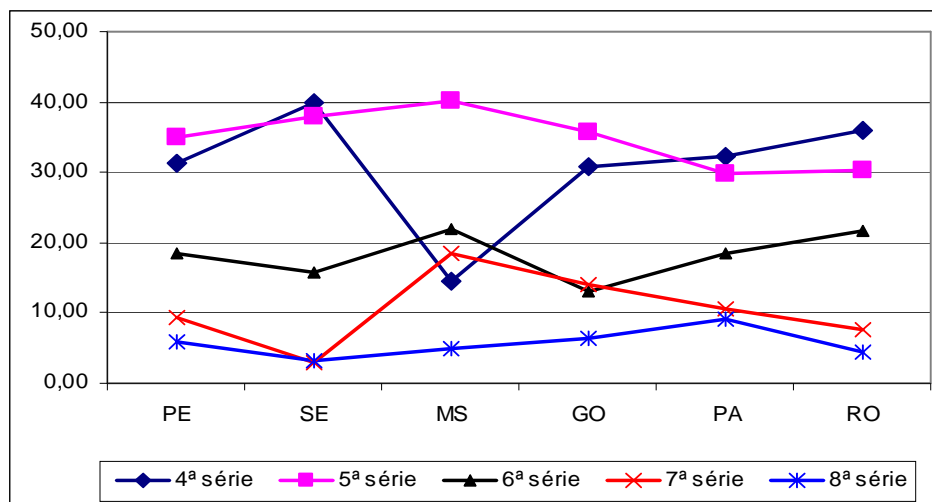
Em 1999 as taxas de reprovação referem-se à 4^a série. Em 2000, aos alunos que efetivaram matrícula na 4^a (reprovados em 1999 e repetentes em 2000) e na 5^a série (promovidos da 4^a para a 5^a série). No ano de 2001, as taxas correspondem aos alunos matriculados na 4^a, 5^a e 6^a série. E assim sucessivamente.

Uma análise segundo a série que deveria ser cursada no ano correspondente revela dois pontos importantes: i) taxas mais elevadas de reprovação são registradas nos estados da Região Nordeste e em Rondônia e; ii) a 5^a série é a que apresenta os maiores valores para o indicador analisado.

Além disso, pode ser constatado que quanto mais defasada a série cursada em relação à série ideal a cada ano específico, mais elevada é a taxa de reprovação. Esse resultado parece sugerir um desestímulo aos estudos por parte do aluno caso ele repita a série continuamente, resultando em nova reprovação.

No GRÁF.13 é apresentado o percentual de reprovação por séries para os alunos das escolas analisadas.

GRÁFICO 13
Percentual de reprovações por série, alunos da Ficha B, segundo UF's, 1999 a 2003



Fonte: Elaboração própria com base na Ficha B - CEDEPLAR/INEP 1999/2003.

As curvas revelam que o comportamento dos alunos em relação ao evento reprovação é bastante diferenciado, tanto entre as UF's quanto entre as séries cursadas. Enquanto os maiores percentuais de reprovação ocorreram nas séries iniciais (4ª e 5ª), os menores percentuais foram observados nas séries mais avançadas (7ª e 8ª), na quase totalidade das UF's.

Na próxima tabela são apresentados o total de reprovados e o percentual de reprovações na coorte segundo o número de reprovações, por UF. Os dados correspondem aos alunos presentes durante todo o período ou até a evasão ou transferência.

TABELA 4
Total e percentual de reprovações por UF, segundo alunos matriculados em 1999, presentes durante o período de acompanhamento: Ficha B, 1999-2003

Situação / UF	PE	SE	MS	GO	PA	RO	Total
01 reprovação	70,83	70,24	74,65	88,80	80,23	66,06	2265
02 reprovações	22,51	24,11	21,59	10,57	17,80	26,17	630
03 reprovações	5,99	5,35	3,76	0,55	1,97	7,00	131
04 reprovações	0,67	0,30	0,00	0,00	0,00	0,77	10
Total	902	336	533	368	511	386	3036

Fonte: Elaboração própria com base na Ficha B - CEDEPLAR/INEP 1999/2003.

De acordo com a última coluna da tabela é possível inferir que do total de reprovados entre a 4ª e a 8ª série nas escolas da amostra, ¼ tiveram uma reprovação, cerca de 21% tiveram duas reprovações e quase 4,7% tiveram entre três e quatro reprovações. Importante lembrar que a coorte inicial tinha 10.562 alunos e que menos de 45% desses alunos estiveram presentes durante todo o período acompanhado. Verifica-se, portanto, que o total de reprovados na coorte foi bastante significativo (cerca de 29%). A análise por UF revela, por um lado, que o maior percentual de alunos com mais de uma reprovação ocorreu nas escolas dos estados de Rondônia, Sergipe e Pernambuco, respectivamente. Por outro lado, o melhor resultado para esse indicador foi verificado entre os alunos matriculados nas escolas de Goiás.

A tabela seguinte apresenta dados sobre a situação de repetência dos alunos da coorte sob estudo.

TABELA 5
Total e percentual de alunos segundo a situação de repetência
entre a 4ª e 8ª série, por UF: Ficha B, 1999 a 2003

Situação / UF	PE	SE	MS	GO	PA	RO	TOTAL
Repetência por reprovação	92,19	92,91	99,73	95,69	96,68	96,92	95,14
Repetência por abandono	7,81	6,76	0,00	4,31	3,32	3,08	4,78
Repetência após aprovação	0,00	0,34	0,27	0,00	0,00	0,00	0,08
Total	781	296	368	232	331	357	2365

Fonte: Elaboração própria com base na Ficha B - CEDEPLAR/INEP 1999/2003.

Como visto anteriormente (Klein, 1995), a repetência numa série k num determinado ano t+1 ocorre em função da ocorrência de reprovação ou afastamento por abandono na série k no ano t. Além disso, também existem casos de repetência na série k no ano t+1 apesar de aprovação na série k no ano t. Na TAB. 5 pode ser verificado que a maioria das repetências, como esperado a priori, deveu-se à reprovação. Um percentual bem menos expressivo de casos foi associado ao afastamento por abandono na série anterior: a maioria dos registros nessa situação ocorreu entre os alunos das escolas situadas nos estados da região Nordeste. Houve apenas dois casos de repetência apesar da aprovação na série anterior. Comparando-se o valor total dessa tabela com o valor total da tabela anterior, percebe-se claramente que nem todos os alunos reprovados repetiram a série. A transferência e a evasão explicam grande parte desse hiato.

Nas tabelas 6 e 7 são apresentadas informações sobre o afastamento por abandono e sobre a evasão, respectivamente. No primeiro caso, o recorte é feito considerando-se a série cursada e, no segundo, considerando-se o evento que antecedeu a evasão (aprovação, reprovação ou afastamento por abandono).

TABELA 6
Total e percentual de alunos segundo a série de ocorrência do afastamento
por abandono entre a 4ª e 8ª série, por UF: Ficha B, 1999-2003

Série / UF	PE	SE	MS	GO	PA	RO
4ª série	23,53	7,81	12,50	27,78	9,21	21,05
5ª série	48,53	56,25	33,33	30,56	47,37	57,89
6ª série	16,91	18,75	41,67	27,78	18,42	21,05
7ª série	10,29	15,63	12,50	13,89	21,05	0,00
8ª série	0,74	1,56	0,00	0,00	3,95	0,00
Total	136	64	24	36	76	19

Fonte: Elaboração própria com base na Ficha B - CEDEPLAR/INEP 1999/2003.

TABELA 7
Total e percentual de alunos segundo a situação de evasão,
entre a 4ª e 8ª série, por UF: Ficha B, 1999-2003

Situação / UF	PE	SE	MS	GO	PA	RO	Total
Evasão após aprovação	66,31	78,67	66,67	70,83	68,69	68,75	69,23
Evasão após afastamento por abandono	31,26	20,85	28,89	24,31	30,30	29,17	28,63
Evasão após reprovação	2,43	0,47	4,44	4,86	1,01	2,08	2,14
Total	659	211	45	144	297	48	1404

Fonte: Elaboração própria com base na Ficha B - CEDEPLAR/INEP 1999/2003.

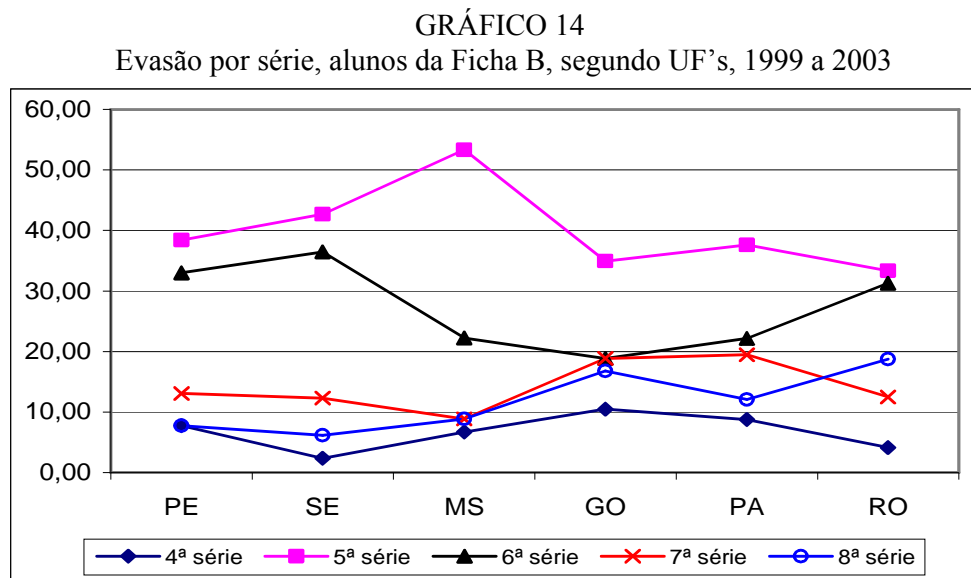
Na TAB. 6 é possível verificar que a 5ª série se destaca por ser a série de maior registro de ocorrências de afastamento por abandono, entre os alunos das escolas da amostra. A exceção ocorre entre os alunos das escolas do Mato Grosso do Sul, cuja série de maior ocorrência do evento foi a 6ª. Importante mencionar que para a 8ª série o percentual de alunos com registros de abandono pode estar subestimado devido ao tratamento dado às informações (conforme capítulo anterior).

Observa-se, na TAB. 7, que os maiores percentuais de casos de evasão na coorte ocorreram após aprovação do aluno numa determinada série, sendo tais percentuais bastante expressivos (sobretudo no estado de Sergipe). Em geral, a evasão é verificada de forma mais intensa após a reprovação do aluno numa determinada série. Que fatores explicam esse resultado observado para a coorte? A expectativa é que os modelos de regressão a serem estimados respondam a esse questionamento.

De acordo com KLEIN (1995) o aluno pode evadir, no ano t+1, após aprovação, reprovação ou afastamento por abandono na série k, no ano t. Pela tabela acima, observa-se que o percentual de evadidos aprovados foi superior ao percentual de evadidos reprovados em todas as UF's. Esse resultado requer uma análise mais aprofundada sobre os determinantes da evasão escolar. Em geral, espera-se que a maioria dos casos de evasão

ocorra após a reprovação numa determinada série, fato não observado entre os alunos da coorte analisada.

No gráfico seguinte são apresentadas as séries com o maior registro de evasão, no período.



Fonte: Elaboração própria com base na Ficha B - CEDEPLAR/INEP 1999/2003.

A visualização gráfica revela um padrão de evasão diferenciado entre as séries e as UF's em estudo. Observa-se que o maior percentual de evasões ocorreu na 5ª série nas escolas das seis UF's. O estado do Mato Grosso do Sul se destacou por apresentar a maior porcentagem de casos, nessa série. A próxima série com o maior percentual de casos foi a 6ª, sendo esse percentual mais significativo nos estados da região Nordeste e no estado de Rondônia.

Na seqüência, são apresentados os resultados da análise de sobrevivência.

6.3 Análise de sobrevivência: resultados e discussão

Na tese, uma hipótese a ser testada é se as funções de sobrevivência para os eventos de interesse são iguais para os alunos das escolas das UF's sob estudo. Pretende-se comparar as curvas estimadas, visando verificar se os alunos matriculados em escolas situadas em UF's mais desenvolvidas têm um maior tempo de sobrevivência aos eventos e se esses alunos apresentam um maior tempo de permanência na escola (seja no estado de aprovado ou não).

Para responder a estas questões, são estimadas as funções de sobrevivência aos eventos: i) reprovação; ii) repetência; iii) afastamento por abandono e; iv) evasão. Enquanto os tópicos (i) e (iii) referem-se a eventos ocorridos no ano letivo corrente que independem do resultado ocorrido no ano anterior, os tópicos (ii) e (iv) são eventos ocorridos no ano t+1, porém, relacionados à ocorrência do resultado escolar observado no ano letivo t. As funções de sobrevivência foram obtidas pelo estimador de Kaplan-Meier. Os resultados dos testes de Log-Rank são apresentados com o objetivo de verificar se as diferenças entre as curvas são estatisticamente significativas.

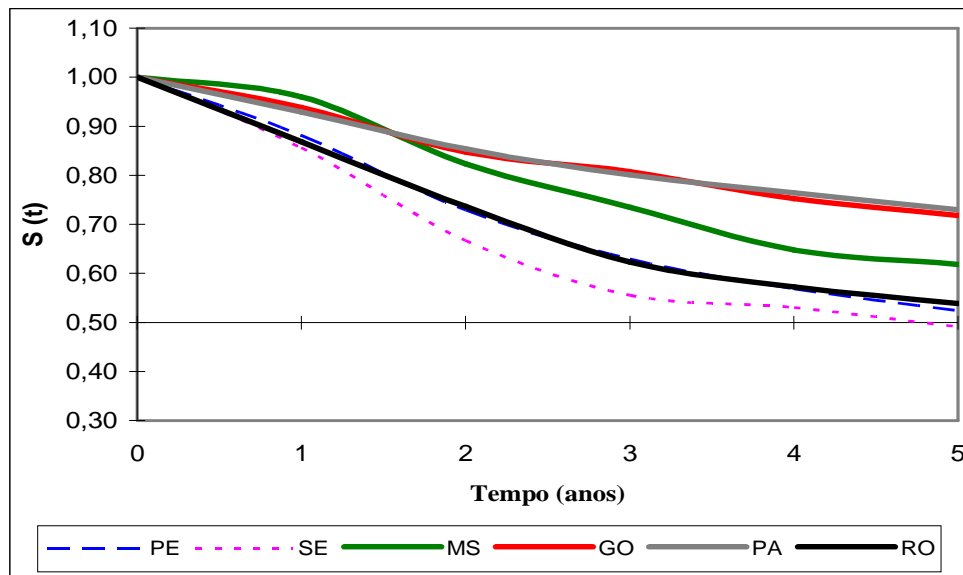
6.3.1 Curvas de sobrevivência à primeira reprovação e à primeira repetência

Foi visto anteriormente que um percentual expressivo de alunos foi reprovado em alguma(s) série(s) entre 1999 e 2003. Em média, aproximadamente $\frac{3}{4}$ desses alunos tiveram apenas uma reprovação. Assim, como a reprovação ocorreu mais de uma vez entre parte dos alunos, deve-se ressaltar que a presente análise compreende apenas a ocorrência do evento pela primeira vez entre a 4ª e a 8ª série do ensino fundamental. Uma vez que em geral a reprovação numa série, num determinado ano, conduz à repetência da série no ano seguinte, é interessante analisar também aspectos relacionados à sobrevivência do aluno à repetência. De forma análoga, a análise de sobrevivência a esse evento compreende somente a sua ocorrência pela primeira vez entre 1999 e 2003.

Nos gráficos 15 e 16 são apresentadas as curvas de sobrevivência à primeira reprovação e à primeira repetência. Estas curvas possibilitam verificar, entre as diversas UF's, aquelas cujos alunos permaneceram mais tempo sem vivenciar tais eventos¹⁹.

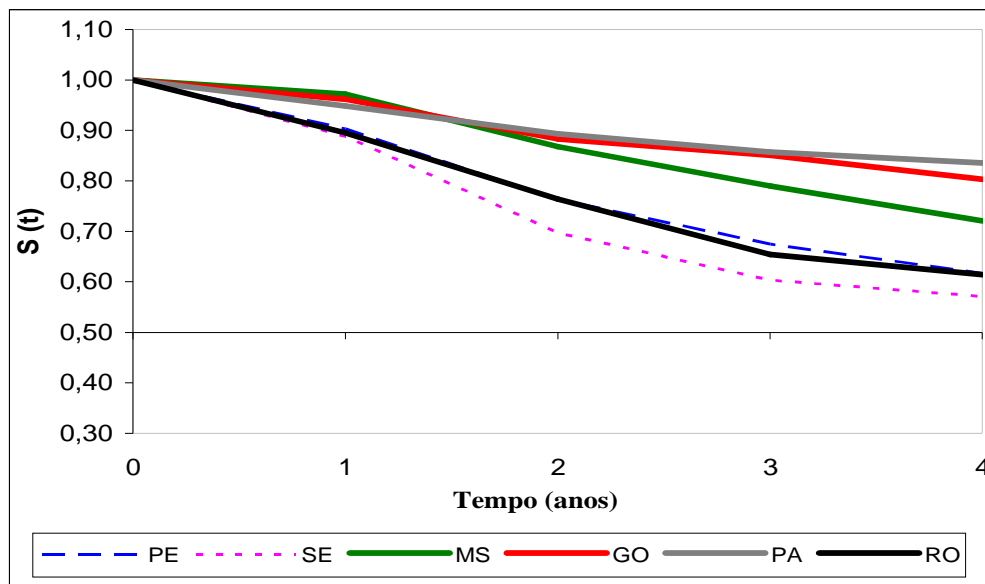
¹⁹ Na análise das funções de sobrevivência analisa-se a probabilidade do aluno sobreviver ao evento em estudo durante o período de acompanhamento da coorte (1999 a 2003) e anos posteriores. No texto, embora muitas vezes não seja mencionada a parte “e anos posteriores”, fica-se subentendido.

GRÁFICO 15
 Curvas de sobrevivência à 1ª reprovação
 entre a 4ª e 8ª série, por UF's, 1999 a 2003



Fonte: Elaboração própria com base na Ficha B – CEDEPLAR/INEP 1999/2003.

GRÁFICO 16
 Curvas de sobrevivência à 1ª repetência
 entre a 4ª e 8ª série, por UF's, 1999 a 2003



Fonte: Elaboração própria com base na Ficha B – CEDEPLAR/INEP 1999/2003.

Nota: 1999/2000: t=1; 2000/2001: t=2; 2001/2002: t=3 e 2002/2003: t=4

Através do GRÁF. 15 constata-se que as curvas de sobrevivência à primeira reprovação decrescem rapidamente para os anos iniciais do estudo para todas as UF's. As menores probabilidades de sobrevivência ao evento foram registradas para os alunos das

escolas de Sergipe, Rondônia e Pernambuco, respectivamente. Em contrapartida, os alunos que estudavam nas escolas de Goiás e do Pará permaneceram mais tempo sem vivenciar o evento, ou seja, até o fim do período de estudo parte significativa desses alunos não tiveram nenhuma reprovação nas séries analisadas.

As curvas de sobrevivência para as escolas de Goiás e Pará, bem como para Pernambuco e Rondônia, praticamente ficaram sobrepostas em todos os anos, indicando uma similaridade no comportamento dos alunos, em relação ao evento. Mas, de uma forma geral, pelo teste de Log-Rank (valor-p = 0,000) as diferenças entre as curvas são significativas, indicando que o tempo de sobrevivência do aluno à primeira reprovação é estatisticamente diferente entre as escolas das UF's.

No gráfico 16 foram mostradas as curvas de sobrevivência à primeira repetência. Estas curvas são estatisticamente diferentes para as escolas das UF's, conforme o teste de Log-Rank (valor-p = 0,000). Como esperado, foram bastante similares os resultados observados referente à primeira reprovação e à primeira repetência, embora nem todos os reprovados no ano t tenham repetido a série no ano t+1 (como visto anteriormente).

