

DANIELLE RAMOS DE MIRANDA PEREIRA

**FATORES ASSOCIADOS AO DESEMPENHO ESCOLAR NAS DISCIPLINAS  
DE MATEMÁTICA E DE PORTUGUÊS NO ENSINO FUNDAMENTAL:  
UMA PERSPECTIVA LONGITUDINAL**

Belo Horizonte, MG  
Centro de Desenvolvimento e Planejamento Regional  
Faculdade de Ciências Econômicas – UFMG  
2006

DANIELLE RAMOS DE MIRANDA PEREIRA

**FATORES ASSOCIADOS AO DESEMPENHO ESCOLAR NAS DISCIPLINAS  
DE MATEMÁTICA E DE PORTUGUÊS NO ENSINO FUNDAMENTAL:  
UMA PERSPECTIVA LONGITUDINAL**

Tese apresentada ao Curso de Doutorado em Demografia 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.

Orientadora: Prof<sup>ª</sup>. Cibele Comini César

Co-orientador: Prof. Eduardo Luiz Gonçalves Rios-Neto

Belo Horizonte, MG  
Centro de Desenvolvimento e Planejamento Regional  
Faculdade de Ciências Econômicas – UFMG  
2006

Ao Marcelo

## **AGRADECIMENTOS**

Agradeço a Deus, que sempre esteve ao meu lado, mesmo nos momentos em que senti solidão. Fazer uma tese é um trabalho bastante solitário, mas tenho certeza de que Ele não me deixou um só momento. Por isto, cheguei aqui, no fim de mais uma “batalha”.

Agradeço ao Marcelo, meu amado marido, primeiro e único amor. Ele foi muito além de marido; foi amigo, companheiro, pai, mãe e irmão. Esteve ao meu lado, especialmente nos momentos mais difíceis em que a vida me colocou. Sereno e sempre paciente, soube me ouvir, compreender, consolar e ajudar. Seu amor e seu apoio fazem com que esta conquista seja nossa.

Agradeço aos meus pais e ao meu irmão, Heverton, pelo estímulo, pela força e pelo amor incondicional.

Bem, então, parece que fazer uma tese não foi um trabalho tão solitário assim. Além da força de Deus, do amor dos meus familiares e amigos, pude contar com as orientações da professora Cibele Comini César e do professor Eduardo Luiz Gonçalves Rios-Neto.

Gostaria de agradecer à professora Cibele, orientadora desta tese. Especialmente, gostaria de ressaltar a sua competência, que muito contribuiu para a realização deste trabalho.

Sou imensamente grata ao professor Eduardo Luiz Gonçalves Rios-Neto, meu co-orientador, pela oportunidade de participar da pesquisa “Avaliação do desempenho: fatores associados”. Se não fosse pela confiança depositada em mim, não teria tido essa oportunidade que tanto contribuiu para o meu amadurecimento como pesquisadora e para a elaboração desta tese.

Agradeço ao professor Roberto Nascimento. Ele foi um amigo. Agradeço, também, a todos os professores do CEDEPLAR, aos funcionários da secretaria, da biblioteca e do setor financeiro e aos meus amigos, em especial, à Luíza.

Agradeço ao CNPq pela bolsa que me foi concedida, proporcionando a oportunidade de me entregar totalmente ao doutorado durante quatro anos. Não poderia deixar de agradecer ao INEP, que em convênio com o CEDEPLAR, cederam os dados da pesquisa “Avaliação do desempenho: fatores associados” para a elaboração desta tese. Sou grata aos participantes desta pesquisa que, de alguma forma, contribuíram para a produção desta tese. Em especial, gostaria de agradecer ao Maurício Lima pelo grande auxílio prestado.

# SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE GRÁFICOS

RESUMO

ABSTRACT

<b>1 INTRODUÇÃO</b>	<b>01</b>
<b>2 REFERENCIAL TEÓRICO SOBRE OS PRINCIPAIS FATORES ASSOCIADOS À PROFICIÊNCIA EDUCACIONAL</b>	<b>06</b>
2.1 Contextualização do desempenho educacional brasileiro	06
2.2 O modelo teórico da proficiência educacional	11
2.3 A importância relativa dos fatores familiares e escolares	13
2.4 Variáveis explicativas a nível da escola	16
2.4.1 Variáveis explicativas relacionadas aos recursos escolares	16
2.4.2 Variáveis explicativas relacionadas ao processo escolar	21
	<b>22</b>
<b>3 METODOLOGIA</b>	
3.1 Banco de dados: histórico e breve descrição da amostragem	22
3.2 Principais características da Teoria Clássica do Teste (TCT) e a da Teoria da Resposta ao Item (TRI) na avaliação de testes educacionais	24
3.2.1 Uma introdução à Teoria Clássica do Teste	24
3.2.2 Introdução à Teoria da Resposta ao Item	26
3.2.2.1 Os modelos mais comuns para uma população (ou grupo)	28
3.2.2.2 Pressupostos	30
3.2.2.3 A escala da habilidade	31
3.2.2.4 A função de informação	32
3.2.2.5 Determinação da qualidade do item	34
3.2.2.6 Modelo para mais de uma população (ou mais de um grupo)	35
3.3 Os principais processos de equalização de acordo com a Teoria da Resposta ao Item	36
3.3.1 Definição de equalização e de sua importância em estudos longitudinais	36
3.3.2 A equalização horizontal (via população) e a equalização vertical (via itens comuns)	37
3.4 Modelos para o desempenho escolar	40
3.4.1 Modelos hierárquicos	41

3.4.2 Modelos hierárquicos longitudinais	42
3.4.2.1 Modelo incondicional	43
3.4.2.2 Modelo condicional	47
<b>4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS</b>	<b>50</b>
4.1 Procedimentos adotados na calibração e na equalização dos itens de português e de matemática	50
4.2 Análise Descritiva	56
4.2.1 A reorganização do banco de dados e o tratamento das variáveis para o ajuste dos modelos	56
4.2.2 Estatísticas descritivas das variáveis resposta e explicativas consideradas nos modelos	60
4.3 Ajuste de modelos hierárquicos longitudinais para o desempenho educacional de português e de matemática	71
4.3.1 Passo a passo do ajuste dos modelos	71
4.3.2- Ajuste do modelo incondicional e estimação do efeito-escola por disciplina	75
4.3.2.1- Efeito-escola de acordo com os modelos hierárquicos longitudinais	78
4.3.3- Ajuste de modelos condicionais para as disciplinas de português e matemática	79
<b>5 CONSIDERAÇÕES FINAIS</b>	<b>86</b>
5.1 Contribuições da tese	86
5.2 Limitações da tese e possibilidades de estudos futuros	90
<b>6 REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>92</b>
<b>ANEXO I</b>	<b>100</b>

## LISTA DE FIGURAS

FIGURA 1– Estrutura da tese	05
FIGURA 2- Modelo insumo-processo-produto com algumas adaptações	12
FIGURA 3 - Modelo Conceitual da Pesquisa	13
FIGURA 4- Curva Característica do Item para um modelo logístico de três parâmetros	30
FIGURA 5- Curva Característica de quatro itens e suas funções de informação correspondentes	34

## LISTA DE GRÁFICOS

GRÁFICO 1- Proficiência média na 4ª série do ensino fundamental em português para o Brasil e suas regiões	07
GRÁFICO 2- Proficiência média na 4ª série do ensino fundamental em matemática para o Brasil e suas regiões	08
GRÁFICO 3- Taxa de atendimento do ensino fundamental brasileiro em 2000	11
GRÁFICO 4- Média dos testes de português por período	62
GRÁFICO 5- Média dos testes de matemática por período	63
GRÁFICO 6- Média dos testes de português por período e estado	65
GRÁFICO 7- Média dos testes de matemática por período e estado	65
GRÁFICO 8- Variância Total x idade centralizada para a disciplina de português	77
GRÁFICO 9- Variância Total x idade centralizada para a disciplina de matemática	77