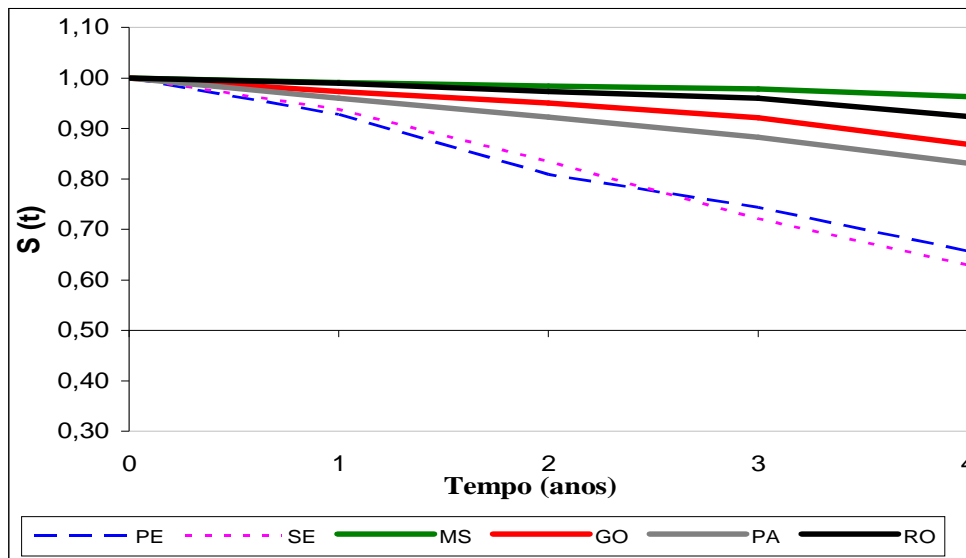
As curvas de sobrevivência à evasão e ao primeiro afastamento por abandono, para os alunos das escolas da amostra, podem ser visualizadas a seguir.

6.3.2 Curvas de sobrevivência à evasão e ao primeiro afastamento por abandono

No GRÁF. 17 observam-se curvas de sobrevivência à evasão bem distintas entre algumas UF's. Pelo teste de Log-Rank (valor-p = 0,000) rejeita-se a hipótese nula de que o tempo de sobrevivência à evasão seja o mesmo entre os alunos das diversas escolas.

A probabilidade de sobrevivência ao evento, entre a 4ª e a 8ª série do ensino fundamental, foi superior a 90% para os alunos das escolas de Goiás, Rondônia e Mato Grosso do Sul. Essa probabilidade foi bem menor para os alunos das escolas de Sergipe e Pernambuco (69% e 73% respectivamente), sendo aproximadamente igual a 87% para os alunos das escolas do Pará.

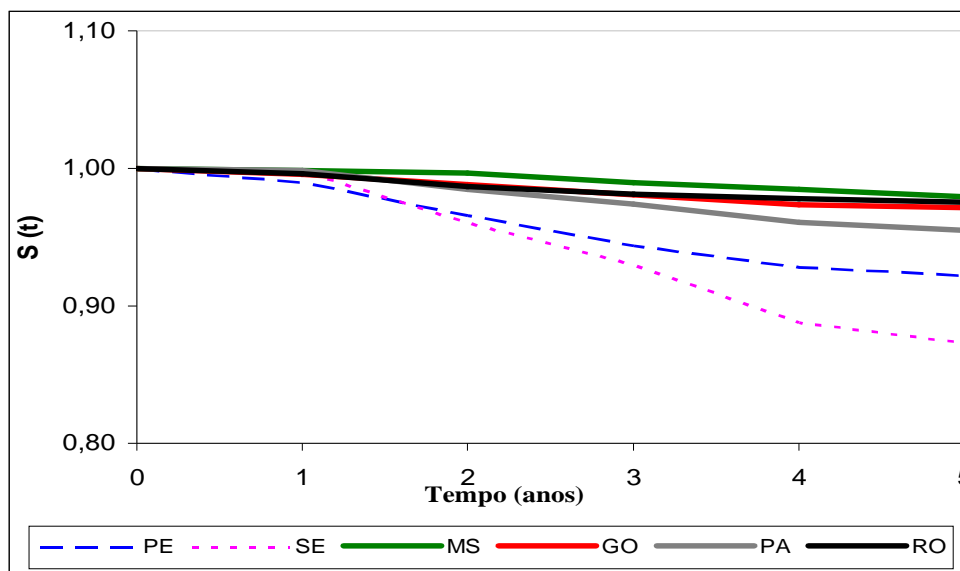
GRÁFICO 17
 Curvas de sobrevivência à evasão
 entre a 4ª e 8ª série, por UF's, 1999 a 2003



Fonte: Elaboração própria com base na Ficha B – CEDEPLAR/INEP 1999/2003.
 Nota: 1999/2000: t=1; 2000/2001: t=2; 2001/2002: t=3 e 2002/2003: t=4

As funções de sobrevivência ao primeiro afastamento por abandono para os alunos da amostra são apresentadas no GRÁF. 18.

GRÁFICO 18
 Curvas de sobrevivência ao 1º afastamento por abandono
 entre a 4ª e 8ª série, por UF's, 1999 a 2003



Fonte: Elaboração própria com base na Ficha B – CEDEPLAR/INEP 1999/2003

O GRÁF. 18 revela que a probabilidade de sobrevivência ao primeiro afastamento por abandono, entre as séries analisadas, foi relativamente elevada entre os alunos das escolas sob estudo. Ao término do período de acompanhamento da coorte, a menor e a maior probabilidade de sobrevivência à ocorrência do evento foram observadas entre os alunos das escolas do estado de Sergipe (87%) e do estado do Mato Grosso do Sul (98%), respectivamente.

A diferença entre as funções de sobrevivência é bem acentuada para a maioria das escolas/UF's. Isso é evidenciado pelo resultado do teste de Log-Rank (valor-p = 0,000). Nos anos finais do estudo evidencia-se um expressivo distanciamento entre as seis curvas, sugerindo um comportamento mais diferenciado dos alunos das escolas das seis UF's com relação ao evento. Evidencia-se também que há uma sobreposição de curvas para os alunos das escolas dos estados de Rondônia e Goiás durante praticamente todo o período de acompanhamento, indicando que os alunos dessas escolas tiveram um comportamento bem parecido relacionado ao abandono.

Comparando-se os eventos evasão e afastamento por abandono, notam-se funções de sobrevivência bastante diferentes entre os alunos das escolas das UF's analisadas. Essa diferença é observada tanto em termos proporcionais quanto relacionados ao formato e velocidade de declínio das curvas.

Outro objetivo da tese a ser contemplado pelo método em pauta é verificar se os alunos matriculados na 4ª série, em 1999, com idade superior à ideal para cursar a série, apresentam funções de sobrevivência diferenciadas em relação aos alunos matriculados na idade adequada, para os eventos de interesse. Esta questão é tratada na seqüência.

6.3.3 Curvas de sobrevivência aos eventos, segundo situação de defasagem idade-série

São apresentadas as curvas de sobrevivência para cada evento em estudo (primeira reprovação, primeira repetência, evasão e primeiro afastamento por abandono, respectivamente), considerando-se a situação de defasagem idade-série observada na 4ª série, no ano de 1999, para a coorte de alunos.

6.3.3.1 Curvas de sobrevivência à primeira reprovação e à primeira repetência, segundo situação de defasagem idade-série.

As funções de sobrevivência à primeira reprovação e à primeira repetência, entre a 4ª e a 8ª série, por situação de defasagem idade-série, podem ser acompanhadas na seqüência.

GRÁFICO 19

Curvas de sobrevivência à 1ª reprovação entre a 4ª e 8ª série, por UF's, 1999-2003: alunos sem defasagem idade-série

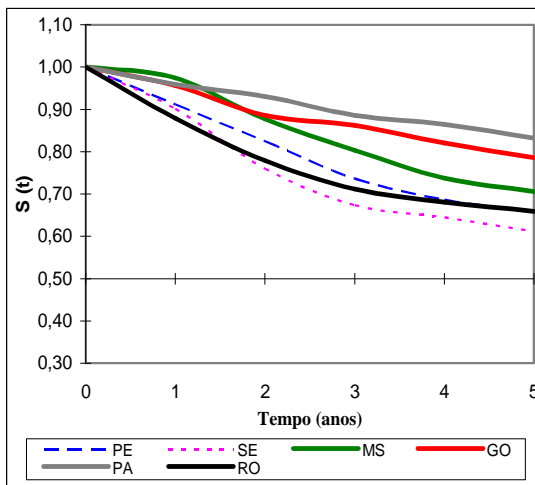
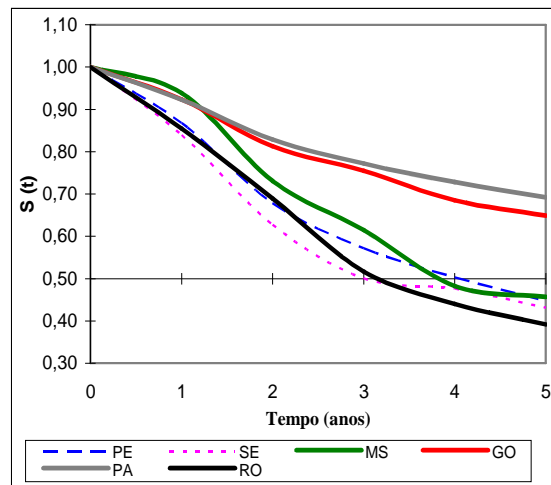


GRÁFICO 20

Curvas de sobrevivência à 1ª reprovação entre a 4ª e 8ª série, por UF's, 1999-2003: alunos com defasagem idade-série



Fonte: Elaboração própria com base na Ficha B – CEDEPLAR/INEP 1999/2003.

GRÁFICO 21

Curvas de sobrevivência à 1ª repetência entre a 4ª e 8ª série, por UF's, 1999-2003: alunos sem defasagem idade-série

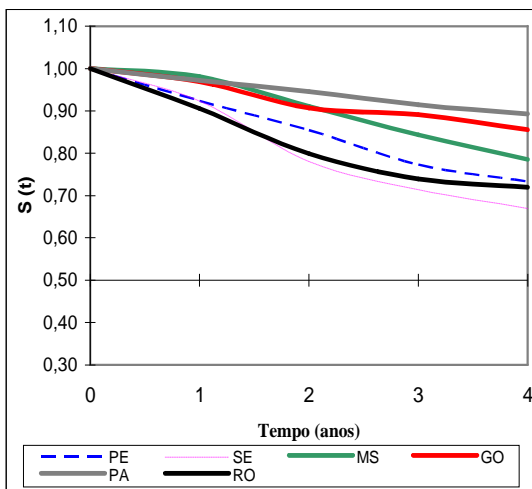
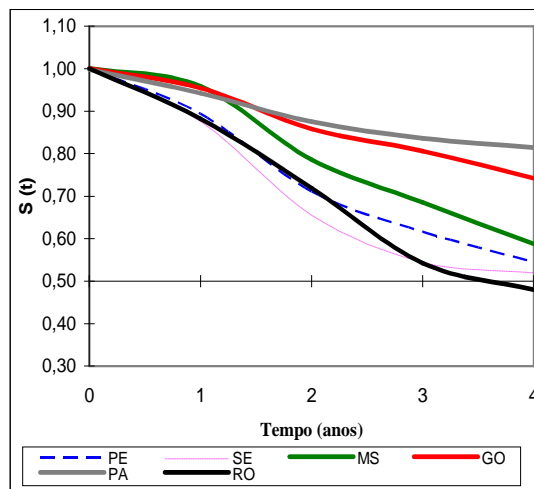


GRÁFICO 22

Curvas de sobrevivência à 1ª repetência entre a 4ª e 8ª série, por UF's, 1999-2003: alunos com defasagem idade-série



Fonte: Elaboração própria com base na Ficha B – CEDEPLAR/INEP 1999/2003.

Devido à similaridade entre as funções de sobrevivência para os eventos reprovação e repetência, a análise é feita conjuntamente.

Considerando-se as situações “alunos sem defasagem idade-série” e “alunos com defasagem idade-série” é possível verificar, no primeiro caso e para os dois tipos de eventos, uma maior proximidade entre as curvas, sendo que no segundo caso estas curvas apresentam inclinações mais diferenciadas, sendo também mais dispersas. Esse resultado sugere que a velocidade de ocorrência dos eventos reprovação e repetência, entre 1999 e 2003, é bem maior para os alunos que se matricularam mais tardiamente na 4ª série, no ano de 1999, sendo que essa velocidade varia entre as UF's. As diferenças entre as curvas de sobrevivência para todos os gráficos foram estatisticamente significativas (valor-p: 0,0000).

A seguir são apresentados os gráficos para os eventos evasão e afastamento por abandono segundo a defasagem idade-série.

6.3.3.2 Curvas de sobrevivência à evasão e ao primeiro afastamento por abandono, segundo situação de defasagem idade-série.

De acordo com os gráficos 23 e 24, se o aluno foi matriculado na idade adequada para cursar a 4ª série, no ano de 1999, sua probabilidade de sobreviver à evasão entre a 4ª e a 8ª série e séries subsequentes é bem maior em relação aos alunos matriculados na 4ª série numa idade superior à ideal à esta série. O teste de Log-Rank (valor-p: 0,0000) apontou que há diferenças significativas entre as curvas.

GRÁFICO 23

Curvas de sobrevivência à evasão entre a 4ª e 8ª série, por UF's, 1999-2003: alunos sem defasagem idade-série

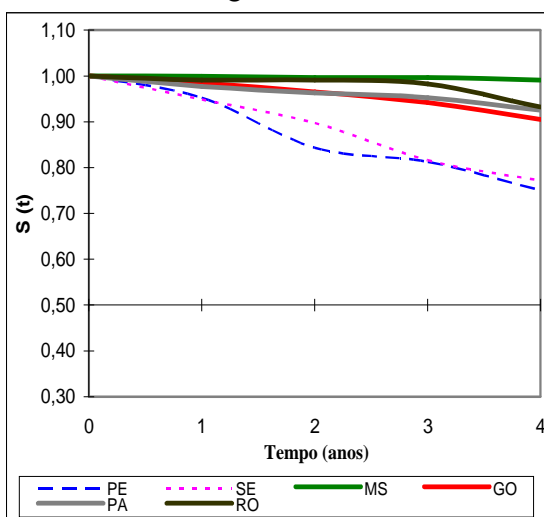
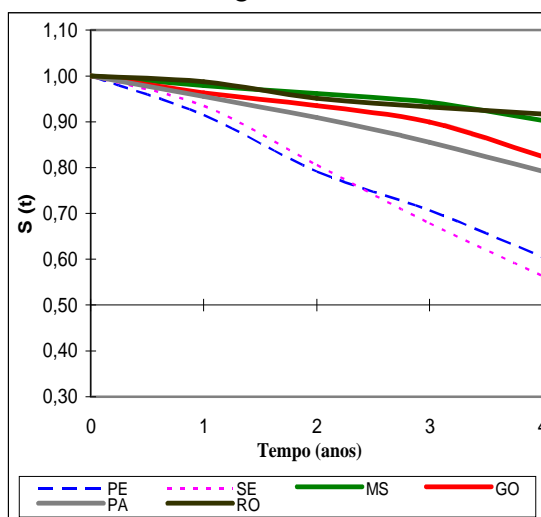


GRÁFICO 24

Curvas de sobrevivência à evasão entre a 4ª e 8ª série, por UF's, 1999-2003: alunos com defasagem idade-série



Fonte: Elaboração própria com base na Ficha B – CEDEPLAR/INEP 1999/2003.

GRÁFICO 25

Curvas de sobrevivência ao 1º abandono entre a 4ª e 8ª série, por UF's, 1999-2003: alunos sem defasagem idade-série

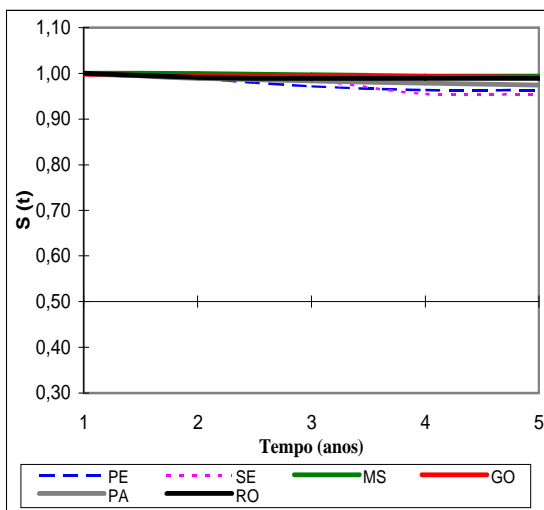
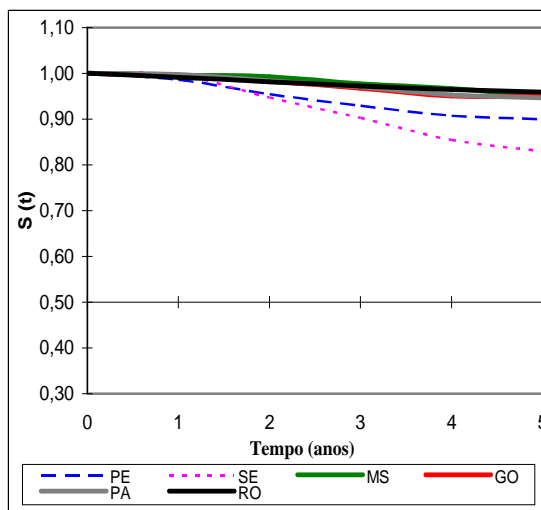


GRÁFICO 26

Curvas de sobrevivência ao 1º abandono entre a 4ª e 8ª série, por UF's, 1999-2003: alunos com defasagem idade-série



Fonte: Elaboração própria com base na Ficha B – CEDEPLAR/INEP 1999/2003.

A visualização dos gráficos 25 e 26 mostra que a probabilidade de sobrevivência do aluno ao primeiro afastamento por abandono, se ele se matriculou na idade adequada na 4ª série em 1999 é bem mais elevada (e próxima a 1) em relação aos alunos matriculados em idade superior à adequada à série. Isso é válido para todas as UF's. Em ambos os casos

nota-se que as curvas de sobrevivência para os alunos das UF's de Mato Grosso do Sul, Rondônia, Pará e Goiás ficaram praticamente sobrepostas ao longo do período. Maiores diferenciais nas curvas foram observados entre os alunos das escolas de Pernambuco e Sergipe, que se destacaram por apresentar as menores probabilidades de sobrevivência ao evento. De uma forma geral, as diferenças entre as curvas foram significativas estatisticamente, pelo teste de Log-Rank (valor-p: 0,0000).

Com relação aos eventos de fluxo analisados, foi possível verificar que a diferença entre as curvas de sobrevivência à primeira repetência foi bem maior do que a diferença entre as curvas de sobrevivência à evasão, entre a 4^a e a 8^a série. Esse aspecto revelado pelo método de análise revela que o comportamento do aluno frente a tais eventos é bem diferenciado entre os alunos das seis UF's. Isso requer um estudo mais aprofundado sobre os fatores responsáveis por esse diferencial. Isso se deve às características dos alunos (passadas e contemporâneas)? Ou aos recursos existentes nas escolas (físicos e humanos)? A resposta a esta questão é dada mediante a estimação dos modelos de regressão logísticos hierárquicos longitudinais, apresentados e discutidos no próximo capítulo.

7. ANÁLISE HIERÁRQUICA LOGÍSTICA LONGITUDINAL: IDENTIFICAÇÃO E DISCUSSÃO DOS DETERMINANTES DA PRIMEIRA REPETÊNCIA²⁰ E DA EVASÃO ENTRE A 4ª E A 8ª SÉRIE NO ENSINO FUNDAMENTAL

Os resultados dos modelos estimados são apresentados e discutidos neste capítulo. Na primeira seção são analisados os modelos cuja variável-resposta corresponde à probabilidade do aluno repetir a série, pela primeira vez, entre a 4ª e a 8ª série do ensino fundamental. Na segunda seção são abordados os modelos cuja variável-resposta equivale à probabilidade de evasão entre essas séries.

A idéia inicial era manter o mesmo conjunto de covariáveis e a mesma especificação dos modelos para ambos os eventos, tendo em vista a comparabilidade dos resultados. Contudo, no processo de estimação foram evidenciadas diferenças significativas no tempo de ocorrência para cada evento específico. Em função dessa especificidade revelada pelos dados, a variável Tempo teve que ser especificada de forma diferenciada. Mesmo assim, foi possível responder ao seguinte questionamento: os fatores determinantes da ocorrência da repetência são os mesmos que determinam a ocorrência da evasão, no ensino fundamental? Um paralelo entre os resultados das regressões para os dois eventos básicos é realizado no encerramento do capítulo. Foi feito também um paralelo entre alguns resultados obtidos com base nos dados descritivos e nos dados resultantes dos modelos estimados, com ênfase sobre a série de ocorrência dos eventos, destacando-se a importância de um estudo longitudinal para os dados educacionais.

7.1 Probabilidade de repetência entre a 4ª e 8ª série do ensino fundamental

No processo de identificação dos fatores determinantes da probabilidade de ocorrência da primeira repetência entre a 4ª e a 8ª série foram consideradas variáveis relacionadas aos alunos e às escolas. Entre as variáveis de alunos, algumas são fixas no tempo, enquanto outras variam ao longo do tempo, o que requer a especificação dos modelos considerando os níveis intra e inter-alunos. A pretensão é identificar, entre esse conjunto de variáveis, aquelas que mais contribuem para explicar a ocorrência do evento. Inicialmente, foi estimado o modelo sem a presença de covariáveis (modelo incondicional), conforme TAB. 8.

²⁰ Para não ficar um texto muito repetitivo, muitas vezes na análise será omitida a palavra “primeira”, mantendo-se a expressão ocorrência da repetência entre a 4ª e a 8ª série, ou probabilidade do evento ocorrer entre a 4ª e a 8ª série.

TABELA 8

Resultado do modelo incondicional para a probabilidade da primeira repetência entre a 4ª e a 8ª série do ensino fundamental: alunos da Ficha B, 1999-2003

Parâmetro	Estimativa	Desvio-padrão	% variabilidade atribuída aos níveis ²¹
* Efeito fixo			
Intercepto: β_{1j}	-2.490	0.070	
* Efeitos aleatórios			
Entre escolas: v_{ij}	0.547	0.082	14,26
Inter-alunos: u_{ij}	3.076	0.117	48,32
Intra-alunos: e_{0ij}	1.000	0.000	-
N = 8.948			

Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005?) e Censo Escolar de 1999.

De acordo com este modelo incondicional, a variabilidade na variável-resposta pode ser atribuída tanto aos fatores relacionados aos alunos quanto às escolas em que estão matriculados. Pode ser notado que a variabilidade entre os alunos dentro das escolas corresponde a mais de três vezes a variabilidade existente entre as escolas.

Após a estimação do modelo incondicional foram estimados dois grupos de modelos. No primeiro (composto por quatro equações), a variável que designa diretamente a trajetória longitudinal do evento é o Tempo, enquanto no segundo (composto por duas equações) é a série cursada. No caso da repetência há uma correspondência entre tempo e série, pois o dado é censurado após a primeira ocorrência. Dada essa equivalência entre as variáveis, a opção por estimar regressões substituindo o tempo pela série teve o objetivo de explicitar a série com maior probabilidade de ocorrência do evento, bem como verificar se o seu efeito sobre a repetência é o mesmo ou não entre as diversas escolas, o que propicia uma maior clareza sobre o evento em estudo.

Os resultados para o primeiro grupo de modelos são apresentados na TAB. 9. O primeiro modelo inclui apenas o tempo associado à ocorrência da repetência. Essa variável - Tempo - foi especificada nas regressões na forma quadrática. Com essa especificação, obteve-se um bom ajuste aos dados. No segundo modelo foram acrescentadas as covariáveis que variam no tempo (trabalho e proficiência anterior), relacionadas ao aluno. As variáveis fixas do aluno foram incluídas no terceiro modelo. Por último, os fatores escolares foram inseridos no quarto modelo.

²¹ A variância de uma distribuição logística cujo fator de escala é igual a 1 ($e_{0ij}=1.000$) é dada por $\pi^2/3 = 3.29$. Cálculo dos coeficientes intra-níveis: estimativa do efeito aleatório para cada nível / (estimativa do efeito aleatório + variância fator de escala).

TABELA 9

Modelos estimados para a probabilidade da primeira repetência entre a 4ª e a 8ª série do ensino fundamental, segundo o Tempo: alunos da Ficha B, 1999-2003.

	Modelo 1		Modelo 2		Modelo 3		Modelo 4	
	Coef.	d.p	Coef.	d.p	Coef.	d.p	Coef.	d.p
*Efeitos fixos								
Intercepto	-4.653*	0.102	-3.652	0.141	-3.301*	0.162	-0.837	0.756
Tempo	2.352*	0.039	0.869*	0.119	0.660*	0.143	1.164*	0.193
Tempo ²	-0.345*	0.007	-0.167*	0.025	-0.157*	0.031	-0.219*	0.035
Trabalho			0.215*	0.067	0.157**	0.072	0.162**	0.073
Proficiência aluno			-0.096*	0.004	-0.092*	0.004	-0.094*	0.004
Cor					-0.101***	0.060	-0.091	0.061
Sexo					-0.360*	0.060	-0.367**	0.061
Indicador NSE_1					0.009	0.030	0.005	0.030
Indicador NSE_2					-0.056**	0.024	-0.049**	0.025
Repetência antes 4ª					0.218*	0.060	0.196*	0.062
Ind.Segurança escola_1							-0.051	0.071
Ind.Segurança escola_2							-0.025	0.067
Ind.Estrutura escola_1							-0.033	0.077
Ind.Estrutura escola_2							-0.195*	0.066
Matrícula 4ª série							0.013*	0.005
Matric.4ª promovidos							-0.016*	0.006
Professor niv. superior							0.004	0.007
Qtde. salas							-0.002	0.016
Norte							-0.375**	0.186
Centro-Oeste							-0.324***	0.185
Nordeste							0.000	-
Profic. média escola							-0.054*	0.016
*Efeitos aleatórios								
Entre escolas: v_{1k}	0.969	0.146	0.631	0.093	0.553	0.082	0.454	0.072
Inter-alunos: u_{1jk}	13.075	0.210	2.462	0.117	0.500	0.108	0.761	0.107
Intra-alunos: e_{0ijk}	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
Var. Intercepto-NSE_2					0.000	0.006	0.000	0.000
Cov. Intercepto-NSE_2					0.048**	0.020	0.000	0.000
Coef.correl:entre escola	22,75	-	16.09	-	14.39	-	12,13	-
Coef.correl:inter-alunos	79,89	-	42.80	-	13.19	-	18.79	-
N = 8.948								

Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005?) e Censo Escolar de 1999.

Notas: i) categorias omitidas: cor (não-branca), sexo (masculino), repetência anterior à 4ª série (não repetiu), trabalho (não trabalha), série (7ª), região (Nordeste);, ii) Coef: coeficiente; d.p: desvio-padrão

***Coeficiente significativo a 10%; **coeficiente significativo a 5%; *coeficiente significativo a 1%.

No modelo 1 evidencia-se uma relação não linear e significativa entre o tempo e a probabilidade de ocorrência da primeira repetência entre a 4ª e a 8ª série²². Esse resultado indica que no início do período analisado (séries iniciais) a probabilidade de ocorrência do evento é crescente, decrescendo ao longo do tempo (das séries).

²² Embora seja mencionado primeira repetência entre a 4ª e a 8ª série, na realidade o evento ocorre entre a 4ª e a 7ª série, pois o estudo é interrompido em 2003.

No modelo 2 verifica-se que as variáveis que variam ao longo do tempo foram significativas, tendo o sinal esperado. Se o aluno trabalhou em algum ano letivo no período analisado, maior a sua probabilidade de repetir alguma série entre a 4ª e a 8ª. Ao incluir a proficiência do aluno centralizada na média da escola, a suposição é que o desempenho específico de um aluno está relacionado ao desempenho médio dos alunos da escola em que está matriculado. Portanto, supõe-se que há uma interação entre a proficiência do aluno e a proficiência média da escola. O sinal negativo para o coeficiente revela uma menor probabilidade de ocorrência do evento para os alunos com proficiência no ano anterior igual ou superior à média da escola. A inclusão dessas variáveis na regressão resultou numa redução expressiva nos valores dos coeficientes das variáveis relacionadas ao Tempo.

Com a inclusão das características fixas do aluno e do *background* familiar no modelo 3 houve algumas alterações nos valores dos coeficientes anteriormente estimados, sendo que a variável Trabalho teve sua significância estatística reduzida. Pode ser verificado que se o aluno tem a cor branca e é do sexo feminino, menor a sua probabilidade de repetir alguma série entre a 4ª e a 8ª. A variável referente ao *background* familiar - nível socioeconômico -, foi mensurada através de dois fatores²³. Apenas o coeficiente do indicador NSE_2 foi estatisticamente significativo, tendo o seu sinal negativo, indicando que alunos cujas famílias não têm a propriedade sequer dos bens básicos (rádio, televisão e geladeira) têm maior probabilidade de vivenciar a repetência. Para um melhor entendimento desse resultado (dado o sinal negativo para o coeficiente), é importante frisar que esse indicador trata-se de uma variável latente. A não posse dos bens básicos tem um peso negativo elevado na composição do fator. Um maior peso negativo está relacionado a um nível socioeconômico mais baixo. Portanto, há uma correlação negativa entre o indicador e a probabilidade de ocorrência do evento, pois quanto menor o nível socioeconômico familiar (medido pela não posse), maior a probabilidade para o evento ocorrer. A variável referente à trajetória escolar passada (repetência antes da 4ª série) foi positivamente correlacionada à probabilidade de repetência, conforme esperado à priori.

As variáveis de escola foram incluídas no modelo 4. Com essa inclusão, os coeficientes estimados do modelo anterior tiveram alterações, porém, mantiveram o sinal e a significância estatística, à exceção do coeficiente da variável cor, que tornou-se insignificante. Entre os fatores relacionados à escola, foram considerados dois índices de

²³ Os valores e as variáveis que compõem esses indicadores podem ser acompanhados no Quadro 5 (Anexos). Outras informações sobre a construção desses indicadores podem ser vistas no Cap.5 (págs. 63 e 66/67).

segurança. Nenhum deles teve significância estatística. De forma análoga, foram construídos dois indicadores associados à estrutura física da escola (com base na presença de quadra, laboratório de informática, sala de tv e vídeo e biblioteca), sendo que somente o segundo indicador foi significativo, tendo seu sinal negativo. Esse indicador é caracterizado, sobretudo, por uma associação forte e positiva entre a existência de laboratório de informática e a segunda dimensão, sinalizando que a existência de laboratório de informática na escola está associada a uma menor probabilidade de repetência. As estimativas para esse modelo indicam ainda que quanto maior o volume de matrículas efetivadas na 4ª série do ensino fundamental em 1999, nas escolas da amostra, maior a probabilidade do aluno repetir algum ano letivo entre 1999 e 2003. Em contrapartida, quanto maior o total de matrículas de alunos promovidos da 3ª para a 4ª série em 1999, menor essa probabilidade. No primeiro caso, duas inferências podem ser feitas: i) a quantidade de matrículas efetivadas envolve tanto alunos promovidos quanto repetentes, havendo indícios que o peso dos repetentes é significativo, no sentido de contribuir para aumentar a probabilidade de acontecimento do evento ou; ii) um maior número de matriculados na escola pode estar associado a maiores turmas, o que pode contribuir para um menor desempenho dos alunos e, conseqüentemente, um aumento no total de repetentes. No segundo caso, evidencia-se uma maior probabilidade de aprovação (principal causa da não repetência) numa determinada série se os alunos matriculados na escola no ano corrente tiverem sido aprovados na série cursada no ano letivo anterior (promoção). Esperar-se-ia que alunos que estudassem em escolas com um maior número de professores com nível superior e com um maior número de salas de aula tivessem uma menor propensão a repetir a série. Todavia, essas variáveis não foram significativas. Com relação às variáveis indicadoras de Região, observa-se que os alunos matriculados nas escolas situadas nas regiões Norte e Centro-Oeste apresentaram uma menor probabilidade de repetir uma determinada série entre a 4ª e a 8ª, em relação aos alunos das escolas do Nordeste.

Nesse modelo foi acrescentada a proficiência média dos alunos da escola. Essa variável capta o efeito do desempenho dos pares sobre a probabilidade de repetência de um aluno específico. O sinal negativo do coeficiente indica que quanto maior a proficiência média da escola, menor a probabilidade de repetência para um determinado aluno.

Uma importante mudança entre os dois últimos modelos estimados foi verificada no intercepto, cujo valor passou de -3.301 para -0.837, sendo que a estimativa que antes

era altamente significativa, deixa de ser. Na realidade, essa brusca alteração deveu-se à introdução da proficiência média da escola no modelo.

Uma vez discutidos os efeitos fixos para os coeficientes estimados, é importante fazer um breve comentário sobre os efeitos aleatórios. Em todos os modelos confirma-se a presença do efeito aleatório para o intercepto, dada a sua significância estatística (parte inferior da tabela). Pode ser notado também que houve uma grande redução na variabilidade associada aos níveis entre escolas e inter-alunos, quando são comparados o primeiro e o último modelo. Esse aspecto evidencia a importância dos fatores incluídos nas regressões visando melhor entender a ocorrência da repetência escolar.

No modelo 3 o coeficiente do indicador NSE_2 foi especificado como tendo efeito aleatório, com o intuito de verificar se o impacto do *background* familiar sobre a probabilidade do aluno repetir alguma série entre a 4ª e a 8ª varia entre as escolas. Foi confirmada a presença do efeito aleatório para o coeficiente, sendo que a covariância entre o intercepto e o indicador foi significativa, tendo sinal positivo. Esse resultado sugere que escolas cujos alunos são oriundos de famílias que não tem nem os bens duráveis básicos apresentam um relacionamento mais forte entre a probabilidade estimada de repetência e o nível socioeconômico familiar. Contudo, no modelo 4, ao se incluir a proficiência média da escola, esse efeito perdeu significância.

As regressões estimadas mostraram que a primeira repetência entre a 4ª e a 8ª série não ocorre de forma arbitrária, ao longo do tempo. Ao contrário, notou-se um comportamento bem definido para a variável Tempo, representado por uma função parabólica. Para uma melhor compreensão desse resultado é interessante associar o tempo de ocorrência do evento à série cursada. Como após a ocorrência do evento pela primeira vez os dados para cada aluno foram censurados, é possível associar os tempos 1 a 4 às séries 4ª à 7ª. Assim, no segundo grupo de modelos apresentado na TAB. 10, o Tempo é substituído pela série cursada. Para não ficar uma análise exaustiva, apenas o modelo 4 do grupo 1 foi re-especificado. Um modelo final - modelo 6 - foi estimado, especificando-se efeitos aleatórios para o coeficiente da 5ª série.

TABELA 10

Modelos estimados para a probabilidade da primeira repetência entre a 4ª e a 8ª série do ensino fundamental segundo a série cursada: alunos da Ficha B, 1999-2003.

	Modelo 5		Modelo 6	
	Coef.	d.p.	Coef.	d.p.
* Efeitos fixos				
Intercepto	0.178	0.974	-0.546	1.015
4ª série	-0.276***	0.166	-0.155	0.172
5ª série	0.285**	0.115	0.354**	0.140
6ª série	0.219**	0.107	0.220**	0.108
7ª série	0.000	-	0.000	-
Trabalho	0.159**	0.072	0.150**	0.073
Proficiência aluno	-0.095*	0.004	-0.095*	0.004
Cor	-0.092	0.061	-0.093	0.061
Sexo	-0.370*	0.061	-0.361*	0.061
Indicador NSE_1	0.004	0.030	0.004	0.030
Indicador NSE_2	-0.049***	0.026	-0.047***	0.026
Repetência antes 4ª	0.196*	0.062	0.196*	0.062
Ind.Segurança escola_1	-0.048	0.071	-0.039	0.070
Ind.Segurança escola_2	-0.027	0.067	-0.033	0.067
Ind.Estrutura escola_1	-0.032	0.077	-0.033	0.076
Ind.Estrutura escola_2	-0.197*	0.066	-0.194*	0.065
Matricula 4ª série	0.013*	0.005	0.014*	0.005
Matric. 4ª promovidos	-0.016*	0.006	-0.017*	0.005
Professor niv. Superior	0.004	0.007	0.005	0.007
Qtde. salas	-0.002	0.016	-0.002	0.016
Norte	-0.380**	0.186	-0.385**	0.184
Centro-Oeste	-0.342***	0.185	-0.383**	0.184
Nordeste	0.000	-	0.000	-
Profic. média escola	-0.050*	0.016	-0.038**	0.016
*Efeitos aleatórios				
Entre escolas: v_{1k}	0.453	0.072	0.488	0.082
Inter-alunos: u_{1jk}	0.933	0.109	0.919	0.109
Intra-alunos: e_{0ijk}	1.000	0.000	1.000	0.000
Variância Inclinação-NSE_2	0.000	0.000	0.000	0.000
Covariância Intercepto-NSE_2	0.000	0.000	0.000	0.000
Variância Inclinação-5ª série			0.742*	0.153
Covariância Intercepto-5ª série			-0.186*	0.085
Covariância NSE2-5ª série			0.000	0.000
Coeficiente correlação: entre escolas	12.10	-	12.92	-
Coeficiente correlação: inter-alunos	22.09	-	21.83	-
N= 8.948				

Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005?) e Censo Escolar de 1999.

Com a substituição do Tempo pela série, no modelo 5, a única variação significativa nas estimativas ocorreu no intercepto (que passou de -0.837 para 0.178). Os coeficientes das demais variáveis, quando não mantiveram seus valores, tiveram alterações mínimas. Esse resultado confirma a relação entre Tempo e série. As estimativas associadas

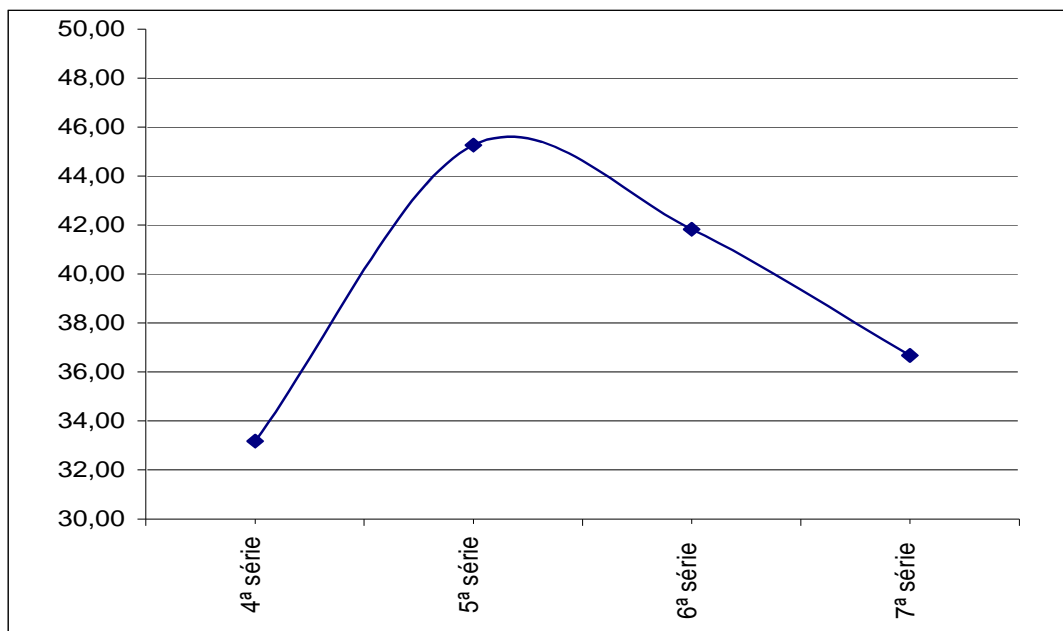
às indicadoras de série confirmam também a associação não linear entre a série cursada e o acontecimento do evento.