## LISTA DE TABELAS

TABELA 1- Classificação dos países segundo a posição da mediana em relação à média da região, por série e disciplina	9
TABELA 2- Distribuição percentual dos efeitos estimados de recursos-chave em rendimento educacional, baseando-se em 376 funções de produção ajustadas	19
TABELA 3- Procedimentos para a calibração e retirada de itens, referentes às provas de português nos períodos de abril de 1999 a novembro de 2003	51
TABELA 4- Esquema longitudinal dos itens das provas de português de abril/1999 a novembro/2003	52
TABELA 5- Procedimentos para a calibração e retirada de itens, referentes às provas de matemática nos períodos de abril de 1999 a novembro de 2003	53
TABELA 6- Esquema longitudinal dos itens das provas de matemática de abril/1999 a novembro/2003	54
TABELA 7- Distribuição dos valores do parâmetro de discriminação nos testes de português e de matemática	55
TABELA 8- Frequência de alunos por número de participação nas rodadas dos testes de português e de matemática	56
TABELA 9- Número de alunos por período e por disciplina	57
TABELA 10- Número de escolas por período de realização das provas	57
TABELA 11- Valores das cargas padronizadas na tentativa inicial de estimação dos índices de infra-estrutura das escolas	59
TABELA 12 - Número de professores por série na disciplina de português	60
TABELA 13 - Número de professores por série na disciplina de matemática	60
TABELA 14- Média e desvio-padrão dos testes de português e de matemática por período	62
TABELA 15- Média e desvio-padrão dos testes de português por período e estado	63
TABELA 16- Média e desvio-padrão dos testes de matemática por período e estado	64

TABELA 17- Percentual de alunos que trabalham por período para a disciplina de português	66
TABELA 18- Percentual de alunos que trabalham por período para a disciplina de matemática	66
TABELA 19- Distribuição dos alunos por sexo e por período para a disciplina de português	67
TABELA 20- Distribuição dos alunos por sexo e por período para a disciplina de matemática	67
TABELA 21- Percentual de alunos que repetiram de ano por período para a disciplina de português	67
TABELA 22- Percentual de alunos que repetiram de ano por período para a disciplina de matemática	68
TABELA 23- Percentual dos valores dos índices de infra-estrutura de conservação por período nas escolas para as disciplinas de português e de matemática	68
TABELA 24- Percentual de escolas, de acordo com o tamanho médio das turmas por período	69
TABELA 25- Média e desvio padrão do tamanho médio das turmas por período	69
TABELA 26- Percentual do tipo de sistema de recuperação das escolas por período	69
TABELA 27- Percentual do número de vezes em que os conselhos de escolas se reuniram por período	71
TABELA 28- Resultados do ajuste do modelo incondicional por disciplina	76
TABELA 29- Resultados do ajuste de modelos condicionais para a disciplina de português	79
TABELA 30 - Resultados do ajuste de modelos condicionais para a disciplina de matemática	80
TABELA 31- Resultados do ajuste de modelos condicionais mais complexos para as disciplinas de português e de matemática	82

## RESUMO

Diante do atual contexto educacional brasileiro, em que a qualidade do ensino fundamental vem ganhando cada vez mais importância, o objetivo principal dessa tese é investigar os fatores associados ao desempenho educacional, enfatizando a importância dos fatores escolares nos resultados educacionais nas disciplinas de português e de matemática em seis estados: Pernambuco, Rondônia, Pará, Sergipe, Goiás e Mato Grosso do Sul.

Neste trabalho, a utilização de dados longitudinais possibilitou a estimação mais adequada do quanto o *status* inicial, a taxa instantânea de crescimento do desempenho e a curvatura desse crescimento explicam a variabilidade dos desempenhos entre as escolas, uma vez que esses efeitos-escola têm caráter longitudinal.

Um dos métodos adotados neste estudo foi o de modelos hierárquicos longitudinais, sendo necessária, anteriormente, a equalização dos desempenhos dos alunos ao longo do tempo por meio do método da Teoria da Resposta ao Item (TRI), para que esses desempenhos pudessem ser diretamente comparáveis.

Os resultados encontrados com o ajuste do modelo condicional final demonstram que apesar da escola ter uma importância fundamental no desempenho educacional dos alunos, não se pode negar a contribuição dos fatores familiares no resultado educacional de seus filhos. Sendo assim, um melhor desempenho educacional parece ser um *mix* de recursos familiares e escolares, podendo-se destacar o trabalho do aluno, a repetência, a escolaridade da mãe, o tamanho da turma e o estado em que a escola pertence.

Diante desses resultados, pode-se afirmar que a quantidade de recursos escolares tem importância para o desempenho escolar, uma vez que menores turmas significam maior número de professores e, portanto, maiores gastos em educação. Dentre os seis estados estudados, o de Mato Grosso do Sul e o de Goiás apresentaram os melhores desempenhos, por outro lado, o estado de Pernambuco tem os piores desempenhos em ambas as disciplinas.

Sendo assim, falando de qualidade da educação, tanto as políticas públicas em nível familiar, quanto aquelas em nível escolar têm um importante papel na contribuição para a melhoria da qualidade educacional brasileira. Desta forma, vale destacar a relevância de políticas públicas como a da bolsa família, por exemplo, apesar de seu custo elevado. Complementarmente, acredita-se que as políticas públicas em nível escolar apresentariam maior eficácia, se a quantidade e a distribuição dos recursos disponíveis levassem em conta as

diferentes realidades e necessidades regionais, de forma que os desempenhos educacionais estudados alcançassem níveis cada vez melhores e mais próximos entre si.

## ABSTRACT

Facing the actual Brazilian educational context - where the quality of the basic education is gaining more and more importance, the main goal of the present thesis is to investigate the factors associated to the educational performance, emphasizing the importance of school factors over the educational results of the following disciplines - Portuguese and mathematics - in six states: Pernambuco, Rondônia, Pará, Sergipe, Goiás and Mato Grosso do Sul.

In the present work, the use of linear data enabled the most appropriate estimate of the *initial status*, once the instantaneous rate of performance growth, and the curvature of that growth explain the variability of the performances among the schools - once those school-effects have a longitudinal character.

One of the methods adopted in this study was the longitudinal hierarchical models. However, it was necessary to previously equate the students' performances along time - which was accomplished through Item Response Theory (IRT) method-, in order for those performances to be directly comparable.

The results found with the adjustment of the conditional final model demonstrate that despite of the school's fundamental importance over the educational performance of the students, one cannot deny the contribution of family factors over the children's educational result. Consequently, a better educational performance seems to be a mix of family and school resources, in which the student's work, the student's repeating, the mother's education, the size of the class per school, and the Brazilian state where the school is located - stand out as the most prominent factors.

Against those results, it can be affirmed that the amount of school resources is important for the school's performance, once smaller classes mean larger number of teachers and, therefore, bigger expenses in education. Among the six states studied, Mato Grosso do Sul and Goiás presented the best performances. On the other hand, the state of Pernambuco showed the worst performances in both disciplines.

Therefore, when focusing quality of education, not only the public policies directed at family level, but as well as the policies aimed at the school (educational) level have a key role in the improvement of the Brazilian educational quality. Hence, it is worth to accentuate the relevance of public policies such as the "Bolsa Família", for instance, in spite of its high cost. In addition, is believed that public policies intended for towards education would be able to

offer a larger effectiveness if the amount and distribution of the available resources take into account the different regional realities and needs, so that the educational performances studied here could reach comparable, closer and better levels of efficiency.