Como visto na análise descritiva, a 5ª série se destaca por conter o maior registro de repetentes. Na regressão estimada essa série se destaca também por apresentar o maior logaritmo da chance de ocorrência da repetência. Portanto, é interessante verificar se o seu efeito sobre a probabilidade do aluno repetir um ano letivo é o mesmo em todas as escolas. Nesse intuito, foi estimado o modelo 6. De acordo com os valores dos componentes da matriz de covariância (parte inferior da tabela), o efeito da 5ª série sobre a probabilidade de acontecimento do evento varia de acordo com as escolas. O sinal negativo da covariância sinaliza que escolas com menor proporção média de alunos repetentes tendem a apresentar um relacionamento mais forte entre a ocorrência do evento e a série. Noutras palavras, nas escolas cuja proporção de repetentes é menor fica mais evidenciada a relação entre a repetência e a 5ª série.

Nesse modelo foi mantido o coeficiente com efeito aleatório para o indicador NSE_2, visando verificar até que ponto a expressiva ocorrência do evento na 5ª série está associada ao nível socioeconômico familiar. Constatou-se um efeito não significativo na interação entre ambas as variáveis. Nesse sentido, as evidências apontam que os fatores relacionados à maior ocorrência do evento nessa série não são relacionados ao *background* familiar. Fatores escolares parecem ser os principais determinantes. Aliás, no modelo 4 foi visto que o efeito aleatório do coeficiente do indicador perdeu significância estatística após a inclusão da proficiência média da escola. É de se esperar que essa proficiência seja afetada por características específicas das escolas.

O GRÁF. 27 apresenta as probabilidades de repetência segundo a série cursada (modelo 6). Essas probabilidades são crescentes até a 5ª série, decrescendo posteriormente, revelando o comportamento parabólico da ocorrência do evento ao longo do tempo. Em relação à 7ª, a probabilidade do evento ocorrer é maior nas duas séries anteriores, sendo menor na 4ª série.

GRÁFICO 27
Probabilidades estimadas da primeira repetência entre a 4ª e a 8ª série
segundo a série cursada: alunos da Ficha B, 1999-2003



Fonte: Tabela 10.

A seguir, são discutidos os valores dos coeficientes do último modelo estimado que foram estatisticamente significativos. Deve-se ressaltar que no processo de estimação os coeficientes obtidos correspondem ao logaritmo da chance de ocorrência da primeira repetência entre a 4ª e a 8ª série para cada aluno (*log-odds*). Nesse estudo, o valor predito do *log-odds* foi convertido no valor predito da probabilidade de ocorrência do evento²⁴. Nesse processo de conversão foram considerados os valores médios para as variáveis incorporadas no modelo. No caso das variáveis contínuas, observou-se o efeito do aumento de um ponto acima desse valor médio sobre a probabilidade de repetência. Para as variáveis categóricas foram considerados os valores dos seus coeficientes. Na TAB. 11 são apresentados os coeficientes e respectivas probabilidades.

²⁴ Cálculos feitos através da função anti-logística para $X\beta$. Fórmula: $p = [1 + \exp(-X\beta)]^{-1}$

TABELA 11

Coefficientes significativos do modelo 6 e respectivas probabilidades da primeira repetência entre a 4ª e a 8ª série do ensino fundamental: alunos da Ficha B, 1999-2003.

Modelo 6			
	Coefficiente	Probabilidade	Prob - Med.
* Efeitos fixos			
Intercepto	-0.546	36.68	-
4ª série	-0.155	33.18	-3.50
5ª série	0.354**	45.26	8.58
6ª série	0.220**	41.82	5.14
7ª série	0.000	-	-
Trabalho	0.150**	40.13	3.45
Proficiência aluno	-0.095*	34.52	-2,15
Cor	-	-	-
Sexo	-0.361*	28.70	-7.98
Indicador NSE_1	-	-	-
Indicador NSE_2	-0.047***	35.66	-1.02
Repetência antes 4ª	0.196*	41.34	4.66
Ind.Segurança escola_1	-	-	-
Ind.Segurança escola_2	-	-	-
Ind.Estrutura escola_1	-	-	-
Ind.Estrutura escola_2	-0.194*	32.30	-4.38
Matricula 4ª série	0.014*	37.05	0.37
Matric. 4ª promovidos	-0.017*	36.35	-0.32
Professor niv. superior	-	-	-
Qtde. salas	-	-	-
Norte	-0.385**	28.29	-8.39
Centro-Oeste	-0.383**	28.29	-8.39
Nordeste	0.000	-	-
Profic. média escola	-0.038*	35.89	-0.79

Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005?) e Censo Escolar de 1999.

Como a função *link logit* não é linear a estimativa para o termo de intercepto (36,68) não corresponde exatamente à média geral, sendo mais adequado afirmar que ela equivale à proporção mediana de repetência numa escola típica entre a 4ª e a 8ª série (para um aluno não-branco, do sexo masculino, que não repetiu nenhuma série antes da 4ª e nem trabalhou no período analisado, que cursou a 7ª série e matriculado numa escola da região Nordeste; com valores para as variáveis contínuas iguais zero).

Na interpretação da probabilidade associada à cada variável as demais têm seus valores igualados a zero. A análise é feita com base na seqüência dos modelos estimados. Inicialmente, são analisadas as variáveis de aluno que mudam no tempo, além da série. Posteriormente, são discutidos os resultados para as variáveis de aluno que são fixas no tempo e, por último, as variáveis de escola.

No primeiro grupo de variáveis constata-se que a variável Trabalho, associada à trajetória contemporânea do aluno, tem forte impacto sobre a probabilidade estimada de repetência, numa escola típica: se o aluno trabalhou em algum ano letivo entre a 4ª e a 8ª série, esse valor correspondeu a 40,13%. Um dos objetivos propostos nessa tese está relacionado à importância do desempenho escolar passado sobre o resultado escolar corrente do aluno, sendo esse desempenho mensurado pela proficiência do aluno. Estimou-se uma probabilidade de repetência numa determinada série entre a 4ª e a 8ª igual a 34,52% para os alunos que tiveram um aumento de um ponto acima da média nos escores referentes à proficiência observada no ano letivo imediatamente anterior ao ano de ocorrência do evento.

Os coeficientes das indicadores de série, à exceção da 4ª, foram estatisticamente significativos. Sendo a categoria omitida a 7ª, verifica-se que a probabilidade de repetência foi bem mais significativa na 5ª série (45,26%). Esse resultado está coerente com as estatísticas existentes para o sistema de ensino brasileiro (Cap. 1) e com as evidências apontadas na análise descritiva da trajetória da coorte (Cap. 6).

Considerando-se o segundo grupo de variáveis, observa-se que entre os fatores demográficos apenas o sexo teve seu coeficiente significativo, sendo que a probabilidade estimada de repetência para as meninas correspondeu a 28,70%, estimativa bem inferior à calculada para os meninos. O *background* familiar também teve impacto significativo sobre a ocorrência do evento: se o escore referente ao nível socioeconômico familiar for aumentado em um ponto acima da média entre 1999 e 2003, estima-se que a probabilidade do aluno repetir a série pela primeira vez entre a 4ª e a 8ª seja equivalente a 35,66%.

Se por um lado o fato de ser do sexo feminino e de ter uma melhoria no nível socioeconômico familiar resulta em menores probabilidades estimadas de repetir o ano letivo, por outro, ter repetido alguma série antes da 4ª série é o fator de maior peso na explicação da ocorrência do evento entre a 4ª e a 8ª série. Esse resultado mostra a forte influência do resultado escolar passado do aluno sobre a sua trajetória escolar corrente.

Deve ser frisado que foi testado o efeito do tempo sobre as variáveis cor e nível socioeconômico. Mas os coeficientes não foram significativos, revelando que a trajetória escolar do aluno, considerando-se o efeito desses fatores sobre a repetência, não sofreu mudança significativa ao longo do tempo.

Com relação aos fatores escolares: entre 1999 e 2003, diante de um aumento de um ponto acima da média nos valores dos escores referente à estrutura física da escola, mensurada, sobretudo, pela presença de laboratório de informática, tem-se uma

probabilidade estimada de repetência para um aluno em torno de 32,30%²⁵. Para uma escola em que o total de matrículas na 4ª série tenha um aumento de uma unidade acima da média, essa estimativa equivaleu a 37,05%. Para igual aumento no total de matrículas de alunos promovidos da 3ª para a 4ª série, esse valor correspondeu a 36,35%. O valor estimado da probabilidade do aluno repetir a série pela primeira vez entre a 4ª e a 8ª foi igual a 28,29% se ele estava matriculado numa escola situada na região Norte ou Centro-Oeste. Essa estimativa foi muito inferior à calculada para os alunos das escolas do Nordeste.

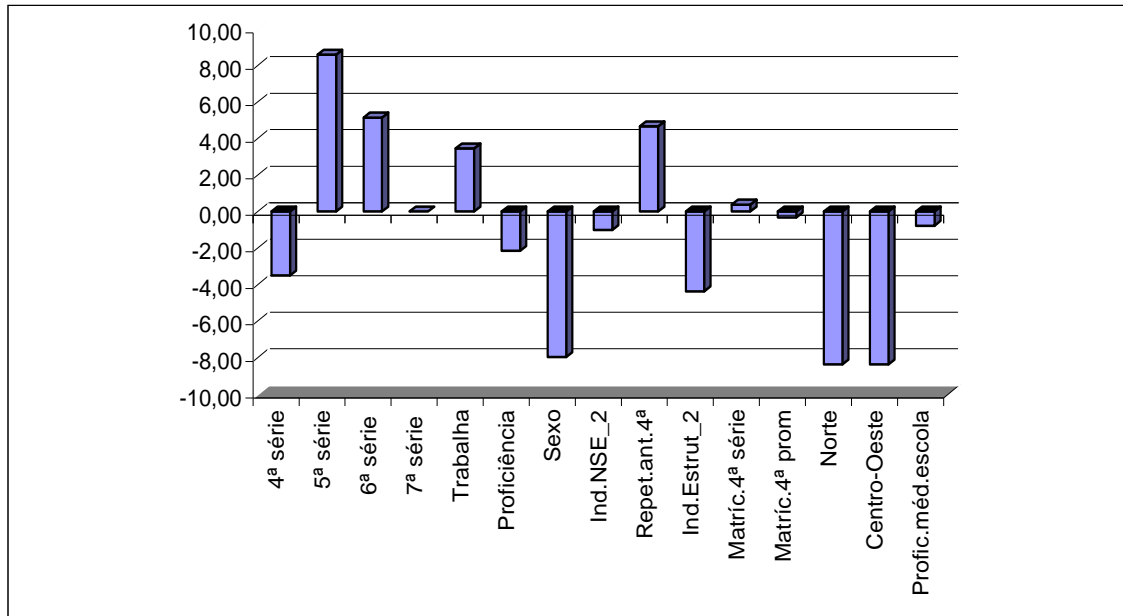
Finalmente, aumentando-se a proficiência média dos colegas da escola em um ponto acima da média, no ano subsequente tem-se uma probabilidade estimada de repetência da série equivalente a 35,89%. Essa estimativa sinaliza que se o aluno estuda numa escola cujos colegas têm um bom desempenho nas disciplinas (no caso específico, português e matemática), menor a sua probabilidade de repetir as séries finais do ensino fundamental.

Na seqüência é apresentado o GRÁF. 28, elaborado a partir da última coluna da TAB. 11. Ele permite uma melhor visualização do efeito de cada fator (inter-aluno, intra-aluno e de escola) sobre a probabilidade de ocorrência da primeira repetência entre a 4ª e a 8ª série para uma escola típica.

²⁵ É bom frisar que aumentos sucessivos no valor de cada variável contínua conduzem a menores reduções (ou aumentos) nas probabilidades estimadas, devido a não-linearidade da função estimada.

GRÁFICO 28

Efeito das variáveis de aluno e escola sobre a probabilidade média da primeira repetência entre a 4ª e a 8ª série do ensino fundamental: Ficha B, 1999-2003.



Fonte: Elaboração própria com base na TAB. 11.

Com base no gráfico, a probabilidade de ocorrência da primeira repetência entre a 4ª e a 8ª série é superior à proporção mediana se: i) a série cursada for a 5ª; ii) o aluno tiver trabalhado em algum ano no período analisado e; iii) o aluno tiver repetido algum ano letivo antes da efetivação da sua matrícula na 4ª série.

Em contrapartida, a probabilidade de ocorrência do evento é inferior à proporção mediana caso: i) a proficiência anterior do aluno seja aumentada; ii) o aluno seja do sexo feminino; iii) haja uma melhoria no nível socioeconômico familiar e; iv) ocorra uma melhoria na estrutura física da escola, bem como uma elevação na proficiência dos colegas da escola.

Além disso, verifica-se que o contexto regional impacta significativamente a ocorrência da repetência, tornando-a bem inferior à proporção mediana caso o aluno estude numa escola não localizada na região Nordeste.

Na próxima seção são apresentados os modelos para a variável-resposta evasão. Para fins comparativos, optou-se por manter o mesmo conjunto de covariáveis e o mesmo número de modelos estimados.

7.2 Probabilidade de evasão entre a 4ª e a 8ª série do ensino fundamental

Tal como no caso da análise da probabilidade de ocorrência da primeira repetência entre a 4ª e a 8ª série do ensino fundamental, na identificação dos determinantes da probabilidade de evasão entre essas séries foram consideradas variáveis de alunos fixas e variáveis no tempo (característica típica de dados longitudinais), além das variáveis de escola. Nessa seção, o mesmo procedimento foi adotado com relação à seqüência dos modelos estimados. O ponto de partida é, portanto, a estimação do modelo incondicional, especificado na TAB. 12.

TABELA 12
Resultado do modelo incondicional para a probabilidade de evasão
entre a 4ª e a 8ª série do ensino fundamental: alunos da Ficha B, 1999-2003.

Parâmetro	Estimativa	Desvio-padrão	% variabilidade atribuída aos níveis
* Efeito fixo			
Intercepto: β_{1j}	-3.181	0.124	
* Efeitos aleatórios			
Entre escolas: v_{ij}	2.117	0.267	39.15
Inter-alunos: u_{ij}	3.242	0.166	49.63
Intra-alunos: e_{0ij}	1.000	0.000	-
N=8.980			

Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005?) e Censo Escolar de 1999.

As estimativas relacionadas aos coeficientes intra-níveis indicam que a variabilidade entre as escolas é elevada, sendo um pouco menor em relação à variabilidade existente entre os alunos dentro das escolas. Como parte significativa da variabilidade na variável-resposta pode ser atribuída aos fatores relacionados à escola freqüentada pelo aluno, é justificada a inclusão do nível 3 nos modelos.

Na TAB. 13 são apresentados os resultados para o primeiro grupo de modelos. Alguns comentários analíticos são omitidos por já terem sido feitos anteriormente.

Deve ser ressaltado que a evasão, tal como a repetência, foi medida em quatro pontos no tempo. Observou-se se o evento ocorreu ou não no segundo ano de cada biênio e a série cursada no ano letivo anterior. Seguindo esse raciocínio, a última série relacionada à evasão é a 7ª, pois foi verificado se o aluno evadiu em 2003 e a série que ele cursou em 2002 (sendo ou não aprovado).

Um comentário preliminar será feito sobre a especificação da variável Tempo. Para a repetência, o melhor ajuste no modelo foi obtido pela inserção do tempo na forma quadrática. No caso presente, a trajetória temporal da evasão foi melhor representada por uma função polinomial de terceira ordem (estrutura cúbica para o tempo).

Na análise descritiva foi observado que a maioria dos casos de evasão na coorte aconteceu após a aprovação na série anterior. Para checar esse aspecto dos dados no modelo, verificou-se se o aluno evadiu no ano t+1 e o resultado escolar ocorrido no ano anterior, t. Mais precisamente, foi verificado o efeito do tempo sobre a ocorrência da evasão, controlando-se pela aprovação na série cursada no ano letivo anterior. Considerou-se também a interação entre o tempo e a aprovação. Ao especificar o modelo com a interação entre ambas as variáveis é possível fazer um importante diagnóstico sobre o comportamento do evento ao longo do período.

TABELA 13
Modelos estimados para a probabilidade de evasão entre a 4ª e a 8ª série do ensino fundamental, segundo o Tempo: alunos da Ficha B, 1999-2003.

	Modelo 1		Modelo 2		Modelo 3		Modelo 4	
	Coef.	d.p	Coef.	d.p	Coef.	d.p	Coef.	d.p
* Efeitos fixos								
Intercepto	-7.619*	0.163	-4.037*	0.605	-3.958*	0.613	-3.379*	1.296
Tempo	6.246*	0.134	3.105*	0.927	3.025*	0.934	2.965*	1.035
Tempo2	-2.092*	0.057	-1.528*	0.428	-1.488*	0.430	-1.459*	0.474
Tempo3	0.237*	0.008	0.227*	0.059	0.222*	0.059	0.215*	0.065
Aprovação			-1.032*	0.165	-1.052*	0.169	-1.107*	0.180
Tempo*aprovação			-0.221*	0.072	-0.205*	0.074	-0.162**	0.083
Trabalho			0.203*	0.084	0.147***	0.088	0.166***	0.098
Cor					-0.065	0.070	-0.039	0.077
Sexo					-0.138**	0.069	-0.150**	0.075
Indicador NSE_1					-0.019	0.035	-0.018	0.037
Indicador NSE_2					-0.060**	0.029	-0.063**	0.031
Repetência antes 4ª					0.118***	0.071	0.071	0.076
Ind.Segurança escola1							0.106	0.122
Ind.Segurança escola2							-0.332*	0.121
Ind. Estrutura escola1							0.155	0.125
Ind. Estrutura escola2							0.024	0.108
Matrículas 4ª série							0.006	0.009
Matric.4ª promovidos							-0.005	0.009
Professor niv. Superior							-0.020***	0.012
Qtde. salas							0.031	0.027
Norte							-1.446*	0.303
Centro-Oeste							-1.807*	0.303
Nordeste							0.000	-
Profic.média escola							0.002	0.023
* Efeito aleatório								
Entre escolas: v_{1k}	2.384	0.320	2.542	0.315	2.452	0.306	1.315	0.196
Inter-alunos: u_{ijk}	21.439	0.316	0.000	0.000	0.187	0.168	0.000	0.000
Intra-alunos: e_{0ijk}	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
Coef.corr: entre escolas	42.02	-	43.59	-	42.70	-	28.56	-
Coef.corr: inter-alunos	86.70		-	-	-		-	-
N=8.980								

Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005?) e Censo Escolar de 1999.

O primeiro modelo permite verificar que no início do período de acompanhamento da coorte a probabilidade de ocorrência da evasão é crescente, decrescendo ao longo do período para retomar o crescimento na etapa final do estudo, isto é, no fim do ensino fundamental.

No segundo modelo, juntamente com o Tempo, foram incluídas as variáveis de aluno que variam no tempo. Além da variável de controle aprovação, foi inserida também uma variável indicadora denotando a interação entre o Tempo e a aprovação. Observa-se uma correlação forte e negativa entre aprovação e evasão. Esse comportamento está de acordo com as expectativas, porém, contradiz os resultados apresentados na análise descritiva. Outra constatação relevante é que ao longo do tempo a probabilidade de evasão é diminuída caso o aluno seja aprovado nas séries anteriores. Esse resultado parece sugerir que o aluno é motivado a permanecer na escola diante de um desempenho escolar favorável. E, conforme esperado, há uma associação positiva e significativa entre a variável trabalho e a evasão escolar.

Os fatores relacionados aos alunos, que são fixos no tempo, foram acrescentados no terceiro modelo. Diante dessa inclusão, a principal alteração ocorrida em relação ao modelo anterior foi a redução da significância estatística da variável trabalho. Como esperado a priori, as meninas apresentaram uma menor probabilidade de evadirem da escola, em relação aos meninos. Caso o aluno tenha repetido alguma série antes da 4ª, maior essa probabilidade. De forma similar ao verificado no caso da repetência, apenas um dos indicadores de nível socioeconômico, o índice NSE_2, teve seu coeficiente significativo. Esse resultado indica que alunos cujas famílias são desprovidas dos bens básicos apresentaram uma maior probabilidade de saída do sistema escolar.

Com a inclusão dos fatores escolares no quarto modelo, as variáveis anteriormente incluídas mantiveram a significância estatística, sendo que a única exceção foi verificada para a repetência anterior à 4ª série, que tornou-se estatisticamente insignificante. Pode ser constatado que entre os fatores diretamente relacionados às escolas, apenas o indicador associado ao sistema de segurança existente (índice de segurança 2) e a escolaridade do professor (nível superior) foram relevantes para explicar a probabilidade da evasão. Como dito antes, o componente de maior poder explicativo nesse indicador é a existência de sistema de proteção contra incêndio. Em geral, escolas com menos recursos não são dotadas desse sistema, o que permite inferir que a sua existência está relacionada às melhores escolas. As estimativas para as indicadoras regionais mostraram que o contexto socioeconômico da região em que a escola está inserida influencia fortemente a decisão do

aluno quanto à permanência ou não na escola. Nesse modelo foi incluída também a proficiência média da escola, cujo coeficiente não teve significância estatística.

Observando-se os valores da parte inferior da tabela, admite-se a existência do efeito aleatório para o intercepto de todos os modelos, comprovado por sua significância estatística. Portanto, há variabilidade significativa entre as escolas no que se refere à probabilidade de evasão dos alunos. Nota-se que essa variabilidade foi reduzida no último modelo em função, basicamente, dos fatores escolares. Deve ser ressaltado que no modelo incondicional e no primeiro modelo estimado a parte aleatória inter-alunos era altamente expressiva. Com a inclusão das variáveis que variam ao longo do tempo, a variabilidade atribuída a esse nível tornou-se insignificante.

Foi testada a presença de efeito aleatório para o nível socioeconômico familiar, porém, verificou-se que o efeito dessa variável sobre a ocorrência da evasão foi o mesmo nas diferentes escolas. Daí, manteve-se o coeficiente como tendo apenas uma parte fixa. Como feito no caso da repetência, testou-se o efeito do tempo sobre as variáveis cor e nível socioeconômico. A idéia foi verificar se com o passar do tempo houve alguma variação no impacto dessas variáveis sobre a evasão. Contudo, os coeficientes estimados não foram estatisticamente significativos. Este teste foi realizado também para a variável trabalho e o resultado foi análogo. Nesse último caso, a expectativa era que no fim do período de estudo o efeito do trabalho sobre a evasão fosse mais significativo. Todavia, o total de alunos que trabalhou nos anos finais reduziu-se expressivamente em relação aos anos iniciais (TAB. 18, anexos). Esse fator, que acompanhou a redução da coorte no período, pode explicar o resultado não significativo obtido.

De acordo com os modelos estimados, a ocorrência da evasão entre a 4ª e a 8ª série apresentou um padrão não linear para a variável Tempo. Analogamente ao caso da repetência, é importante relacionar o tempo de ocorrência do evento à série cursada. Porém, diferentemente do caso anterior, não é possível associar o tempo à série de forma seqüencial (ou seja, tempo 1 = 4ª série, tempo 2 = 5ª série, etc), pois o aluno pode repetir a série “n” vezes até a saída do sistema escolar. De qualquer forma, a substituição do Tempo pela série cursada pode expressar o comportamento do evento ao longo das séries. Estimou-se um segundo grupo de modelos com essa substituição, cujos resultados são apresentados na TAB. 14. Como feito na seção anterior, apenas o modelo 4 foi re-estimado.

No modelo 5, apesar da não correspondência direta entre Tempo e série, observam-se valores bem parecidos para os coeficientes estimados, sinalizando que a discrepância

entre ambas as variáveis não é tão grande ao ponto de inviabilizar a análise. Uma vez que os valores dos coeficientes não sofreram alterações importantes, foi estimado o modelo 6, também com inserção da série cursada, sendo que a variável de controle aprovação foi substituída pela proficiência anterior do aluno, como feito para a repetência. A idéia é tornar mais comparável os resultados para ambos os eventos. Diante dessa substituição, importantes mudanças foram verificadas nas estimativas. Em geral, os coeficientes estimados tornaram-se mais significativos, exceto para o caso da escolaridade do professor, cujo coeficiente perdeu significância estatística.

TABELA 14
Modelos estimados para a probabilidade de evasão entre a 4ª e a 8ª série do ensino fundamental segundo a Série cursada: alunos da Ficha B, 1999-2003.

	Modelo 5		Modelo 6	
	Coef.	d.p.	Coef.	d.p.
* Efeitos fixos				
Intercepto	-1.651	1.395	-2.033	1.417
4ª série	-0.116	0.220	-0.137	0.221
5ª série	-0.179	0.146	-0.058	0.145
6ª série	-0.559*	0.144	-0.642*	0.182
7ª série	0.000	-	0.000	-
Aprovação	-1.410*	0.086	-	-
Proficiência aluno			-0.024*	0.005
Trabalho	0.165***	0.098	0.191**	0.099
Cor	-0.036	0.077	-0.072	0.077
Sexo	-0.149**	0.075	-0.197**	0.076
Indicador NSE_1	-0.015	0.037	-0.001	0.037
Indicador NSE_2	-0.064**	0.031	-0.083**	0.031
Repetência antes 4ª	0.071	0.076	0.089	0.076
Ind.Segurança escola1	0.108	0.122	0.066	0.113
Ind.Segurança escola2	-0.333*	0.121	-0.347*	0.113
Ind. Estrutura escola1	0.156	0.124	0.171	0.115
Ind. Estrutura escola2	0.025	0.108	0.031	0.100
Matrículas 4ª série	0.006	0.009	0.007	0.008
Matric. 4ª promovidos	-0.005	0.009	-0.007	0.008
Professor niv. Superior	-0.020***	0.012	-0.017	0.011
Qtde. salas	0.032	0.027	0.028	0.025
Norte	-1.445*	0.303	-1.412*	0.278
Centro-Oeste	-1.817*	0.302	-1.876*	0.283
Nordeste	0.000	-	0.000	-
Profic.média escola	0.007	0.022	-0.008	0.023
* Efeito aleatório				
Entre escolas: v_{1k}	1.310	0.195	1.241	0.191
Inter-alunos: u_{ijk}	0.000	0.000	0.000	0.000
Intra-alunos: e_{0ijk}	1.000	0.000	1.000	0.000
Variância Inclinação-6ª série			0.940	0.342
Covariância Intercepto-6ª série			-0.548	0.201
Coef.corr: entre escolas	28.48	-	27.39	-
Coef.corr: inter-alunos	-	-	-	-
N=8.980				

Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005?) e Censo Escolar de 1999.

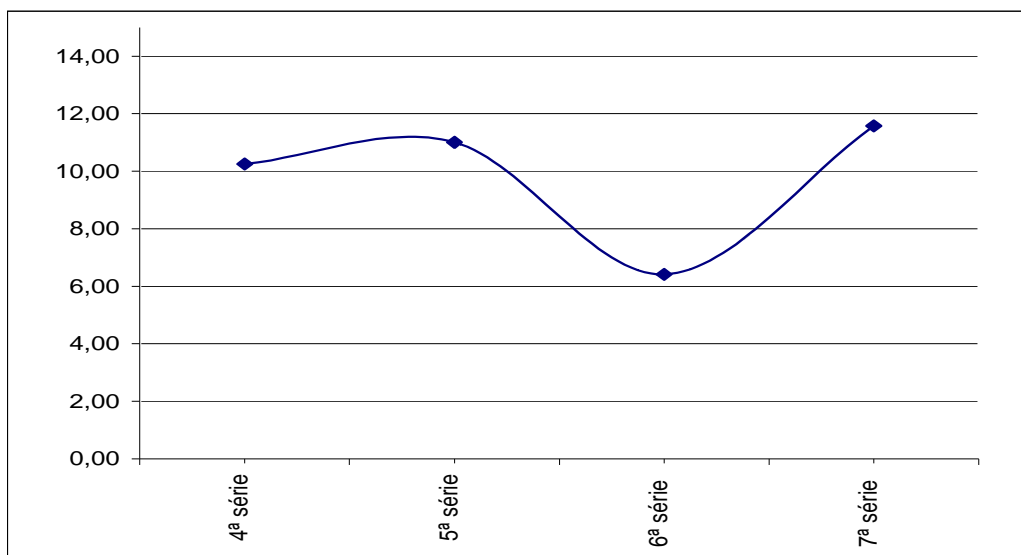
No modelo 5 evidencia-se que a probabilidade de ocorrência da evasão é menor nas séries iniciais, apesar do único coeficiente significativo ser o associado à 6ª série. Ou seja, apesar de não se poder associar plenamente o tempo à série cursada, nota-se uma determinada equivalência entre ambas as variáveis ao constatar-se uma maior probabilidade do evento ocorrer nas séries mais avançadas.

Um dos objetivos do último modelo é verificar se a proficiência anterior do aluno tem impacto significativo sobre a sua decisão quanto à permanência na escola. De acordo com o sinal negativo do coeficiente associado à variável, verifica-se que quanto maior essa proficiência, menor a probabilidade do aluno evadir do sistema escolar. Em geral, alunos com proficiência mais elevada compartilham um ambiente externo (econômico, cultural) que favorece a aprendizagem. Noutras palavras, esse resultado pode estar associado ao nível socioeconômico familiar.

Esse modelo final foi especificado de forma a verificar se o impacto da série cursada sobre a evasão varia de acordo com as escolas. Testou-se esse impacto no coeficiente da 6ª série, por ser a única a apresentar significância estatística. A parte de efeitos aleatórios mostrada nas últimas linhas da TAB. 14 indica que o efeito dessa variável sobre a ocorrência do evento não é o mesmo entre as escolas. Como no caso da repetência, o sinal negativo para a covariância entre o intercepto e a 6ª série indica que nas escolas com uma menor proporção média de evadidos é mais nítida a relação entre a ocorrência do evento e a série.

O comportamento da evasão ao longo das séries pode ser melhor visualizado no gráfico seguinte, elaborado com base nas probabilidades estimadas.

GRÁFICO 29
Probabilidades estimadas de evasão entre a 4ª e a 8ª série
segundo a série cursada: alunos da Ficha B, 1999-2003.



Fonte: Tabela 14.

Pode ser constatado que o formato da curva segue a mesma tendência não linear revelada pelo modelo com inclusão da variável Tempo. Nota-se que a probabilidade de

evasão aumenta ligeiramente entre a 4ª e a 5ª série, decresce substancialmente na 6ª, sendo bem mais elevada na 7ª série. Esse resultado pode estar sugerindo uma associação mais forte entre evasão e mercado de trabalho, pois nas séries finais do ensino fundamental o aluno tem uma idade relativamente mais avançada, fato que favorece sua inserção nesse mercado.

Na TAB. 15 são apresentadas as probabilidades relacionadas aos coeficientes significativos do último modelo estimado. De forma similar ao caso anterior, ao se analisar o coeficiente de uma variável, iguala-se as demais variáveis à zero.

TABELA 15
Coeficientes significativos do modelo 6 e respectivas probabilidades de evasão entre a 4ª e a 8ª série do ensino fundamental: alunos da Ficha B, 1999-2003.

Modelo 6			
	Coefficiente	Probabilidade	Prob – Méd.
Intercepto	-2.033	11.58	
4ª série	-0.137	10.25	-1.33
5ª série	-0.058	11.01	-0.57
6ª série	-0.642*	6.42	-5.16
7ª série	0.000	-	-
Proficiência aluno	-0.024*	11.30	-0.27
Trabalho	0.191**	13.71	2.13
Cor	-	-	-
Sexo	-0.197**	9.71	-1.87
Indicador NSE_1	-	-	-
Indicador NSE_2	-0.083**	10.72	-0.86
Repetência antes 4ª	-	-	-
Ind.Segurança escola1	-	-	-
Ind.Segurança escola2	-0.347*	8.47	-3.11
Ind. Estrutura escola1	-	-	-
Ind. Estrutura escola2	-	-	-
Matriculas 4ª série	-	-	-
Matric. 4ª promovidos	-	-	-
Professor niv. superior	-	-	-
Qtde. salas	-	-	-
Norte	-1.412*	3.08	-8.50
Centro-Oeste	-1.876*	1.96	-9.61
Nordeste	0.000	-	-
Profic. média escola	-	-	-

Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005?) e Censo Escolar de 1999.

O valor obtido para o intercepto (11,58) corresponde à proporção mediana de evasão numa escola típica, entre a 4ª e a 8ª série (para um aluno não-branco, do sexo masculino, que não repetiu nenhuma série antes da 4ª e nem trabalhou no período sob análise, que cursou a 7ª série e matriculado numa escola da região Nordeste; cujos valores para as variáveis contínuas são igualados a zero).

Como feito no caso da repetência, a análise inicial foca as covariáveis de nível 1 e 2 (variáveis ao longo do tempo e fixas no tempo, respectivamente) para, em seguida, contemplar as covariáveis de nível 3 (relacionadas às escolas).

No grupo das variáveis que variam ao longo do tempo, pode ser notado que o Trabalho foi o único fator a aumentar a probabilidade de ocorrência da evasão escolar numa escola típica: se o aluno trabalhou em algum ano letivo essa probabilidade equivaliu a 13,71%.

Uma das pretensões desse modelo foi avaliar o papel da proficiência anterior do aluno sobre a probabilidade de evasão. A idéia é verificar até que ponto um melhor desempenho inibe a ocorrência do evento. Estimou-se uma probabilidade de evasão entre a 4ª e a 8ª série igual a 11,30% para os alunos que têm um aumento de um ponto acima da média nos escores referentes à proficiência observada no ano letivo anterior à ocorrência do evento.

Pode ser observado que a única série com coeficiente significativo foi a 6ª. O sinal negativo indica que a probabilidade do aluno concluir essa série e evadir da escola foi bem menor em relação à 7ª série, equivalendo a 6,42%.

Com relação ao grupo das variáveis fixas dos alunos, verifica-se que somente o sexo e o nível socioeconômico familiar foram estatisticamente significativos no modelo. A probabilidade estimada de evasão para as meninas foi inferior à observada para os meninos, sendo da ordem de 9,71%. Constatou-se uma forte associação entre o *background* familiar e a ocorrência do evento. Para um aumento no escore referente ao nível socioeconômico familiar de um ponto acima da média, a probabilidade estimada do aluno evadir foi inferior à média, sendo igual a 10,72%. A variável associada à trajetória escolar passada (repetência antes da 4ª série) não teve efeito significativo sobre o acontecimento do evento.

Entre os fatores de nível 3, mais especificamente os diretamente relacionados à escola, nota-se que apenas o indicador do sistema de segurança da escola (Ind.Segurança escola2) teve impacto sobre a evasão. Diante de um aumento de um ponto acima da média nos valores dos escores referentes ao sistema de segurança, a probabilidade estimada de um aluno sair do sistema escolar foi equivalente a 8,47%. A proficiência média dos colegas não teve efeito sobre a ocorrência do evento.

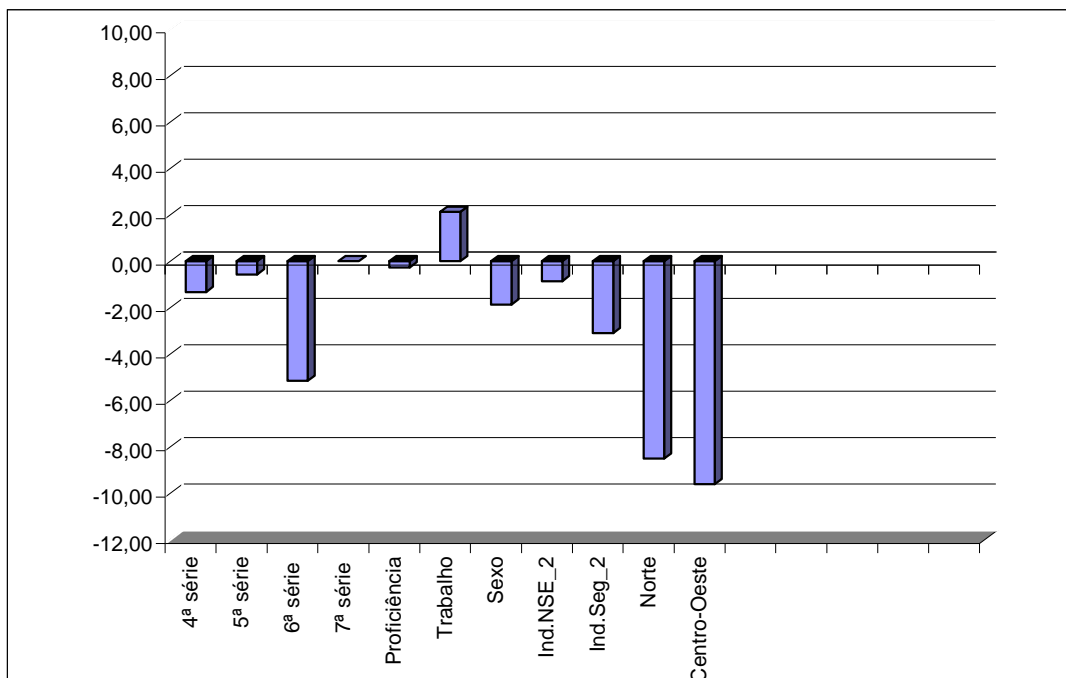
Os resultados mostram também a importância dos fatores regionais sobre a decisão de saída do aluno do sistema escolar: a probabilidade estimada do aluno evadir da escola foi igual a 3,08% se ele estava matriculado numa escola situada na região Norte e igual a

1,96% se ele estava matriculado numa escola da região Centro-Oeste, valores muito inferiores ao estimado para os alunos pertencentes às escolas do Nordeste.

Essa seção é finalizada com a apresentação do GRÁF. 30, que possibilita visualizar melhor o efeito dos fatores de aluno e da escola sobre a probabilidade estimada de evasão.

GRÁFICO 30

Efeito das variáveis de aluno e escola sobre a probabilidade média de evasão entre a 4ª e a 8ª série do ensino fundamental: Ficha B, 1999-2003.



Fonte: Elaboração própria com base na TAB. 15.

As considerações anteriores são nitidamente mostradas no gráfico. A probabilidade estimada de evasão foi menor do que a proporção mediana numa escola típica nas séries iniciais (entre a 4ª e a 8ª). Ela foi menor também nas seguintes situações: i) se o aluno aumentou a sua proficiência no ano letivo anterior; ii) se o aluno era do sexo feminino; iii) diante de uma melhoria no nível socioeconômico familiar; iv) caso ocorra uma melhoria no sistema de segurança da escola (aspecto relacionado a uma boa escola) e; v) caso o aluno frequente escolas localizadas nas regiões Norte e Centro-Oeste.

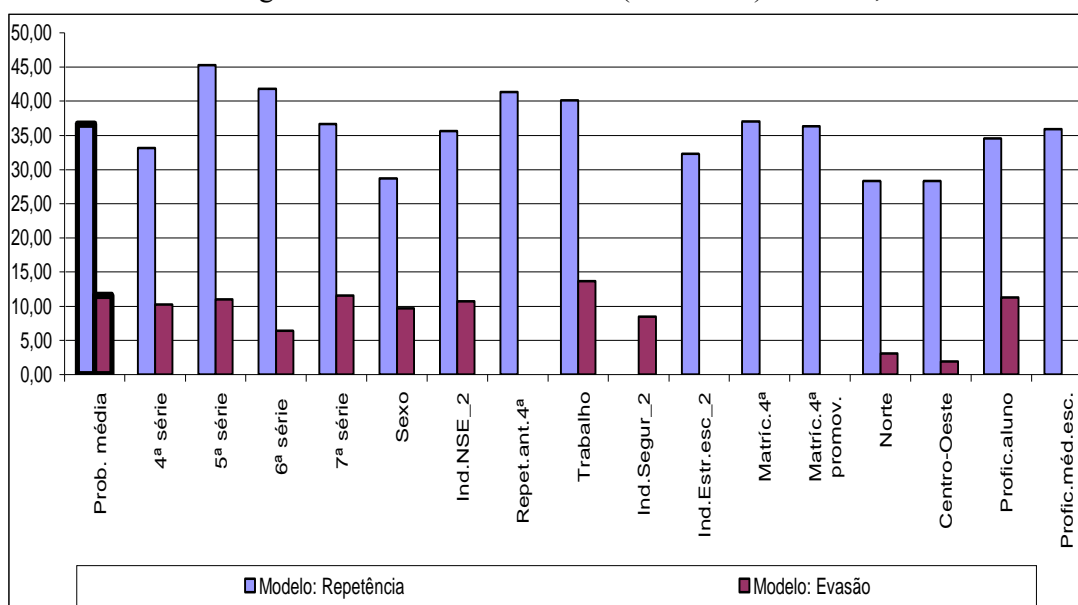
Contrariamente, a probabilidade estimada de ocorrência do evento foi bem superior à proporção mediana se o aluno trabalhou durante algum ano letivo, no período de análise.