## 1 - INTRODUÇÃO

Nos últimos anos, a avaliação educacional vem apresentando uma importância crescente nos diferentes níveis de ensino no Brasil (FRANCO & BONAMINO, 2001). Complementarmente, LÜDKE (2001) ressalta que essa avaliação tem centrado seu discurso em resultados, para que a educação possa conhecer melhor os seus problemas e os meios para buscar soluções. Desta forma, um argumento relevante para justificar a importância da avaliação educacional seria o melhor conhecimento dos problemas escolares e, conseqüentemente, a maior possibilidade de encontrar soluções para os problemas que afetam a qualidade da educação, tais como a insuficiência de recursos escolares.

A qualidade da educação tem sido medida por meio de testes padronizados do rendimento ou desempenho educacional (HANUSHEK, 2002). Este desempenho depende da habilidade (ou seja, da proficiência) do aluno e das características dos itens que compõem o teste. Segundo HAMBLETON (1993), a habilidade de um indivíduo, em qualquer dimensão do conhecimento humano, não deve ser assumida como inata, mas sim como uma habilidade cognitiva. Assim, a habilidade é uma característica latente, devendo ser estimada por meio de modelos estatísticos de variáveis latentes, como os modelos da Teoria da Resposta ao Item. Diante dessas considerações, esta tese adota como sinônimos a habilidade, a proficiência e o desempenho, embora se tenha conhecimento de que o desempenho não seja exatamente definido como a proficiência ou a habilidade.

No contexto brasileiro, pode-se afirmar que a preocupação com a avaliação educacional tornou-se mais intensa a partir dos anos 90, com a criação do Sistema Nacional de Avaliação de Educação Básica (SAEB). Com esta iniciativa, os formuladores de políticas públicas educacionais passaram a contar com melhores subsídios para formulação e monitoramento de políticas direcionadas à melhoria da qualidade da educação (SAEB, 1999).

Além do SAEB, o Brasil tem participado de avaliações internacionais, tal como a promovida em 1997 pela Organização das Nações Unidas para Educação, Ciências e Cultura (UNESCO) em parceria com a OREALC em onze países latino-americanos. No mesmo sentido, segundo INEP (2004), o Brasil vem, também, participando do Programa Internacional de Avaliação de Estudantes (PISA), organizado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE).

A análise dos resultados desses trabalhos permite identificar que a educação brasileira vem passando por uma grave crise, cuja solução deve estar direcionada à resolução de

problemas que afetam a qualidade do ensino brasileiro. Dessa forma, podem ser apontadas algumas questões no sentido de justificar a importância dos estudos que auxiliam os gestores na formulação de políticas públicas.

De início, vale destacar a relevância dos trabalhos que buscam auxiliar a melhoria da qualidade do ensino, uma vez que a cobertura do ensino fundamental já se encontra em patamares bastante elevados em todo o Brasil (SAEB, 1999). Assim, como o grave problema da educação brasileira está relacionado à qualidade da educação, é especialmente importante conhecer os fatores associados à proficiência educacional, visando auxiliar os formadores de políticas públicas na identificação de distorções presentes no sistema educacional, facilitar a elaboração de políticas que procurem corrigir estas distorções, buscando melhorar a qualidade da educação no Brasil.

Neste contexto, em que o foco das atenções tem sido direcionado para a qualidade da educação, é importante enfatizar a “janela de oportunidades” criada pelas condições recentes da dinâmica demográfica brasileira, em decorrência da queda da fecundidade e da desaceleração do nível do crescimento populacional (CERQUEIRA, 2004).

A “janela de oportunidades” foi um termo citado por CARVALHO & WONG (1995) para indicar a diminuição da pressão por acesso à escola em decorrência da redução na participação relativa da população em idade de cursar o ensino fundamental, ressaltando a importância de um re-direcionamento das políticas públicas educacionais.

Outro argumento reforça a necessidade de estudar os fatores escolares associados à proficiência educacional, num contexto tão desigual em termos de recursos, como o Brasil. Conforme afirmam BUCHMANN & HANNUM (2001), os recursos básicos da escola têm mais importância na proficiência dos alunos em contextos onde há maiores desigualdades nos recursos educacionais, como no caso dos países em desenvolvimento, incluindo, portanto, o Brasil.

Neste ponto, vale ressaltar que a utilização de dados longitudinais tem muito a contribuir com as avaliações da qualidade do ensino, uma vez que permite a análise da trajetória da proficiência dos alunos ao longo do tempo e, portanto, a análise do crescimento do desempenho ao longo do tempo.

Outra contribuição dos dados longitudinais diz respeito à estimação mais adequada do efeito-escola, de acordo com a abordagem de modelos hierárquicos longitudinais. Este efeito-escola é, geralmente, composto por três efeitos: 1) a porcentagem da variação do *status* inicial explicada pela variabilidade entre as escolas; 2) a porcentagem da variação da taxa instantânea de crescimento do desempenho explicada pela variabilidade entre as escolas; e 3)



a porcentagem da variação da taxa de aceleração (ou da curvatura) do crescimento explicada pela variabilidade entre as escolas (RAUDENBUSH & BRYK, 2002).

O banco de dados utilizado nesta tese, além de ser longitudinal, tem ainda a vantagem de ser o primeiro banco de dados longitudinal brasileiro na área de avaliação educacional, podendo-se afirmar que este estudo será um dos primeiros, no Brasil, a apresentar estimativas mais adequadas do efeito-escola em contextos brasileiros com os piores rendimentos educacionais: Nordeste, Norte e Centro-Oeste.

De uma forma geral, a literatura da proficiência educacional, incluindo especialmente áreas como Economia e Educação, estabelece três níveis de fatores aos quais a proficiência está associada: nível do aluno, nível da turma e nível da escola. Isto porque, em geral, os alunos estão aninhados em turmas e estas, em escolas.

À estrutura hierárquica presente nos dados educacionais é adequado o ajuste de modelos hierárquicos longitudinais, a fim de se estimar os fatores associados ao desempenho escolar. A utilização deste modelo hierárquico longitudinal é, principalmente, justificada pela estrutura de dependência presente nos dados educacionais longitudinais. Esta dependência, conforme descrição de WILLETT (1997), caracteriza-se, especialmente, por dois níveis hierárquicos, que constituem uma estrutura hierárquica natural, em que as características intra-alunos estão aninhadas nas características inter-alunos, fornecendo um importante arcabouço teórico para medidas de mudança. Nesta tese, um terceiro nível, o nível da escola, foi acrescentado a essa hierarquia, já que os alunos se encontram aninhados em escolas.

Diante do exposto, considerando a necessidade de se conhecer melhor os fatores associados ao desempenho escolar, visando a contribuir de alguma forma para a melhoria do sistema educacional brasileiro, esta tese pretende investigar quais os fatores escolares associados ao desempenho educacional entre a quarta e a oitava séries do ensino fundamental nas disciplinas de matemática e de português, controlando pelos fatores individuais ou familiares.

A estratificação educacional no Brasil em termos de fatores familiares foi citada por CAMARGO & BARROS (1991) ao afirmarem que o baixo nível educacional e a maior pobreza desencorajam os pais a investirem na educação, devido ao elevado custo de oportunidade da retirada dos seus filhos do mercado de trabalho. Entretanto, a escolha pela ênfase nos fatores escolares, controlando pelos fatores individuais ou familiares, está relacionada a uma maior dificuldade das políticas públicas em promover alterações nesses últimos fatores.

Tendo como metodologia a abordagem dos modelos hierárquicos longitudinais, a primeira hipótese desta tese é a grande importância da escola na explicação da variabilidade do desempenho educacional no ensino fundamental da quarta até a oitava série nas disciplinas de português e de matemática.

A segunda hipótese é a de que o tamanho da turma está negativamente associado ao desempenho escolar, uma vez que os gastos em educação são dados pelo tamanho da turma e pelo salário dos professores (KRUEGER, 2003). Desta forma, de acordo com o referido autor, acredita-se que a quantidade de recursos escolares afeta o desempenho educacional, indo de encontro ao pensamento de que a forma de utilização, e não a quantidade, dos recursos está associada ao desempenho educacional, conforme afirma HANUSHEK (2003). Vale ressaltar que os principais trabalhos sobre os efeitos do tamanho da turma no desempenho educacional foram realizados nos Estados Unidos, não se tendo notícias da realização de trabalho sobre esse tema de grande repercussão no Brasil.

A terceira hipótese é a de que o desempenho educacional está associado às regiões nas quais se encontram as escolas, uma vez que a política educacional dos estados pode ser a mesma, afetando diferentemente o desempenho escolar. Neste sentido, BUCHMANN & HANNUM (2001) alerta que os recursos básicos da escola têm mais importância na proficiência dos alunos em contextos onde há maiores desigualdades nos recursos educacionais.

Em síntese, o objetivo central dessa tese é investigar quais os fatores escolares associados ao desempenho educacional entre a quarta e a oitava séries do ensino fundamental nas disciplinas de matemática e de português, controlando pelos fatores individuais ou familiares. Este objetivo está relacionado à estimação adequada do efeito-escola, contando com a adoção das abordagens dos modelos hierárquicos longitudinais.

Conforme a FIG.1 que ilustra a estrutura da tese, no capítulo 1 foram apresentadas as justificativas, assim como, a importância de se conhecer um pouco mais sobre a qualidade da educação brasileira. Ademais, a questão a ser investigada foi definida e as hipóteses foram formuladas.

O capítulo dois apresenta a contextualização do desempenho educacional brasileiro. Destaca-se o modelo a partir do qual se adaptou o modelo teórico dessa tese, discutindo os principais fatores associados à proficiência educacional.

O capítulo três, inicialmente, apresenta o histórico do banco de dados e uma breve descrição da amostra. A seguir, descreve a metodologia utilizada, para que se possa atingir os objetivos propostos e verificar as hipóteses enunciadas. O primeiro método adotado foi a

Teoria da Resposta ao Item (TRI) para estimar a variável resposta, de forma que os desempenhos pudessem ser diretamente comparáveis. Este capítulo especifica ainda os modelos hierárquicos adotados na análise longitudinal.

O capítulo quatro mostra os resultados, desde o tratamento de variáveis explicativas, passando pela análise descritiva das variáveis resposta e explicativas até o ajuste dos modelos hierárquicos longitudinais. Por fim, o último capítulo apresenta as conclusões do estudo, abordando suas contribuições e limitações, assim como sugestões para trabalhos futuros.

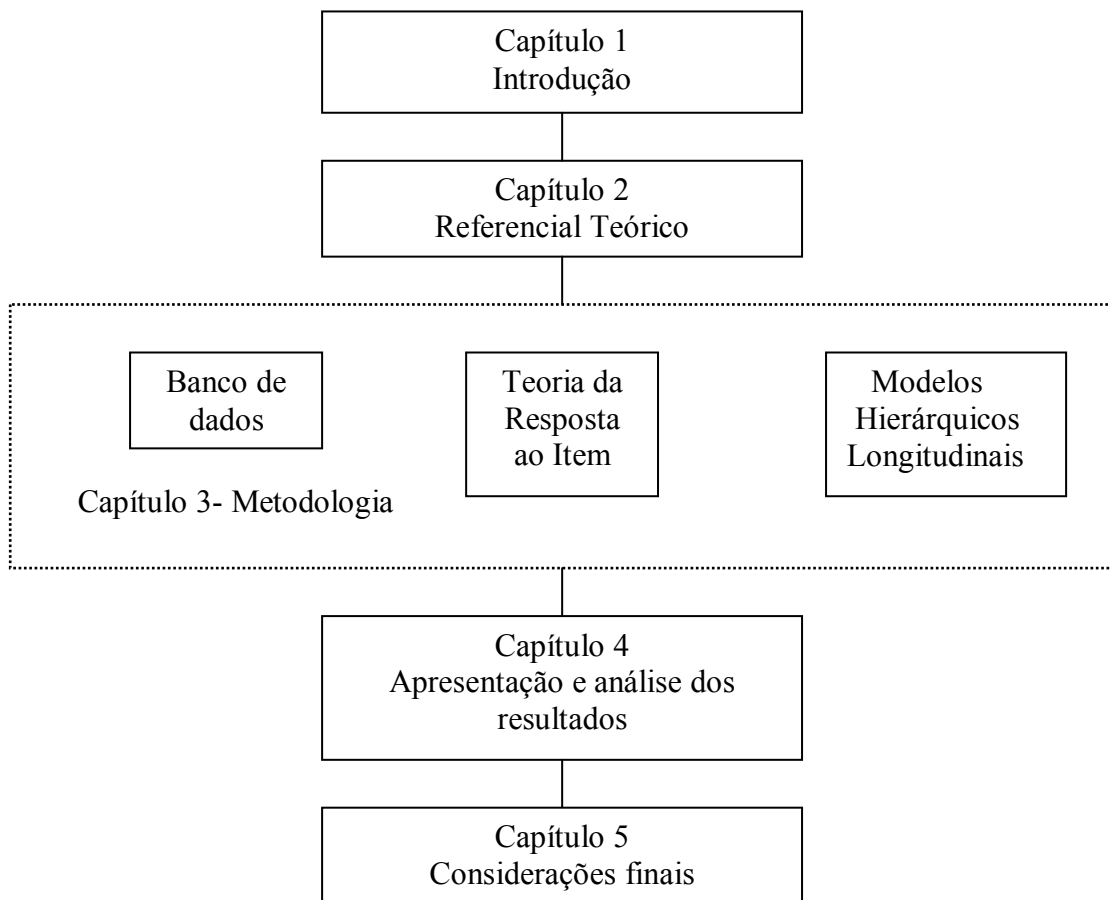


FIGURA 1– Estrutura da tese

## **2- REFERENCIAL TEÓRICO SOBRE OS PRINCIPAIS FATORES ASSOCIADOS À PROFICIÊNCIA EDUCACIONAL**

Este capítulo apresenta uma contextualização do desempenho educacional brasileiro em nível nacional e internacional. A seguir, destaca-se o modelo teórico a partir do qual se adaptou o modelo conceitual desse trabalho, discutindo os principais fatores associados à proficiência educacional.

### **2.1- Contextualização do desempenho educacional brasileiro**

A preocupação com a avaliação educacional no Brasil tornou-se mais intensa a partir da década de 90, com a criação do Sistema Nacional de Avaliação de Educação Básica (SAEB), que tem como objetivo fornecer elementos para a formulação e o monitoramento de políticas direcionadas à melhoria da qualidade da educação (SAEB, 1999).

O SAEB testa amostras de alunos, em todos os estados e no distrito federal, da 4ª e 8ª séries do ensino fundamental e da 3ª série do ensino médio de escolas públicas ou particulares. Até 1997, as disciplinas testadas foram matemática, português e ciências, que no ensino médio correspondem à física, química e biologia. A partir de então, foram incluídas as disciplinas de história e geografia. Em todas as turmas sorteadas, todos os alunos respondem a uma das provas (FRANCO & BONAMINO, 2001; SAEB, 2002). A avaliação do SAEB vem acontecendo a cada dois anos e desde 1995 passou a adotar a abordagem da Teoria da Resposta ao Item, possibilitando expressar em uma mesma escala a proficiência dos alunos nas séries e nas disciplinas testadas (ANDRADE, TAVARES & VALLE, 2000).