7.3 Repetência e evasão no ensino fundamental: um paralelo entre os resultados obtidos e algumas considerações sobre a importância do estudo longitudinal para os eventos estudados

A discussão inicial aqui realizada refere-se aos modelos finais estimados para a repetência e a evasão, simultaneamente. A pretensão é identificar diferenciais e similaridades quanto aos principais fatores associados à ocorrência de cada evento. Essa discussão é baseada no GRÁF. 31.

GRÁFICO 31

Probabilidades estimadas da primeira repetência e da evasão entre a 4ª e 8ª série do ensino fundamental segundo as variáveis relevantes (modelos 6): Ficha B, 1999-2003.



Fonte: Elaboração própria com base nas TAB. 11 e 15.

A visualização gráfica revela que a proporção mediana de repetência - primeira barra do gráfico, mensurada pelo intercepto – foi muito superior à proporção mediana de evasão numa escola típica. Esse resultado está condizente com as estatísticas inicialmente apresentadas nessa tese, que mostram a supremacia das taxas de repetência em relação às taxas de evasão, bem como o maior percentual de alunos repetentes em relação ao de evadidos. Quanto ao tempo médio de ocorrência da repetência e da evasão, observa-se claramente que ele foi bastante diferenciado: para a repetência, as barras apresentam o formato curvilíneo, com o pico na 5ª série. Para a evasão, nota-se oscilação entre as séries, sendo a probabilidade estimada de ocorrência maior na 7ª série.

Entre as variáveis relacionadas ao aluno, pode ser notado que o sexo, o *background* familiar e as variáveis relacionadas à trajetória escolar passada e contemporânea (repetência antes da 4ª série e trabalho em algum ano letivo) foram importantes preditoras da ocorrência da repetência. No caso da evasão, os principais determinantes foram o sexo, o *background* familiar e, sobretudo, o fato do aluno ter trabalhado em algum ano letivo entre 1999 e 2002.

As variáveis diretamente relacionadas à escola tiveram maior impacto sobre a repetência. De uma forma geral, observou-se que a existência de uma melhor estrutura física e de um bom sistema de segurança esteve inversamente relacionada à probabilidade de ocorrência dos eventos, indicando que alunos matriculados em escolas com mais recursos físicos têm maior propensão à aprovação (lembrando que a não-aprovação é a principal causa da repetência) e à permanência na escola.

Os fatores socioeconômicos regionais apresentaram forte efeito sobre a probabilidade de ocorrência de ambos os eventos. Alunos matriculados em escolas situadas na região Nordeste foram mais propensos a repetirem uma série, ou evadirem do sistema escolar entre a 4ª e a 8ª série em relação aos alunos das escolas das regiões Norte e Centro-Oeste.

Por último, constata-se que a proficiência anterior do aluno, apesar de ter sido significativa nos modelos estimados para os dois eventos, teve um papel mais determinante no caso da repetência. Além disso, a proficiência média da escola só foi significativa na explicação da ocorrência desse evento: o desempenho dos colegas impactou positivamente o resultado escolar de um aluno específico, reduzindo sua probabilidade de repetir uma determinada série.

Para concluir esse capítulo, é importante frisar a importância de um estudo longitudinal baseado em técnicas específicas para a análise dos dados educacionais. Parte dos comentários é feita com base nos gráficos seguintes:

GRÁFICO 32

Percentual de casos de repetência entre a 4ª e 8ª série: Ficha B, 1999-2003.

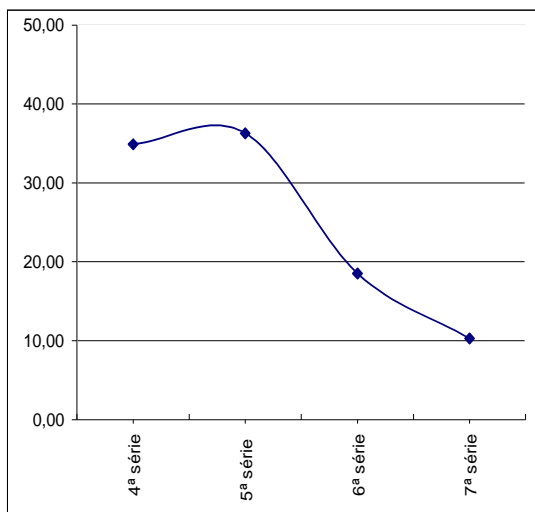
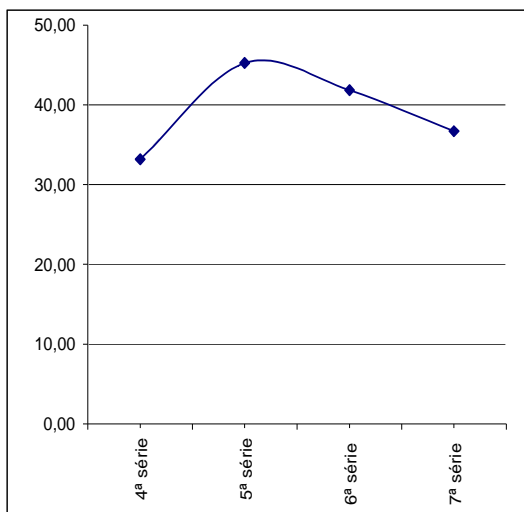


GRÁFICO 33

Probabilidades de 1ª repetência entre a 4ª e 8ª série: Ficha B, 1999-2003.



Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005). Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005).

Como na Ficha B consta o resultado escolar do aluno ao fim de cada ano letivo, desde sua matrícula na 4ª série em 1999 até 2003, é possível calcular o total de casos de repetência e de evasão ocorridos na coorte, entre a 4ª e a 8ª série do ensino fundamental.

O GRÁF. 32 retrata o percentual de casos de primeira repetência registrados para a coorte, segundo a série cursada. De acordo com esse gráfico, quando se considera a totalidade de registros ocorridos entre 1999 e 2003 e se verifica em quais séries eles foram predominantes, conclui-se que o evento ocorreu de forma mais significativa nas séries iniciais (4ª e 5ª).

No GRÁF. 33 são mostradas as probabilidades estimadas de ocorrência do evento ao longo do período. A visualização gráfica mostra que, ao se considerar a trajetória temporal do evento controlando-se para os fatores intra e inter-alunos, essas probabilidades foram maiores nas séries finais.

Um paralelo entre os dois gráficos revela curvas bastante distintas, apesar de se constatar que o evento ocorreu de forma mais acentuada na 5ª série, sendo esse aspecto mais evidente no segundo gráfico. Com base no GRÁF. 32 é possível verificar que o ponto máximo relacionado à ocorrência da primeira repetência situa-se entre a 4ª e a 5ª série, enquanto no GRÁF. 33 esse ponto situa-se entre a 5ª e a 6ª série.

Os gráficos seguintes, relacionados à evasão escolar, contemplam os mesmos indicadores.

GRÁFICO 34

Percentual de casos de evasão entre a 4ª e 8ª série: Ficha B, 1999-2003.

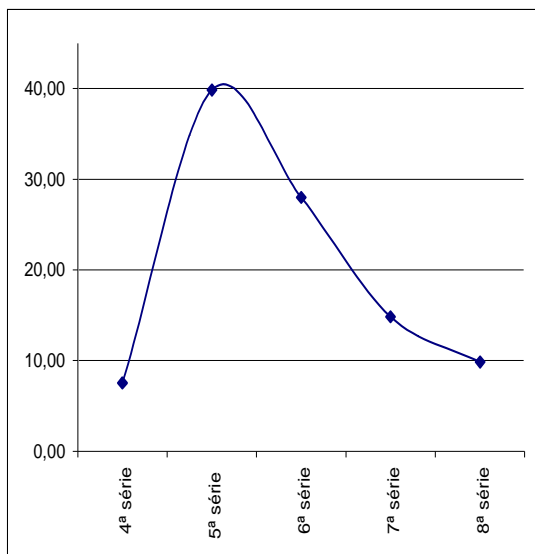
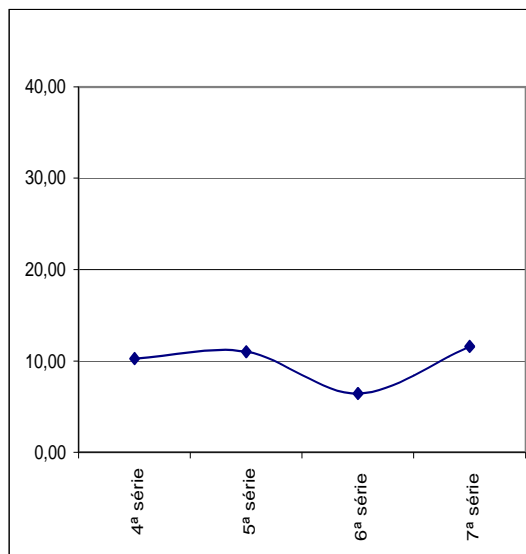


GRÁFICO 35

Probabilidades de evasão entre a 4ª e a 8ª série: Ficha B, 1999-2003.



Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005). Fonte: Elaboração própria. Dados básicos: CEDEPLAR (2005)

O GRÁF. 34 mostra o percentual de registros de evasão observados para a coorte, segundo a série cursada. Quando se considera todos os casos ocorridos entre 1999 e 2003, verifica-se que as séries com maior predomínio da evasão escolar foram as intermediárias (5ª e 6ª), sendo que o evento ocorreu de forma menos expressiva na 4ª série e nas séries finais do ensino fundamental.

No GRÁF. 35 são apresentadas as probabilidades estimadas de ocorrência do evento ao longo das séries. Ao se considerar a trajetória temporal do evento, controlando-se para os fatores intra e inter-alunos, nota-se que a probabilidade de ocorrência é mais elevada na 7ª série do ensino fundamental.

As curvas apresentadas são bastante divergentes: no primeiro caso, o comportamento do evento pode ser descrito por uma função parabólica, enquanto no segundo, por uma função polinomial de grau 3. De forma análoga ao caso da repetência, fica evidenciada a necessidade de se utilizar técnicas específicas de análise longitudinal para os dados educacionais, de forma a se obter resultados mais condizentes com a realidade do fenômeno sob estudo.

Além dessas considerações, outro importante ponto deve ser mencionado: na análise descritiva foi visto que a maioria dos casos de evasão na coorte aconteceu após a aprovação do aluno numa determinada série. Esse fato induz a importantes questionamentos, pois contradiz os estudos existentes que mostram que o evento ocorre

principalmente após a reprovação numa determinada série. Contudo, ao se utilizar um método específico de análise longitudinal (modelo longitudinal), observou-se uma correlação forte e negativa entre a aprovação e a ocorrência do evento e que, à medida que passa o tempo e o aluno é aprovado, menor a sua probabilidade de evadir do sistema escolar.

Evidencia-se, portanto, que os eventos escolares são melhor compreendidos quando analisados através de métodos adequados aos dados longitudinais, nesse caso, modelos de regressão hierárquicos longitudinais.

CONSIDERAÇÕES FINAIS

Uma característica dos estudos longitudinais é a perda de acompanhamento dos indivíduos sob estudo. Essa tese manteve essa peculiaridade. Foi verificada uma redução bastante expressiva da coorte, entre 1999 e 2003. Apenas 42,5% dos alunos permaneceram nas escolas da amostra, durante todo o período. Essa redução foi atribuída, basicamente, ao grande número de transferências (para outras escolas ou para a EJA/Supletivo Seriado) e casos de evasão. A maioria das transferências ocorreu na 4ª e na 5ª série do ensino fundamental, enquanto a evasão foi predominante no ano de 2001.

Os resultados descritivos relacionados à trajetória da coorte são comentados na seqüência. Esses resultados são contrastados com algumas estatísticas apresentadas no primeiro capítulo da tese. A idéia é identificar possíveis similaridades nos eventos básicos ocorridos entre os alunos da coorte e os indicadores regionais e nacionais apresentados.

Os dados evidenciaram uma retenção significativa de alunos em séries anteriores à série que deveria ser cursada no ano letivo correspondente. Considerando-se os alunos presentes durante todo o período (4.488), somente 69,4% concluíram com aprovação a 8ª série. A maior causa dessa retenção foi a reprovação, sendo esse evento mais significativo na 5ª série, nas escolas de todas as UF's. Esse último resultado está condizente com as estatísticas apresentadas no primeiro capítulo, que mostraram que as maiores taxas de reprovação, no país, foram observadas naquela série.

Durante o período de acompanhamento, quase 30% dos alunos matriculados nas diversas escolas analisadas tiveram pelo menos uma reprovação. A repetência por reprovação foi registrada para cerca de 18% dos alunos. Evidenciou-se que um percentual significativo de alunos reprovados não repetiu a série. Esse fato é atribuído à transferência e à evasão. O afastamento por abandono foi registrado para aproximadamente 3,5% dos alunos, ocorrendo principalmente na 5ª série. No cap. I, a análise segundo as regiões brasileiras revelou que a série com maior ocorrência do evento foi a 5ª para as regiões Norte e Nordeste e a 8ª para as regiões Centro-Oeste, Sul e Sudeste. Nesse estudo, não foi possível analisar a ocorrência do evento nessa última série, pois o acompanhamento da coorte foi interrompido em 2003. No tratamento das informações (Cap. 5) se o aluno cursou a 7ª série e não retornou à escola em 2003, o caso foi considerado evasão. Esse tratamento pode ter conduzido a uma subestimação nos registros de afastamento por abandono ou a uma sobrestimação nos registros de evasão para a 8ª série. Finalmente, com relação à saída do aluno do sistema escolar, foi elevado o percentual de evadidos na coorte.

Diferentemente do que foi verificado para o país, em que as taxas de evasão aumentaram gradativamente ao longo das séries, na análise descritiva foi visto que os percentuais de casos registrados nas regiões analisadas oscilaram bastante entre a 4ª e a 8ª série.

Embora a análise descritiva dos eventos tenha se baseado em percentuais e não em taxas, como tratado no primeiro capítulo, foi possível verificar que o comportamento dos eventos, na coorte, seguiu praticamente o mesmo padrão observado para o país. Além disso, constatou-se que a região Nordeste, representada pelos estados de Pernambuco e Sergipe, se sobressaiu por apresentar as piores estatísticas educacionais (percentual de casos de reprovação, repetência, abandono e evasão) entre a 4ª e a 8ª série do ensino fundamental.

No Brasil, os dados sobre reprovação, repetência, abandono e evasão têm sido estudados, basicamente, através da construção e análise de indicadores específicos. Apesar de ser possível extrair informações relevantes com base nesses indicadores, considerar apenas o dado num determinado ponto ou mesmo em dois pontos no tempo não permite captar certas peculiaridades do evento em questão. Essa tese tem a vantagem de ser desenvolvida a partir de dados longitudinais, que possibilitam o uso de metodologias específicas para o estudo de trajetória de indivíduos. Portanto, diante do acompanhamento da coorte de alunos matriculados na 4ª série do ensino fundamental nas escolas da amostra, nas seis UF's brasileiras, no período de 1999 a 2003, novas facetas dos eventos sob estudo podem ser descobertas. Os resultados de uma das técnicas utilizadas são discutidos a seguir.

Com a análise de sobrevivência - que contemplou os eventos primeira reprovação, primeira repetência, primeiro afastamento por abandono e evasão - pretendeu-se verificar a plausibilidade de três hipóteses: i) a igualdade das funções de sobrevivência estimadas para os alunos das escolas das seis UF's; ii) a observância de um maior tempo de sobrevivência à cada evento de interesse para os alunos matriculados em escolas situadas em UF's mais desenvolvidas e; iii) a observância de uma probabilidade de sobrevivência diferenciada para todos os eventos de interesse, considerando-se dois grupos na coorte: um, composto por alunos matriculados na 4ª série, em 1999, com idade defasada em relação a essa série e o outro composto por alunos matriculados na idade ideal, naquela série e naquele ano.

Com relação à primeira hipótese, constataram-se diferenças estatisticamente significativas entre as funções estimadas para a totalidade dos eventos, para os alunos matriculados nas escolas das seis UF's (teste de Log-rank). Infelizmente, o método analítico não permite fazer inferências sobre os fatores responsáveis por essas diferenças.

Extrapolando a questão da hipótese levantada, um comentário será feito sobre os eventos reprovação e repetência. Embora o comportamento das curvas para ambos os eventos, por UF, tenha sido ligeiramente parecido, a priori esperar-se-iam probabilidades de sobrevivência também parecidas, pois em geral a reprovação (principal fator que conduz à repetência) numa série num determinado ano letivo conduz à repetência na série no ano letivo seguinte. No entanto, esse aspecto não foi verificado, pois parcela significativa dos alunos reprovados não repetiu a série em função da transferência ou da evasão.

A segunda hipótese é confirmada: os alunos matriculados nas escolas das UF's menos desenvolvidas apresentaram um menor tempo de sobrevivência aos eventos. O desenvolvimento regional foi medido através do Índice de Desenvolvimento Humano (IDH). A confirmação da conjectura levantada foi feita com base na tabela abaixo.

Tabela 16
IDH por UF's selecionadas e posição em relação às demais UF's brasileiras, 2000.

UF	IDH	POSIÇÃO (em relação ao país)
MS	0,778	7°
GO	0,776	8°
RO	0,735	14°
PA	0,723	15°
PE	0,705	19°
SE	0,682	23°

Fonte: Elaboração própria com base no Atlas do Desenvolvimento Humano no Brasil, 2000.

No estudo, os alunos pertencentes às escolas dos estados de Pernambuco e Sergipe, sobretudo deste último estado, tiveram as menores probabilidades de sobrevivência aos eventos, em relação aos alunos dos demais estados. Pela tabela verifica-se que embora as seis UF's em estudo sejam classificadas como tendo nível médio de desenvolvimento (IDH entre 0.500 e 0.799), essas duas UF's tiveram os menores valores para o indicador, no ano de 2000. Entre todos os estados brasileiros, Pernambuco e Sergipe ocuparam a 19ª e a 23ª posição em termos de IDH, respectivamente. Uma vez comprovada a hipótese, um importante questionamento deve ser feito: quais os fatores responsáveis pelos resultados obtidos através da técnica de análise? Mais especificamente, o que explica menores tempos de sobrevivência para os alunos das regiões economicamente menos desenvolvidas? Embora os fatores regionais não sejam contemplados nesse estudo, em geral as regiões mais desenvolvidas são dotadas de escolas cuja qualidade da infra-estrutura e dos recursos humanos é superior à existente nas regiões menos desenvolvidas. É esta uma das principais

propostas dessa tese: verificar o papel dos recursos escolares sobre o resultado escolar do aluno.

Finalmente, sobre a última hipótese, foram observadas probabilidades de sobrevivência bastante diferenciadas para os alunos com e sem defasagem idade-série, para os eventos de interesse. A probabilidade de sobrevivência aos eventos, a partir da 4ª série, foi superior para os alunos matriculados na 4ª série, em 1999, na idade adequada à série. Esse resultado sugere, em primeiro lugar, a necessidade de se ingressar na escola na idade adequada. Aliás, no Brasil, esta questão está praticamente resolvida, em função da ampla cobertura atualmente existente no ensino fundamental. Em segundo lugar e de forma mais prioritária, está a questão da implementação de medidas favoráveis à manutenção do aluno na escola, na situação de aprovado. Isso é particularmente importante diante da constatação que o resultado escolar passado influencia o resultado escolar futuro.

A análise de sobrevivência revelou importantes aspectos relacionados aos eventos estudados. Mas importantes questionamentos não foram respondidos. O método não permite saber quais os fatores explicativos da menor ou maior probabilidade de sobrevivência do aluno a um evento específico. Noutras palavras, o método não permite identificar os fatores explicativos do resultado escolar. Assim, o estudo precisa ser complementado, mediante outra metodologia que permita responder as questões que permaneceram em aberto. É esta a intenção ao se estimar os modelos hierárquicos logísticos, tendo por variáveis-resposta os eventos representativos do fluxo escolar na coorte: a repetência e a evasão.

Interessante lembrar que os modelos estimados tiveram por base a especificação original da FPE, que foi adaptada de forma a inserir na equação geral as especificidades dos dados longitudinais utilizados. Essa equação foi constituída por variáveis associadas ao tempo, alunos e respectivas famílias (níveis 1 e 2) e variáveis associadas às escolas (nível 3).

Com a estimação dos modelos pretendeu-se identificar, entre as diversas variáveis, aquelas mais fortemente relacionadas à ocorrência de cada evento. Mais especificamente, pretendeu-se responder aos questionamentos feitos na parte inicial dessa tese. A discussão seguinte é baseada nas respostas à tais questões.

Foram levantadas seis questões. As duas primeiras referem-se ao efeito da proficiência anterior do aluno e da proficiência média da escola sobre a ocorrência de ambos os eventos. A proficiência do aluno mensurada a cada ano letivo, embora tenha sido significativa para a redução dos dois eventos, contribuiu mais efetivamente para reduzir a

probabilidade estimada de repetência. Esse resultado condiz com o que se espera, pois se num determinado ano letivo o resultado obtido pelo aluno nos testes de português e matemática foi favorável, a tendência é que ele seja aprovado nesse ano em todas as disciplinas ministradas. Quanto à proficiência média da escola, seu efeito só foi significativo no caso da repetência. Foi visto que o resultado escolar de um aluno específico é influenciado pelo desempenho dos seus colegas de escola. Assim, alunos matriculados em escolas cujos colegas tinham uma proficiência mais elevada apresentaram uma maior probabilidade de serem aprovados às sucessivas séries. No caso da evasão, ainda que uma maior proficiência do aluno reduza a sua probabilidade de evadir do sistema escolar, seu efeito sobre a decisão de continuidade na escola é pequeno, sendo que o baixo nível socioeconômico familiar tem um maior peso sobre essa decisão. Essa constatação é reforçada pelo fato da proficiência dos colegas da escola não ter tido nenhum efeito sobre a ocorrência do evento.

A terceira questão está associada à influência (positiva ou negativa) das variáveis associadas à trajetória escolar passada e contemporânea do aluno sobre a probabilidade de ocorrência dos eventos. As variáveis consideradas foram a repetência antes da 4ª série e a situação de trabalho do aluno (se ele trabalhou em algum ano letivo entre 1999 e 2003). Essa investigação explicitou, no primeiro caso, que a trajetória escolar passada do aluno não está dissociada da sua trajetória escolar corrente. Se o aluno repetiu algum ano letivo no início do ensino fundamental, é grande a sua probabilidade de repetir entre a 4ª e a 8ª série. Talvez possa ser inferido que a dificuldade de aprendizagem no passado continua repercutindo na vida escolar corrente, com as deficiências não sanadas constituindo barreiras ao aprendizado. Um ponto positivo revelado pelo estudo é que o forte impacto negativo da repetência nas séries iniciais sobre a trajetória escolar contemporânea do aluno pode ser parcialmente diluído pelo efeito do desempenho dos colegas da escola. Esse resultado é interessante, pois revela que a repetência no ensino fundamental pode ter uma redução caso o aluno repetente seja incluído numa turma cujos colegas tenham uma melhor proficiência. Na análise da evasão, era de se esperar que se o aluno tivesse repetido algum ano letivo antes da efetivação da sua matrícula na 4ª série, maior a sua probabilidade de evadir entre a 4ª e a 8ª série, uma vez que estudos (incluindo-se esse) têm mostrado o impacto negativo da repetência nas séries iniciais sobre o resultado escolar futuro. Contudo, essa expectativa não foi confirmada. No segundo caso, foi constatada a influência negativa do trabalho sobre o resultado escolar. Caso o aluno tenha trabalhado em algum ano letivo entre 1999 e 2002, maior a sua probabilidade de repetência e menor a

de permanência na escola, revelando a dificuldade em se conciliar trabalho e estudo. Importante lembrar que os alunos das escolas amostradas estudavam no turno diurno. Entretanto, é preciso ter cautela com relação à análise do efeito do trabalho sobre a repetência, pois não pode ser descartada a possibilidade do aluno estar com um desempenho ruim na escola, independentemente da sua situação de inserção no mercado laboral.

O quarto ponto levantado referiu-se à importância dos fatores de *background* familiar e escolares para a redução da probabilidade de repetência numa série. Além disso, questionou-se se o resultado obtido é válido também para a evasão. Com base nos modelos estimados para cada evento é possível afirmar, por um lado, que os fatores escolares tiveram maior impacto sobre a repetência do que sobre a evasão. O corpo discente é favorecido se está matriculado em melhores escolas, apresentando uma maior probabilidade de aprovação às sucessivas séries. É razoável supor que as escolas da amostra que se caracterizaram por ter um maior número de matriculados promovidos da 3ª para a 4ª série em 1999 e uma maior proficiência média do seu corpo discente podem ser consideradas boas escolas. Esse conceito é reforçado pela presença de uma melhor infraestrutura. Essas variáveis foram negativamente correlacionadas com a probabilidade de repetência. Vale lembrar que o principal componente do indicador de estrutura foi a presença de laboratório de informática na escola. Pode ser argumentado que o uso de computadores nas aulas favorece a aprendizagem, pois o aluno é motivado a realizar pesquisas de cunho acadêmico com mais qualidade, sendo criados e aperfeiçoados os hábitos de leitura, além de serem desenvolvidas outras habilidades e aumentada a sua reflexão crítica. Portanto, infere-se que a qualidade da escola contribui de forma significativa para reduzir a probabilidade de repetência para os alunos nela matriculados. Por outro lado o *background* familiar, mensurado pelo nível socioeconômico, teve um peso importante na determinação de ambos os eventos. Especificamente com relação à evasão, as regressões mostraram que alunos cujas famílias não tinham nem os bens duráveis básicos na sua residência estavam mais sujeitos a sair do sistema escolar. Os resultados obtidos permitem associar esta saída à participação em atividades laborativas. Ou seja, a situação socioeconômica da família está fortemente relacionada à evasão escolar. Assim, enquanto os fatores escolares tiveram um efeito mais significativo sobre a repetência, na análise da evasão os dados mostraram que ainda que o aluno esteja matriculado numa boa escola, sua probabilidade de evadir do sistema escolar é mais fortemente afetada por sua situação socioeconômica. Esse resultado observado para a coorte confirma conclusão

presente no Relatório da UNESCO (2008) sobre a evasão, para as escolas localizadas nos países em desenvolvimento. Segundo o Relatório, mesmo as escolas bem estruturadas desses países não conseguem reter o aluno caso ele esteja submetido a uma situação de pobreza.

O quinto questionamento é aqui transcrito: no ano de 2003, entre a 4ª e a 8ª série, a 5ª série se sobressaiu por apresentar as maiores taxas de repetência do país, enquanto as taxas de evasão aumentaram gradualmente ao longo das séries (MEC/INEP 2003). Para a coorte em estudo, a probabilidade de repetência é mais acentuada também na 5ª série? E a evasão, segue o mesmo comportamento observado para o país? Por um lado, os resultados obtidos revelaram que a probabilidade de um aluno da coorte repetir o ano letivo é maior na 5ª série, o que condiz com o comportamento das taxas de repetência observadas em 2003 mostradas no Cap.1. Por outro lado, embora as probabilidades estimadas de evasão por série, para a coorte, não tenham seguido a mesma trajetória temporal crescente vista naquele capítulo, elas foram mais elevadas nas últimas séries do ensino fundamental.

O último questionamento levantado referiu-se ao efeito do contexto socioeconômico regional sobre a ocorrência da repetência e da evasão no ensino fundamental. Verificou-se que o ambiente socioeconômico da região em que a escola está inserida tem forte impacto sobre a ocorrência de ambos os eventos. A probabilidade de repetência e de evasão é bem maior para os alunos matriculados em escolas da região Nordeste, em relação às demais regiões analisadas. Mais uma vez, esse aspecto evidenciado pelas regressões está atrelado ao nível socioeconômico familiar.

Esta tese teve o mérito de utilizar, pela primeira vez no país, a junção de duas ricas bases de dados longitudinais, a Avaliação de desempenho: fatores associados e a Ficha B. A primeira base já foi utilizada na elaboração de uma monografia, uma dissertação e uma tese, no CEDEPLAR (LUZ, 2005; MACEDO, 2004 e PEREIRA, 2006). A segunda é utilizada pela primeira vez nessa tese. O uso conjunto de ambas as bases (complementadas com algumas informações do Censo Escolar de 1999) trouxe importantes elucidações sobre dois grandes problemas existentes no sistema de ensino brasileiro, mais especificamente, no ensino fundamental: a repetência e a evasão.

Dado o caráter longitudinal dos dados, foi possível a utilização de duas importantes metodologias de análise: a análise de sobrevivência e os modelos hierárquicos logísticos longitudinais. A primeira técnica, embora atualmente muito adotada no meio acadêmico brasileiro, ainda não tinha sido empregada no estudo de dados educacionais. A segunda técnica, apesar de ter o seu uso mais difundido nos últimos anos na área da educação,

ainda não fora utilizada para o estudo dos determinantes da repetência e da evasão escolar, devido à inexistência de dados longitudinais para estes eventos. É bom frisar que o último estudo longitudinal realizado no Brasil tendo como variável-resposta a repetência foi realizado em 1996, por Gomes-Neto e Hanushek. Ainda assim a análise referiu-se à década de 80, se restringindo às áreas rurais da região Nordeste.

De uma forma geral, é possível inferir que os dois métodos de análise aqui empregados se complementam e que seus resultados se mostraram compatíveis. Na análise de sobrevivência foi visto que a probabilidade de sobrevivência do aluno à primeira repetência entre a 4ª e a 8ª série foi muito inferior à probabilidade de sobrevivência à evasão. Em contrapartida, os modelos hierárquicos estimados mostraram que a probabilidade média de repetência foi bastante superior à probabilidade média de evasão para uma escola típica.

Os dois métodos revelaram também que o resultado escolar passado impacta de forma expressiva a trajetória escolar presente do aluno. Na análise de sobrevivência o resultado escolar passado foi mensurado pelo indicador de defasagem idade-série. Na maioria das vezes a defasagem idade-série acontece em função da reprovação. O método permitiu ver o grande diferencial existente nas curvas de sobrevivência aos eventos para os alunos que em 1999 tinham idade defasada em relação à 4ª série, as quais eram bem mais baixas em relação às curvas estimadas para os alunos com idade ideal àquela série. Analogamente, as regressões estimadas revelaram o forte efeito negativo da repetência anterior à 4ª série sobre o curso de vida escolar do aluno.

Por último, ambas as técnicas mostraram que os piores resultados ao longo da trajetória escolar da coorte foram observados para os alunos matriculados nas escolas da região Nordeste. Embora na análise de sobrevivência o recorte tenha sido feito por UF, em geral os alunos das escolas de Pernambuco e Sergipe se destacaram por apresentar a menor sobrevivência a cada evento. Ainda que a primeira técnica não permita identificar os fatores responsáveis por esse resultado, foi visto que os estados daquela região apresentaram os piores indicadores de IDH, em relação às demais UF's. Surge o questionamento: o que explica menores tempos de sobrevivência para os alunos das regiões economicamente menos desenvolvidas? É possível inferir que em geral as escolas das regiões mais desenvolvidas têm mais e melhores recursos físicos e humanos. A última técnica permitiu a identificação de alguns desses recursos e o efeito de cada um deles sobre a ocorrência dos eventos.

Acredita-se não ser pretensioso afirmar que esta tese contribuiu de forma efetiva para o melhor entendimento dos eventos em questão, mediante o uso de dados e técnicas longitudinais. A maioria dos resultados obtidos para a coorte corrobora os estudos e as estatísticas existentes em nível nacional. Outros, no entanto, mostraram-se aparentemente contraditórios, revelando importantes facetas dos eventos, anteriormente não reveladas. Para não ficar um texto muito repetitivo, serão feitos aqui apenas dois comentários sobre o evento evasão. Primeiro: enquanto a análise descritiva revelou que o evento, na coorte, aconteceu na maioria das vezes após a aprovação do aluno numa série, a análise de trajetória revelou que se o aluno é aprovado a cada série sucessiva, menor sua probabilidade de evadir da escola. Segundo: com base nos dados descritivos, poder-se-ia supor que a evasão é bem mais elevada nas séries iniciais, entre a 4^a e a 8^a. Contudo, no acompanhamento da coorte ao longo dos cinco anos percebeu-se que, ao se controlar pela aprovação, a tendência é que a probabilidade do aluno evadir do sistema escolar seja maior após a conclusão da 7^a série. Ainda que a literatura e as estatísticas existentes de certa forma apontaram alguns desses aspectos, essas contradições entre os resultados obtidos, vindas à tona nessa tese, confirmam a importância do uso de uma técnica de análise longitudinal para os dados educacionais.

Entre as analogias associadas aos resultados descritivos observados e aos resultados estimados, menção será feita novamente à série cursada. Porém, dessa vez, a variável-resposta é a repetência. Tanto na análise descritiva quanto na hierárquica foi confirmado que o grande gargalo relacionado à repetência no ensino fundamental centra-se na 5^a série. A análise hierárquica teve o mérito de revelar que esse aspecto é mais evidente nas escolas cuja probabilidade média de repetência é menor. Supõe-se ser possível associar menor ocorrência da repetência às melhores escolas (com mais e melhores recursos físicos e humanos). Os resultados mostram que mesmo nessas escolas a repetência é mais expressiva na 5^a série. Tal série marca a transição entre os métodos de organização curricular. Nesse sentido, percebe-se que o aluno é fortemente afetado pelas mudanças na grade curricular ocorridas entre a 4^a e a 5^a série. A sugestão é que haja uma revisão no projeto pedagógico de cada escola de forma a minimizar o efeito dessa mudança sobre o desempenho acadêmico.

Importante frisar algumas limitações deste estudo. As bases de dados utilizadas contemplam apenas os alunos matriculados nos turnos matutino e vespertino. Sabe-se que nas regiões menos desenvolvidas (como as contempladas nessa tese) é bastante significativo o percentual de alunos que trabalham durante o dia e estudam à noite. Esta

limitação impossibilita análises visando a determinação dos diferenciais entre os alunos, por turnos de estudo. Além disso, as informações existentes referem-se às séries 4^a a 8^a. Ou seja, parte importante do ciclo de vida escolar do aluno (no ensino fundamental) não é coberta. Outra limitação é que todas as escolas da amostra estão situadas nas áreas urbanas. Estudos como o de FLETCHER e RIBEIRO (1989) mostram que os indicadores e o desempenho educacional divergem bastante quando se considera a situação de domicílio urbano/rural. Esses aspectos listados induzem a uma maior cautela na generalização dos resultados obtidos.

Ainda que importantes lacunas sobre ambos os eventos tenham sido preenchidas nesta tese, é fato que muitas questões permanecem em aberto. As sugestões de aprofundamento desse estudo levam em consideração as bases de dados aqui utilizadas. Dada a variedade de informações nelas existentes, é interessante que sejam melhor exploradas em estudos futuros.

Especificamente no caso da evasão, foi visto que um percentual significativo da sua variabilidade (cerca de 39,15%) pode ser atribuído aos fatores escolares. Entretanto, as variáveis diretamente associadas à escola tiveram pouco poder explicativo sobre o acontecimento do evento. Foi constatado que o contexto regional teve impacto significativo sobre a probabilidade do aluno evadir do sistema escolar. Os alunos cujas famílias têm um menor poder aquisitivo, em geral, residem nas UF's menos desenvolvidas. E, pelo fato de serem menos desenvolvidas, não criam as condições necessárias para a melhoria do *status* socioeconômico familiar, de forma a assegurar a permanência do aluno na escola, e permanência com promoção às sucessivas séries. É a perpetuação do círculo vicioso da pobreza. Dada a importância do contexto regional sobre a evasão escolar, uma sugestão para investigações futuras que visem o aprofundamento desse estudo é a incorporação, nas equações hierárquicas, de um nível representado pelas regiões.

Nesta tese, após a ocorrência da primeira repetência, os dados foram censurados, ou seja, o acompanhamento do aluno foi interrompido. A análise descritiva revelou que, entre os alunos reprovados, cerca de 21% tiveram mais de uma reprovação entre a 4^a e a 8^a série. Esse estudo pode ser aprofundado mediante a incorporação desses casos na análise, que podem contribuir para trazer maiores esclarecimentos sobre os determinantes da repetência.

Outro motivo de interrupção do acompanhamento do aluno foram os registros de transferência para outra escola. Em muitos desses casos, consta na ficha do aluno a escola para a qual ele foi transferido e o resultado escolar nela obtido. Em geral, esta transferência

ocorreu entre as escolas da amostra. Certamente, interessantes resultados podem ser obtidos caso se dê sequência a esse acompanhamento. Os diferenciais existentes entre as escolas (de origem e destino) se tornarão mais evidentes, sendo possível avaliar com maior clareza o efeito dos recursos nelas existentes sobre o resultado escolar. Além disso, é provável que o resultado escolar dos alunos transferidos - desconsiderado nessa tese após a transferência - altere parte dos resultados obtidos nessa tese. Daí, a importância de um estudo que considere o resultado escolar do aluno também após sua transferência para outra escola da amostra.

A expectativa é que o acatamento de tais sugestões por parte dos pesquisadores da área educacional resulte em importantes ganhos aos raros estudos existentes que contemplam os principais eventos da trajetória escolar do aluno no ensino fundamental. Inclusive este.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBERNAZ, A.; FERREIRA, F. H. G.; FRANCO, C. Qualidade e equidade no ensino fundamental brasileiro. *Pesquisa e Planejamento Econômico*. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada, v. 32, n. 3, p. 453-476, dez. 2002.

ALLISON, P. D. *Survival analysis using the SAS system: a practical guide*. Cary,: SAS Institute, 1995.

ANDRADE, D. F. *Plano da amostra. Amostragem realizada para a pesquisa avaliação do desempenho: fatores associados*. Belo Horizonte: [s.n], 1999. Mimeografado.

BARBOSA, M. E. F.; FERNANDES, C. A escola brasileira faz diferença? Uma investigação dos efeitos da escola na proficiência em matemática dos alunos da 4ª série. In: FRANCO, C. (Org.). *Avaliação, ciclos e promoção na educação*. Porto Alegre: ArtMed, 2001. p. 155-172.

BLOSSFELD, H. P.; HAMERLE, A.; MAYER, K. U. *Event history analysis: statistical theory and application in the social sciences*. Hillsdale: L. Erlbaum Associates, 1989.

BRYK, A. S.; RAUDENBUSH, S. W. *Hierarchical linear models: applications and data analysis methods*. 2nd. ed. Newbury Park, California: Sage, 2002.

CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL *Avaliação de desempenho- fatores associados*. [Belo Horizonte]: [CEDEPLAR/UFMG], [2005?] Relatório entregue ao INEP. Mimeografado.

CERQUEIRA, C. A. Determinação de fatores ligados às taxas de distorção idade/série, taxa de evasão escolar e taxa de repetência. In: RIOS NETO, E. L. G.; RIANI, J. L. R (Org.). *Introdução à demografia da educação*. Campinas: ABEP, 2004.

CÉSAR, C. C.; SOARES, J. F. Desigualdades acadêmicas induzidas pelo contexto escolar. *Revista Brasileira de Estudos Populacionais*, Campinas, v.18, n.1/2, p.97-110, jan./dez. 2001.

COLEMAN, J. S. *et al. Equality of educational opportunity*. Washington: D.C.:U.S. Government Printing Office, 1966. apud HANUSHEK, E. A. The failure of input-based schooling policies. *The Economic Journal*, Malden, v. 113, n. 485, p.64-98, Feb. 2003.