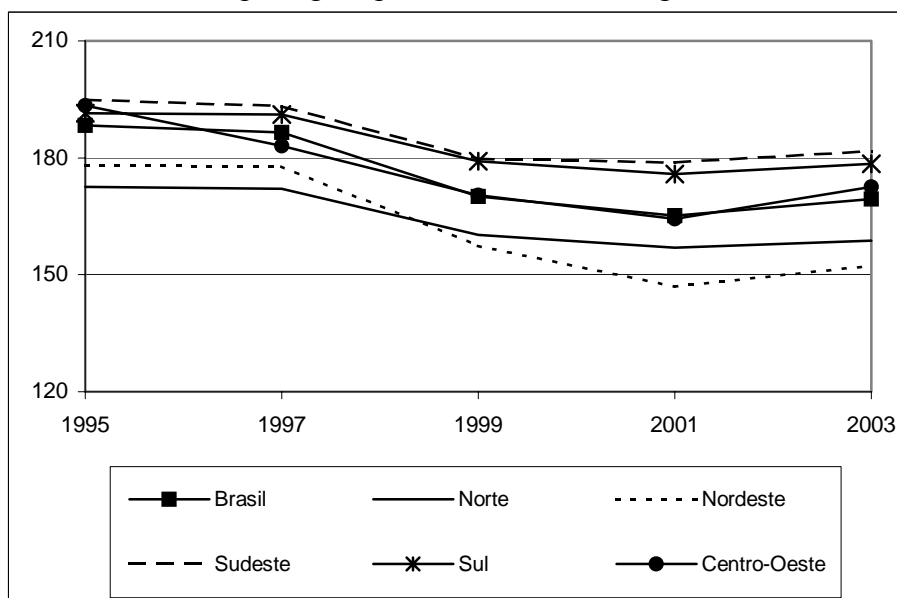
Uma análise dos dados publicados pelo SAEB traz algumas conclusões importantes. Os GRAF.1 e 2 mostram a evolução da proficiência educacional para a 4ª série do ensino fundamental, em português e matemática, para o Brasil e suas regiões. Uma proficiência de 200 é considerada próxima ao adequado nas duas disciplinas (SAEB, 2004).

Como já era esperado, nas duas disciplinas, a proficiência educacional é maior nas regiões Sudeste e Sul. A seguir, vem a Centro-Oeste e, por último, as regiões Nordeste e Norte. Na disciplina de português, “depois de três períodos de avaliação mostrando quedas consecutivas, a proficiência em leitura dos estudantes da 4ª série do ensino fundamental apresenta uma pequena inversão de tendência” (SAEB, 2004, p.7). O SAEB testou ainda se

essas diferenças entre as proficiências médias eram estatisticamente significativas entre os anos de 2001-2003. Somente o Brasil como um todo e as regiões Nordeste e Centro-Oeste apresentaram proficiências médias em português estatisticamente diferentes neste período.

Em matemática, a proficiência média do Brasil entre 1995 e 1997 é praticamente a mesma. De 1997 a 2001, houve uma queda de aproximadamente 8%, passando a proficiência média de 190,8 para 176,3. Em 2003, a proficiência foi de 177,1, de forma que os testes estatísticos não rejeitaram a hipótese de igualdade entre as proficiências de 2001 e 2003.

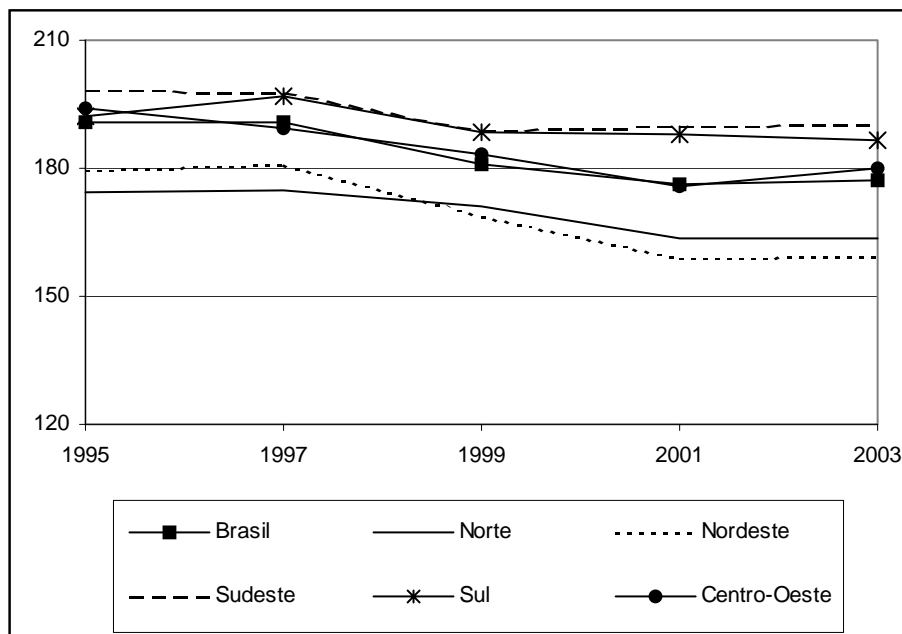
GRÁFICO 1- Proficiência média na 4ª série do ensino fundamental em português para o Brasil e suas regiões



FONTE: SAEB (2004)

Considerando essas variações no nível da proficiência média em matemática, as regiões têm aproximadamente a mesma estrutura na evolução de suas proficiências médias, com exceção da região Sul, onde foi observado um aumento de proficiência entre 1995 e 1997. Para as regiões nenhuma diferença significativa foi encontrada entre 2001 e 2003 (SAEB, 2004).

GRÁFICO 2 - Proficiência média na 4ª série do ensino fundamental em matemática para o Brasil e suas regiões



FONTE: SAEB (2004)

O relatório do SAEB (2004) também apresentou resultados da proficiência, classificando-os, por meio das habilidades desenvolvidas pelos alunos, em quatro estágios: adequado, intermediário, crítico e muito crítico. Por exemplo, um aluno num estágio crítico ou muito crítico desenvolveu habilidades muito elementares para a série cursada. Em 2003, os alunos brasileiros que se encontravam em estágio crítico ou muito crítico, referente às habilidades em português, correspondiam a 55,4% dos alunos pesquisados. Para o mesmo período em matemática, esse percentual 51,6% não é muito diferente do obtido em português.

Analisando a situação de cada uma das regiões do País, ARAÚJO & LUZIO (2004) constataram uma realidade marcada por fortes desigualdades. No Nordeste, a soma dos níveis muito crítico e crítico, em português, corresponde a 75% dos alunos da 4ª série, enquanto no Sul, esta soma totaliza 47% e no Sudeste, 44%. Em matemática, o Nordeste concentra 69% dos estudantes nesses mesmos estágios, enquanto no Sul esse percentual é de 41% e no Sudeste, de 39%.

Além do SAEB, o Brasil tem participado de avaliações internacionais, tal como a promovida em 1997 pela Organização das Nações Unidas para Educação, Ciências e Cultura (UNESCO) em parceria com a OREALC em onze países latino-americanos. A TAB.1 mostra um resumo dos resultados dessa avaliação, classificando os países segundo a posição da

mediana em relação à média da região, por série e disciplina. No documento da UNESCO, “os resultados realçam a liderança educacional de Cuba e são consistentemente favoráveis aos sistemas educacionais do Brasil e da Bolívia e desfavoráveis em relação à Venezuela” (FRANCO & BONAMINO, 2001, p.25).

TABELA 1- Classificação dos países segundo a posição da mediana em relação à média da região, por série e disciplina

	Língua		Matemática	
	3ª série	4ª série	3ª série	4ª série
1º	Cuba (343)	Cuba (349)	Cuba (351)	Cuba (353)
2º	Argentina (263) Chile (259) Brasil (256)	Chile (286) Argentina (282) Brasil (277) Colômbia (265) México (252) Paraguai (251)	Argentina (251)	Argentina (269)  Brasil (269)  Chile (265)  Colômbia (258)  México (256)
3º	Venezuela (242) Colômbia (238) Bolívia (232) Paraguai (229) México (224) R.Dominicana (220) Honduras (216)	Venezuela (249) Honduras (238) Bolívia (233) R.Dominicana (232)	Brasil (247) Chile (242) Colômbia (240) Bolívia (240) México (236) Paraguai (232) R.Dominicana (225) Venezuela (220) Honduras (218)	Paraguai (248)  Bolívia (245)  R.Dominicana (234)  Honduras (231)  Venezuela (226)

FONTE: FRANCO & BONAMINO (2001).

O Brasil vem também participando do Programa Internacional de Avaliação de Estudantes (PISA) organizado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE). O PISA é uma avaliação internacional mais abrangente que, no início, contou com a participação de 31 países (INEP, 2004). São eles: Alemanha, Austrália, Áustria, Bélgica, Brasil, Canadá, Coreia do Sul, Dinamarca, Espanha, Estados Unidos, Finlândia, França, Grécia, Hungria, Irlanda, Islândia, Itália, Japão, Letônia, Liechtenstein, Luxemburgo,

México, Nova Zelândia, Noruega, Polônia, Portugal, Reino Unido, República Tcheca, Rússia, Suécia, Suíça.

Este Programa avaliou amostras de jovens com 15 anos de idade que estavam matriculados em escolas. O cronograma estabelecia que a primeira avaliação ocorreria em 2000 com ênfase em leitura. A segunda avaliação estava prevista para ocorrer em 2003 com ênfase em matemática e, por último, a avaliação de 2006 daria ênfase a ciências (PISA, 2001; FRANCO & BONAMINO, 2001).

Entretanto, em 2001 foram incluídas mais dez nações como participantes do PISA- Albânia, Argentina, Chile, Bulgária, Hong-Kong, China, Indonésia, Israel, Macedônia, Peru, Tailândia - o que se convencionou chamar de “Pisa Ampliado”(INEP, 2004).

De acordo com INEP (2004), com a inclusão desses dez países, o Brasil passou de último lugar entre os 31 participantes para 37º entre os 41 países do “Pisa Ampliado”, no que se refere à proficiência em leitura (proficiência de 396, numa escala de 0 a 800).

Contudo, nas provas de matemática (proficiência de 334) e de ciências (proficiência de 375), o Brasil ficou em penúltimo lugar, superando apenas a posição do Peru. O INEP (2004) salienta ainda que na média das três áreas avaliadas, a proficiência brasileira também ficou em penúltimo lugar.

Os resultados do PISA só confirmam o que o SAEB já vem divulgando sobre a baixa proficiência dos alunos brasileiros. Por exemplo, conforme dito anteriormente, em 2003 o SAEB constatou que 55,4% e 51,6% dos alunos da 4ª série não tinham desenvolvido habilidades básicas relacionadas, respectivamente, à matemática e ao português.

Diante do exposto, pode-se dizer que a educação brasileira vem passando por uma grave crise, cuja solução deve estar direcionada à resolução de problemas que afetam a qualidade do ensino brasileiro. Dessa forma, podem ser apontadas algumas questões no sentido de justificar a importância dos estudos que auxiliam os gestores na formulação de políticas públicas.

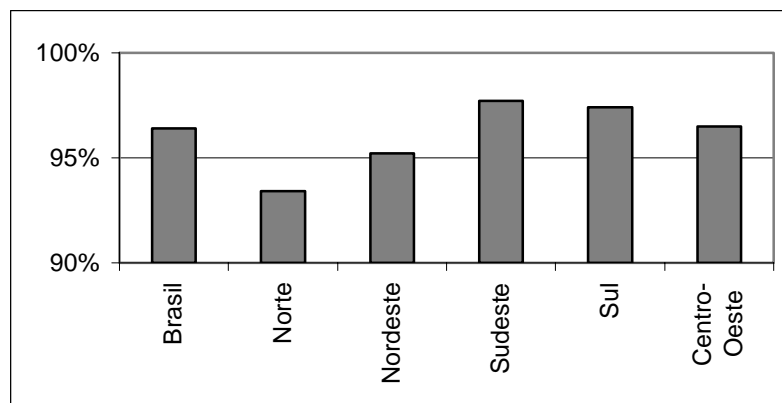
De início, a preocupação dos trabalhos que buscam auxiliar a melhoria da qualidade do ensino é importante, uma vez que a cobertura do ensino fundamental já se encontra em patamares bastante elevados em todo o Brasil. Essa cobertura é fruto de um processo de expansão da matrícula que, conforme SAEB (1999), reflete a implementação bem sucedida de políticas públicas orientadas para a consecução das metas de universalidade do atendimento escolar.

Conforme mostra o GRAF. 3, os resultados da taxa de atendimento (indicador que permite avaliar o acesso da população ao sistema educacional e representa o percentual da



população de idade escolar que frequenta a escola, podendo ser calculado para a faixa etária de 7 a 14 anos), indicam que o acesso à educação no Brasil já pode ser considerado universal, apesar da pequena diferença da taxa de atendimento na Região Norte em relação a essa taxa nas demais regiões.

GRÁFICO 3- Taxa de atendimento do ensino fundamental brasileiro em 2000



FONTE: MEC/INEP e IBGE (2004)

Assim, como o grave problema da educação brasileira está relacionado à qualidade da educação, é essencial conhecer os fatores associados à proficiência educacional, auxiliando os formadores de políticas públicas na identificação de distorções presentes no sistema educacional, facilitando a elaboração de políticas que procurem corrigir estas distorções, buscando, assim, melhorar a qualidade da educação no Brasil.

## 2.2- O modelo teórico da proficiência educacional

Inicialmente, o modelo teórico utilizado para a proficiência educacional era o modelo insumo-produto que considerava apenas os insumos (recursos) e o produto (desempenho educacional), não separando processo e recursos escolares nem reconhecendo a estrutura hierárquica do sistema escolar. Contudo, os processos escolares foram se mostrando cada vez mais importantes no estudo do desempenho escolar, tanto que o conhecido modelo insumo-produto evoluiu para o modelo insumo-processo-produto (WILLMS, 1992; MURILLO, 1999; TORRECILLA, 2000).

O modelo insumo-processo-produto é um melhoramento do modelo insumo-produto, porque ele reconhece a estrutura hierárquica do sistema escolar e separa processo escolar de

fatores que estão fora do controle dos professores e diretores (WILLMS, 1992). Este modelo insumo-processo-produto divide os recursos (*inputs*) e os processos de aprendizagem em nível de aluno, de classe, de escola e de comunidade. A FIG. 2 mostra o modelo teórico insumo-processo-produto proposto por WILLMS (1992) com algumas adaptações. O modelo teórico de WILLMS (1992) inspirou a criação do modelo teórico adotado neste trabalho, conforme ilustra a FIG. 3.