COSTA RIBEIRO, S. A educação e a inserção do Brasil na modernidade. *Cadernos de Pesquisa*, São Paulo, n. 84, p. 63-82, fev. 1993.

FERRÃO, M. E.; BELTRÃO, K. I.; SANTOS, D. P. *Política de não-repetência e a qualidade da educação: evidências obtidas da modelagem dos dados da 4a. série do SAEB-99*. [2002?]. Disponível em: <<http://www.mat.ubi.pt/~meferrao/qualidade.pdf>>. Acesso em: 10 out. 2007.

- FERRÃO, M. E.; BELTRÃO, K. I; FERNANDES, C. *Aprendendo sobre escola eficaz: evidências do Saeb-99*. Brasília: INEP/MEC, 2002. Disponível em: <<http://www.mat.ubi.pt/~meferrao/INEP.pdf>>. Acesso em: 10 out. 2007.
- FLETCHER, P. *As dimensões transversal e longitudinal do Modelo Profluxo*. Brasília, DF.: Ministério da Educação e Cultura, 1997. Mimeografado.
- FLETCHER, P. R; RIBEIRO, S. C. *Modelling education system performance with demographic data: an introduction to the Profluxo Model*. Brasília: [s.n.], 1989.
- GOLDSTEIN, H. *Multilevel statistical models*. 2nd ed. London: Edward Arnold, 1995.
- GOMES NETO, J. B.; HANUSHEK, E. A. The causes and effects of grade repetition. In: BIRDSALL, N.; SABOT, R. H. (Ed.) *Opportunity forgone: education in Brazil*. Washington, DC: Inter-American development Bank, 1996. p. 425-460.
- GOUVEA, G. F. P. A educação básica no Brasil. *São Paulo Perspectiva*, São Paulo, v. 14, n. 1, jan./mar. 2000.
- HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. *Fundamentals of item response theory*. Newbury Park: Sage Publications, 1991.
- HANUSHEK, E. A. Conclusions and controversies about the effectiveness of school resources. *Economic Policy Review*, New York, v. 4, n. 1, p. 11-28, Mar.1998.
- HANUSHEK, E. A. *The evidence on class size..* University of Rochester ,W. Allen Wallis Instituto of Political Economy., 1998b. (Occasional Paper, 98-1).
- HANUSHEK, E. A. The failure of input-based schooling policies. *The Economic Journal*, Malden, v. 113, n. 485, p. 64-98, Feb. 2003.
- HANUSHEK, E. A. Publicly provided education. In: AUERBACH, A. J.; FELDSTEIN, M. (Ed.). *Handbook of public economics*. Amsterdam: Elsevier, 2002. v. 4, p. 2045-2141.
- HANUSHEK, E. A.; GOMES NETO, J. B.; HARBISON, R. W. Efficiency-enhancing investments in school quality. In: BIRDSALL, N.; SABOT, R. H. (Ed.). *Opportunity forgone: education in Brazil*. Washington, D. C.: Inter-American development Bank,1996. p. 385-424.
- HANUSHEK, E. A. *et al. Does peer ability affect student achievement?* Cambridge: National Bureau of Economic Research, 2001. (NBER Working Paper, 8502).
- HANUSHEK, E. A.; LAVY, V.; HITOMI, K. *Do students care about school quality? Determinants of dropout behavior in developing countries*. Cambridge: National Bureau of Economic Research, 2006. (NBER Working Paper, W12737).
- HANUSHEK, E. A.; LUQUE, J. A. *Efficiency and equity in schools around the world*. Cambridge: National Bureau of Economic Research, 2002. (NBER Working Papers, 8949)

HEYNEMAN, S. P.; LOXLEY, W. A. The effect of primary school quality on academic achievement across twenty-nine high- and low – income countries. *American Journal Sociology*, Chicago, v. 88, n. 6, p.1162-1194, May 1983.

HOUGH, G. C. *et al. Multistate life tables and event history analysis: new tables of working life*. 1994. 228 f. Tese (Doutorado) - Faculty of the Graduate School, Universidade de Texas, Austin, 1994.

HOX, J. *Multilevel analysis: techniques and applications*. London: Lawrence Erlbaum Associates, 2002.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo Demográfico de 2000: documentação dos microdados da amostra*. 2 ed. Rio de Janeiro, 2003. CD-ROM.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo demográfico de 1991: documentação dos microdados da amostra*. Rio de Janeiro, 1996.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento. *Pesquisa Nacional por Amostra de Domicílios 2005-2006*. Rio de Janeiro, 2006.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Pesquisa nacional por amostra de domicílios: síntese de indicadores 2006*. Rio de Janeiro, 2007.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS. *Geografia da educação brasileira*. Brasília, 2000.

KLEIN, R. *Produção e utilização de indicadores educacionais: metodologia de cálculo de indicadores do fluxo escolar da educação básica*. [S.l]: INEP, 1995. Mimeografado.

KRUEGER, A. B. Economic considerations and class size. *The Economic Journal*, Malden, v. 113, n. 485, p.34-63, Feb. 2003.

KRUEGER, A. B. Experimental estimates of education production functions. *The Quarterly Journal of Economics*, Cambridge, v. 114, n. 2, p. 497-532, May. 1999.

KRUEGER, A. B. Reassessing the view that american schools are broken. *Economic Policy Review*, New York, v. 4, n. 1, p. 29-46, Mar. 1998.

KRUEGER, A. B. A response to Eric Hanushek's "Evidence, politics and the class size debate". In: LAWRENCE, M.; ROTHSTEIN, R. (Ed). *The class size debate*. Washington: Economic Policy Institute, 2002.

LAZEAR, E. P. Educational production. *Quartely Journal of Economics*, Cambridge, v. 116, n. 3, p. 777-803, Aug. 2001.

LEE, V. E.; BRYK, A. S. A multilevel model of the social distribution of high school achievement. *Sociology of Education*, Washington, v. 62, n. 3, p. 172-192, July 1989.

LUZ, L. S. *Os determinantes do desempenho escolar: a estratificação educacional e o efeito valor adicionado*. 2005. 52 f. Monografia (Graduação em Economia) – Faculdade de Ciências Econômicas, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

MACEDO, G. A. *Fatores associados ao rendimento escolar de alunos da 5ª série (2000): uma abordagem do valor adicionado e da heterogeneidade*. 2004. 124 f. Dissertação (Mestrado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

MAGALHÃES, D. J. A. V. *Uma abordagem multinível sobre localização e mobilidade residenciais na Região Metropolitana de Belo Horizonte*. 2002. 265 f. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2002.

MARTELETO, L. J. *Desigualdade intergeracional de oportunidades educacionais: uma análise da matrícula e escolaridade das crianças brasileiras*. Belo Horizonte: UFMG/Cedeplar, 2004. (Texto para Discussão, 242).

MAYER, K. U.; TUMA, N. B. *Event history analysis in life course research*. Madison: The University of Wisconsin Press, 1990.

NERI, M. *et al.* The effects of idiosyncratic shocks to father's income on child labor, school drop-outs and repetition rates in Brazil. In: ENCONTRO BRASILEIRO DE ECONOMETRIA, 22., 2000, Campinas. *Anais...* Campinas: SBE, 2000.

OLIVEIRA, A. M. H. C. *Acumulando informações e estudando mudanças ao longo do tempo: análises longitudinais do mercado de trabalho brasileiro*. 2002. 151 f. Tese (Doutorado em Demografia) - Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2002.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA. *Marco estratégico para a Unesco no Brasil*. Brasília, 2006.

ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA. *Repetición escolar: una perspectiva global*. Ginebra, 1996.

PAL, S. *Child schooling in Peru: evidence from a sequential analysis of school progression*, 2003 (Labor and Demography, 0309001). Disponível em: <<http://129.3.20.41/eps/lab/papers/0309/0309001.pdf>>. Acesso em: 12 mar. 2008.

PEREIRA, D. R. M. *Fatores associados ao desempenho escolar nas disciplinas de matemática e de português no ensino fundamental: uma perspectiva longitudinal*. 2006. 104 f. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

PSACHAROPOULOS, G. Child labor versus educational attainment: some evidence from Latin America. *Journal of Population Economics*, Heidelberg, v. 10, n. 4, p.377-386, Oct. 1997.

RASBASH, J. *et al.* *A user's guide to MLwin: version 2.1d for use with MLwiN 1.10*.

University of London, Centre for Multilevel Modelling, 2000. Disponível em: <<http://www.cmm.bristol.ac.uk/team/userman.pdf>>. Acesso em: 11 JAN. 2008.

RAUDENBUSH, S.; BRYK, A. *Hierarchical linear models applications and data analysis methods*. New York: Sage Publications, 2002.

RIGOTTI, J. I. R.; CERQUEIRA, C. A. As bases de dados do INEP e os indicadores educacionais: conceitos e aplicações. In: RIOS NETO, E. L. G.; RIANI, J. L. R. (Org.). *Introdução à demografia da educação*. Campinas: ABEP, 2004.

RIANI, J. L. R. *Determinantes do resultado educacional no Brasil: família, perfil escolar dos municípios e dividendo demográfico numa abordagem hierárquica e espacial*. 2005. 218 f. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

RIANI, J. L. R.; GOLGHER, A. B. Indicadores educacionais confeccionados a partir de bases de dados do IBGE. In: RIOS NETO, E. L. G.; RIANI, J. L. R. (Org.). *Introdução à demografia da educação*. Campinas: ABEP, 2004.

RIOS NETO, E. L. G.; CÉSAR, C. C.; RIANI, J. L. R. Estratificação educacional e progressão escolar por série no Brasil. *Pesquisa e Planejamento Econômico*. Rio de Janeiro, v. 32, n. 3, p. 395-415, dez. 2002.

RIOS NETO, E. L. G. O método probabilidade de progressão por série. In: RIOS NETO, E. L. G.; RIANI, J. L. R. (Org.). *Introdução à demografia da educação*. Campinas: ABEP, 2004.

ROGERS, A. *Multiregional demography: principles, methods and extensions*. New York: Wiley, 1995.

SCHWARTZMAN, S. Acceso y retrasos en la educación en América Latina. Buenos Aires: SITEAL, [200?]. Debate I. Equidad en el acceso y la permanencia en el sistema educativo. Disponível em: <http://www.siteal.iipe-oei.org/modulos/DebatesV1/upload//deb_23/art_13/ART_SimonSchwartzman.pdf>. Acesso em: 07 ago. 2007.

SCHWARTZMAN, S.; BROCK, C. (Ed.). *Os desafios da educação no Brasil*. Rio de Janeiro: Nova Fronteira, 2005.

SINGER, J. D.; WILLETT, J. B. *Applied longitudinal data analysis: modeling change and event occurrence*. New York: OXFORD, 2003.

SNIJDERS, T.; BOSKER, R. *Multilevel analysis: an introduction to basic and advanced multilevel modelling*. New York: Sage, 1999.

SOARES, J. F.; CÉSAR, C. C.; MAMBRINI, J. Determinantes de desempenho dos alunos do ensino básico brasileiro: evidências do SAEB de 1997. In: FRANCO, C. (Org.) *Avaliação, ciclos e promoção na educação*. Porto Alegre: ARTMED, 2001.

SOARES, S.; LIMA, A. F. *A mensuração da educação nas PNAD's da década de 1990*. Rio de Janeiro: IPEA, 2002 (Texto para Discussão, 928).

SOUZA, D. T. *Teacher professional development and the argument of incompetence: the case of in-service elementary teacher education in São Paulo-Brazil*. 2001. 297 f. Tese (PHD) - Institute of Education, University of London, 2001.

TOOD, P. E.; WOLPIN, K. I. On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, Malden, v. 113, n. 485, p. 3-33, Feb. 2003.

TORRES, R. M. *Repetición escolar: ¿falla del alumno o falla del sistema?* 1995. (Evaluación, Aportes para la Capacitación, 1). Disponível em: <<http://www.fronesis.org/documentos/torres1995repeticion.pdf>>. Acesso em: 20 ago.2007.

TRUSSELL J.; HANKINSON, R.; TILTON, J. *Demography applications of event history analysis*. New York: Oxford University, 1992.

UNESCO. Organização das Nações Unidas para a Educação, a Ciência e a Cultura *Alfabetização para a vida: relatório de monitoramento global de educação para todos – EPT 2006*. Paris: UNESCO, 2005.

UNITED NATIONS EDUCATIONAL. Scientific and Cultural Organization. *Education for all: status and trends*. Paris, 1993.

UNITED NATIONS EDUCATIONAL. Scientific and Cultural Organization. Education for all by 2015: will we make it? In __. *EFA Global Monitoring Report 2008*. [2008?]. Disponível em: <<http://appablog.wordpress.com/2007/11/30/efa-global-monitoring-report-2008-regional-overview-sub-saharan-africa/>>. Acesso em: 15 nov. 2007.

UNITED NATIONS EDUCATIONAL. Scientific and Cultural Organization. EFA progress: where do we stand? In: __. *EFA Global Monitoring Report 2006*. [2006?]. cap. 2, p.37-73. Disponível em: <http://portal.unesco.org/education/es/ev.php-URL_ID=43283&URL_DO=DO_TOPIC&URL_SECTION=201.html> Acesso em: 15 nov. 2007.

WILLETT, J. B. Measuring change: what individual growth modeling buys you. In: AMSEL, E.; RENNINGER, K. A. (Ed). *Change and development: issues of theory, method and application*. Mahwah,: Lawrence Erlbaum Associates, 1997. cap. 11, p. 213-243.

WILLETT, J. B. Some results on reliability for the longitudinal measurement of change: implications for the design of studies of individual growth. *Educational and Psychological Measurement*, Durham v. 49, n. 3, p. 587-602, 1989. Disponível em: <<http://www.gseacademic.harvard.edu/~willettj/paperdownloads.htm>>. Acesso em: 16 dez. 2007.

ANEXOS

TABELA 17

Percentual de alunos segundo variáveis fixas no tempo e Região. Ficha B, 1999-2003.

Regiões/covariáveis	Sexo		Cor		Repetência antes 4ª	
	Feminino	Masculino	Branca	Outra	Sim	Não
Norte	49,72	50,28	39,30	54,72	48,22	40,96
Nordeste	51,15	48,85	33,05	62,07	52,42	39,10
Centro-Oeste	49,03	50,97	49,58	47,30	35,80	54,78

Fonte: Ficha B - CEDEPLAR/INEP 1999/2003.

TABELA 18

Percentual de alunos segundo a situação de trabalho e Região: Ficha B, 1999-2003

Covariável	Norte			Nordeste			Centro-Oeste		
	Sim	Não	n	Sim	Não	n	Sim	Não	N
Trabalho									
Tempo 1	20,40	79,60	3.187	14,06	85,94	3.180	19,65	80,35	3.689
Tempo 2	17,58	82,42	2.128	13,39	86,61	1.905	20,19	79,81	2.442
Tempo 3	11,75	88,25	1.659	12,50	87,50	1.416	10,62	89,38	1.958
Tempo 4	11,09	88,91	1.172	11,53	88,47	1.015	12,99	87,01	1.532

Fonte: Ficha B - CEDEPLAR/INEP 1999/2003.

TABELA 19

Percentual de alunos por nível socioeconômico, proficiência e Região: Ficha B, 1999-2003

Covariáveis	Norte			Nordeste			Centro-Oeste		
	Mín.	Máx.	Méd.	Mín.	Máx.	Méd.	Mín.	Máx.	Méd.
*INVARIÁVEL									
Ind. Nível Socioec_1	-3,23	5,74	-0,031	-3,39	2,83	-0,254	-2,54	3,24	0,046
Ind.Nível Socioec_2	-8,72	2,16	-1,046	-8,51	2,16	-1,039	-8,81	2,16	-1,078
*VARIÁVEL									
Proficiência anterior									
Tempo 1	-23,43	22,76	0,044	-20,17	26,85	0,058	-24,18	24,09	-0,048
Tempo 2	-23,68	22,19	-0,258	-23,16	25,98	0,065	-24,75	21,82	-0,267
Tempo 3	-20,75	18,84	-0,0006	-23,19	23,56	-0,015	-21,79	20,43	-0,157
Tempo 4	-19,51	21,39	-0,026	-20,31	25,80	0,180	-24,35	21,62	-0,153

Fonte: Ficha B - CEDEPLAR/INEP 1999/2003.

TABELA 20

Variáveis de escola segundo Regiões: Ficha B e Censo Escolar de 1999

Covariáveis	Norte			Nordeste			Centro-Oeste		
	Mín.	Máx.	Méd.	Mín.	Máx.	Méd.	Mín.	Máx.	Méd.
*FIXAS									
Matriculas 4ª série	39	266	139,35	32	300	117,36	29	210	102,57
Matric. 4ª Prom.	24	266	119,22	28	270	101,55	26	184	93,35
Professor niv sup.	1	49	20,57	6	51	29,10	1	51	22,77
Salas	5	40	14,10	6	24	13,61	6	23	12,80
Ind.Estrutura_1	-2,36	1,63	-0,20	-2,36	1,63	-0,18	-2,36	1,63	0,02
Ind.Estrutura_2	-1,48	3,18	0,02	-1,48	2,03	-0,20	-1,48	3,18	0,08
Ind.Segurança_1	-2,36	1,57	0,10	-2,36	1,57	0,29	-2,59	1,79	-0,43
Ind.Segurança_2	-1,32	2,55	-0,21	-1,32	2,55	-0,02	-1,57	2,93	0,22
*VARIÁVEIS									
Profic Méd. Esc									
Tempo 1	38,43	56,49	79,00	41,30	55,78	66,86	43,37	59,44	63,67
Tempo 2	46,92	59,93	79,00	47,90	59,89	66,86	53,12	63,36	61,61
Tempo 3	52,58	64,12	79,00	49,56	64,21	68,22	54,09	66,36	62,62
Tempo 4	52,15	64,63	80,88	50,29	64,78	66,86	56,22	70,56	62,62

Fonte: Ficha B - CEDEPLAR/INEP 1999/2003.

Quadro 3: Coordenadas das categorias nas duas dimensões, sistema de segurança na escola: Ficha B, 1999.

VARIÁVEL	CATEG.	DIMENSÃO	
		1	2
Policial/vigia	não	0,475	0,534
	sim	- 0,807	- 0,948
Controle de entrada estranhos	não	1,282	- 0,029
	sim	- 0,361	0,007
Sistema proteção contra incêndio	não	0,252	- 0,217
	sim	- 1,855	1,746

Fonte: Ficha B - CEDEPLAR/INEP 1999/2003.

Quadro 4: Coordenadas das categorias nas duas dimensões, Estrutura básica da escola: Censo Escolar 1999.

VARIÁVEL	CATEG.	DIMENSÃO	
		1	2
Quadra	Não	0,696	0,369
	Sim	- 0,479	- 0,254
Biblioteca	Não	0,957	-0,182
	Sim	- 0,516	0,098
Laboratório de Informática	Não	0,162	- 0,258
	Sim	- 1,426	2,274
Sala TV/Vídeo	Não	0,378	0,374
	Sim	- 0,763	- 0,755

Fonte: Censo Escolar de 1999.

Quadro 5: Coordenadas das categorias nas duas dimensões,
Nível socioeconômico familiar: Ficha B, 1999

VARIÁVEL	CATEG.	DIMENSÃO	
		1	2
Geladeira	Não	- 0,750	- 1,470
	Sim	0,073	0,146
Aspirador de pó	Não	- 0,267	0,138
	Sim	1,385	- 0,743
Freezer	Não	- 0,461	0,195
	Sim	0,844	- 0,360
Rádio	Não	- 1,229	- 3,592
	Sim	0,083	0,218
TV	Não	- 1,295	- 1,298
	Sim	0,099	0,094
Vídeo-cassete	Não	- 0,580	0,035
	Sim	0,738	- 0,048
Máquina de Lavar	Não	- 0,656	- 0,071
	Sim	0,430	0,048
Automóvel	Não	- 0,451	0,053
	Sim	0,796	- 0,096
Empregada doméstica	Não	- 0,208	0,113
	Sim	1,083	- 0,602

Fonte: Ficha B - CEDEPLAR/INEP 1999/2003.