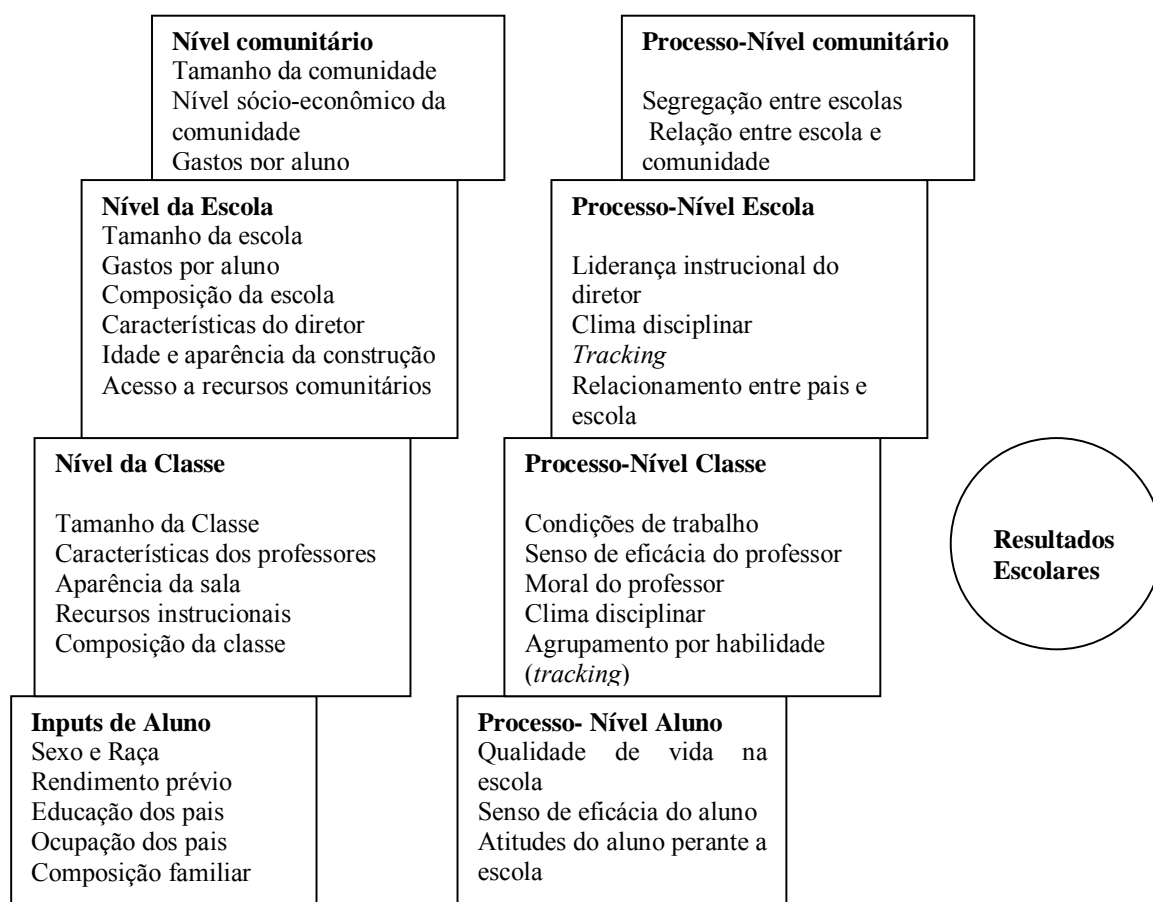


FIGURA 2- Modelo insumo-processo-produto com algumas adaptações

FONTE: WILLMS (1992)

Neste ponto, vale salientar que não é intenção do trabalho contemplar todos os possíveis determinantes da proficiência educacional, até mesmo porque grande parte dos fatores mencionados no modelo insumo-processo-produto de WILLMS (1992), especialmente aqueles relacionados ao processo de aprendizagem, não estão disponíveis no banco de dados.

Assim sendo, embora seja reconhecida a importância dos processos de aprendizagem para o resultado educacional (LEE & BRYK, 1989; WILLMS, 1992; HANUSHEK, 2002a,

2003), não se dispõe de dados suficientes para estimar mais adequadamente seus efeitos na proficiência educacional.

É importante ressaltar, que este trabalho, apesar dos objetivos principais estarem voltados à estimação dos efeitos das variáveis escolares na proficiência e do efeito-escola, pretende-se, também, considerar os insumos ou recursos relacionados ao aluno como variáveis controle no ajuste dos modelos hierárquicos longitudinais.

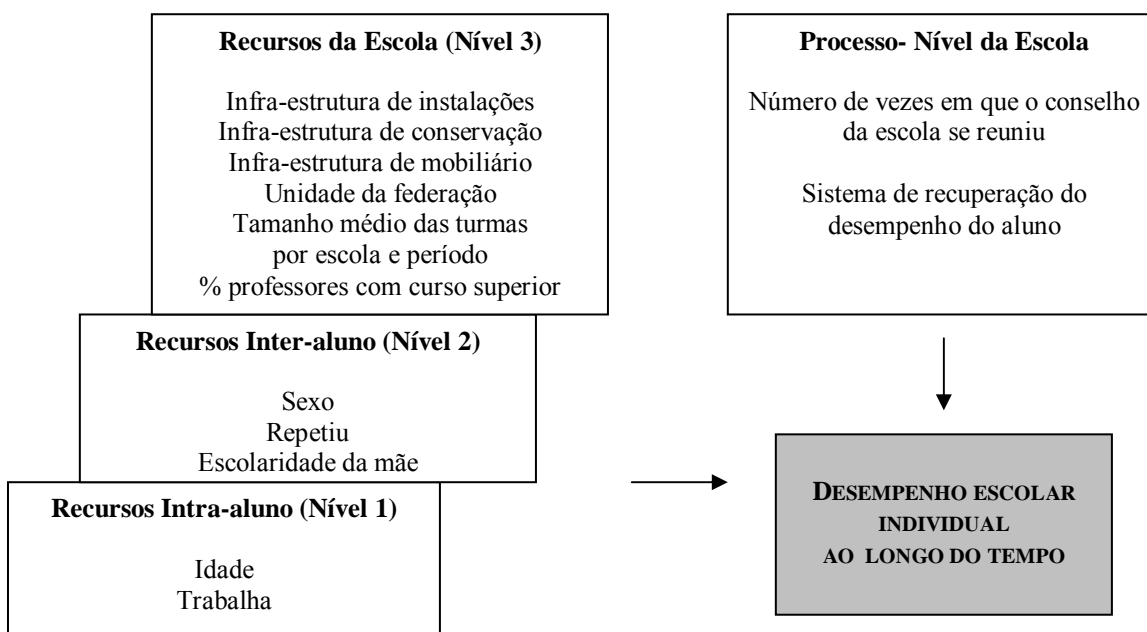


FIGURA 3 - Modelo Conceitual da Pesquisa

FONTE: WILLMS (1992)

### 2.3- A importância relativa dos fatores familiares e escolares

Esta seção tem a finalidade de introduzir o leitor no contexto da importância relativa entre os fatores familiares e os escolares como determinantes da proficiência do aluno.

Um dos primeiros estudos que buscou pesquisar a qualidade das escolas públicas e privadas nos Estados Unidos foi o Relatório de Coleman, publicado nos meados da década de 60 (HANUSHEK, 2003). Segundo o autor, este relatório apresentou um estudo sobre os fatores associados à proficiência educacional, mostrando que os efeitos dos recursos<sup>1</sup> da

<sup>1</sup> HANUSHEK (2002a) divide estes recursos, em resumo, da seguinte forma:

- 1) recursos da turma: educação e experiência do professor, tamanho da classe ou razão professor-aluno;
- 2) recursos financeiros: gastos por aluno e salário do professor.

escola na proficiência dos estudantes eram pequenos. Os fatores mais importantes para a proficiência seriam os familiares, seguido pelo efeito dos pares (agrupamento por habilidade).

Complementarmente, segundo HANUSHEK (2003), o relatório de Coleman foi o estudo que estimulou a produção de centenas de pesquisas empíricas sobre a qualidade da educação, tanto em Economia como na área de Educação, e instigou o debate sobre a importância dos recursos escolares para a proficiência do aluno.

Este debate sobre a importância dos recursos escolares na proficiência do aluno pode ser representado, principalmente, pelas opiniões de dois autores Eric Hanushek e Alan Krueger (TOOD & WOLPIN, 2003).

Para HANUSHEK (2003), a importância dos recursos escolares na proficiência não está associada à quantidade de recursos direcionados à escola. Portanto, para esse autor políticas baseadas na quantidade de recursos escolares possuem pouca chance de estarem sendo efetivas na melhoria da proficiência dos alunos, que estaria associada a uma utilização mais eficiente dos recursos.

Por outro lado, KRUEGER (2003) discorda dos resultados de Hanushek e da conclusão de que a quantidade de recursos da escola não são importantes para a proficiência do estudante. Esta diferença de opinião vem acompanhando o trabalho desse dois economistas, de forma que eles estão sempre debatendo suas opiniões em artigos publicados em periódicos da área, principalmente nos últimos dez anos (TOOD & WOLPIN, 2003).

Sobre a importância dos recursos escolares na proficiência, um outro estudo importante é o relatório preparado para a UNESCO por WILLMS & SOMERS (1999). Esses autores adotaram modelos hierárquicos para avaliar os resultados escolares de alunos da 3ª e 4ª séries do ensino fundamental, em língua e em matemática, na América Latina. Os resultados desse estudo sugerem que, controlando pelo *background* familiar, vários são os fatores escolares relacionados à proficiência educacional dos alunos na América Latina. Dentre tais fatores, os autores incluem os altos níveis de recursos da escola, como a infraestrutura e os materiais instrucionais.

Outra questão interessante no que tange à importância dos recursos escolares na proficiência foi levantada BUCHMANN & HANNUM (2001). Esses pesquisadores concluem que os recursos básicos da escola (biblioteca e qualificação dos professores, etc) parecem ser mais importantes em contextos em que há uma maior desigualdade de recursos educacionais, como é o caso dos países em desenvolvimento. Por outro lado, tais recursos básicos parecem ser menos importantes em contextos em que pelo menos um nível mínimo de recursos básicos é atingido, como ocorre nos países desenvolvidos.

No Brasil, geralmente, os estudos sobre os determinantes da proficiência educacional, utilizando modelos de regressão hierárquicos, têm conseguido na maioria das vezes mostrar a importância da escola na proficiência dos alunos (BARBOSA & FERNANDES, 2001; SOARES, CÉSAR & MAMBRINI, 2001; ALBERNAZ, FERREIRA & FRANCO, 2002).

Nesta tese, apesar dos fatores associados ao nível do aluno serem utilizados no ajuste dos modelos como controles, uma vez que se pretende enfatizar o efeito das variáveis relacionadas à escola no desempenho educacional, faz-se necessário realizar um breve comentário sobre essas variáveis, quais sejam: sexo, repetência, escolaridade da mãe e indicadora de que o aluno trabalha.

Em geral, a literatura considera que o sexo masculino apresenta uma proficiência superior em matemática, enquanto o sexo feminino supera a proficiência masculina em português, como pode ser verificado nos estudos de WILLMS & SOMERS (1999), SOARES, CÉSAR & MAMBRINI (2001), CÉSAR & SOARES (2001), ALBERNAZ, FERREIRA & FRANCO (2002), entre outros. Quanto à variável repetência, a reprovação é um dos fatores que interferem diretamente na aprendizagem dos alunos no ensino fundamental, nas disciplinas de português e de matemática (SAEB, 2004).

Alguns estudos utilizam a educação do pai como fator associado ao desempenho escolar, como por exemplo, o trabalho de BARBOSA & FERNANDES (2001). Contudo, vários são aqueles que ressaltam a escolaridade da mãe como um fator mais importante do que a escolaridade do pai na determinação do resultado educacional dos seus filhos, dentre eles, RIOS-NETO, CESAR & RIANI (2002) e RIANI (2004). Este fato justifica a opção por trabalhar com a variável escolaridade da mãe nessa tese.

Por sua vez, o fato do aluno trabalhar reduz seu tempo disponível ao estudo, podendo prejudicar o desempenho escolar. Ademais, o trabalho infantil tem sido combatido em todo o mundo, porque as crianças estão expostas ao trabalho perigoso e em condições inadequadas, levando à perda da infância e à diminuição do aprendizado (KASSOUF, 2002; SILVA & KASSOUF, 2002). Sobre o efeito do trabalho no rendimento do aluno, CÉSAR & SOARES (2001) mostram os resultados do ajuste de um modelo hierárquico, utilizando os dados do teste de matemática da 8ª série do SAEB-99, que considera as variáveis explicativas do aluno. Segundo esses autores, controlando por sexo, cor, filho e nível sócio-econômico, o fato do aluno trabalhar está associado a uma queda de aproximadamente 9 pontos no seu escore de matemática na oitava série.

## **2.4- Variáveis explicativas a nível da escola**

### **2.4.1-Variáveis explicativas relacionadas aos recursos escolares**

#### **a) Infra-estrutura das escolas**

Em suas análises sobre os efeitos da escola na proficiência de alunos da América Latina, WILLMS & SOMERS (1999) salientam a importância de altos níveis de recursos escolares, incluindo a infra-estrutura da escola e seu material instrucional. No mesmo sentido, o SAEB (2004) ressalta a importância de salas de aula adequadas, do acesso à biblioteca, laboratórios e quadras esportivas.

Em um estudo que procurava verificar, em especial, o efeito de variáveis de infra-estrutura na proficiência escolar, BARBOSA & FERNANDES (2001) propuseram, por meio de análise fatorial, quatro fatores relacionados à composição da infra-estrutura da escola e de seus equipamentos: conservação do prédio, condições de funcionamento dos espaços laboratoriais e de apoio, mobiliário e equipamento mínimo e, por último, instalações, áreas externas e de recreação. No modelo ajustado, controlando pelo *background* do aluno e da turma, esses quatro fatores se mostraram importantes para um melhor desempenho escolar.

Complementarmente, o estudo de ALBERNAZ, FERREIRA & FRANCO (2002) destaca a importância da infra-estrutura da escola na melhoria dos rendimentos escolares. Para os alunos da 8ª série no Brasil, considerando as quatro disciplinas do SAEB-99 (matemática, português, geografia e história), os resultados mostraram que existem escolas em que os estudantes estão aprendendo menos do que deveriam. Esses rendimentos mais baixos estão associados à insuficiência de recursos financeiros, à insuficiência de professores e sua baixa escolaridade, a salas barulhentas e/ou abafadas.

#### **b) Escolaridade do professor**

Vários trabalhos citam o nível educacional dos professores como um dos recursos escolares importantes no desempenho educacional, dentre eles, BARBOSA & FERNANDES (2001), ALBERNAZ, FERREIRA & FRANCO (2002) e SAEB (2004). Nesta tese, o percentual de professores com curso superior é uma das variáveis explicativas a ser considerada no ajuste dos modelos hierárquicos longitudinais.

### c) Tamanho da classe

A variável tamanho da turma tem causado grande divergência de opinião entre Hanushek e Krueger sobre os efeitos da quantidade de recursos escolares na proficiência. A princípio, a tendência é acreditar que reduções no tamanho da turma tenham um efeito positivo na proficiência do aluno. KRUEGER (1998, 2002, 2003) e HANUSHEK (1998a, 1998b, 2002a, 2002b, 2003), os maiores representantes da literatura sobre este tema, ao longo da última década vêm debatendo a efetividade da redução do tamanho da turma no rendimento escolar. Os principais trabalhos sobre os efeitos do tamanho da turma foram realizados nos Estados Unidos, não se tendo notícias da realização de trabalho sobre o tema de grande repercussão no Brasil.

KRUEGER (2003) defende a efetividade das reduções do tamanho da turma na proficiência, porque o tamanho da turma é o principal determinante dos gastos em educação e, assim sendo, o conhecimento de seus efeitos sobre a proficiência tem implicações diretas nas políticas públicas. Por outro lado, para HANUSHEK, (1998b), essas reduções não são efetivas para a melhoria da proficiência, uma vez que levariam a um excessivo aumento nos custos do ensino.

Em seus trabalhos, geralmente, Krueger e Hanushek adotam a metodologia de meta-análise para verificar suas hipóteses. O ponto de maior discordância entre eles reside na forma de operacionalização da meta-análise, o que estaria causando diferentes resultados sobre a efetividade da relação entre redução do tamanho da turma e da proficiência do aluno.

Essa discussão sobre o tamanho da turma e sua relação com o rendimento do aluno é muito complexa, porque, além das divergências citadas acima, ela pode apresentar interações com *tracking*<sup>2</sup> (efeito de pares).

Entretanto, esta tese não pretende trabalhar essas interações, uma vez que o efeito de pares em dados educacionais são difíceis de serem adequadamente estimados, em especial, devido à natureza simultânea das interações de pares, tornando a separação de impactos causais muito complexa (HANUSHEK *et al*, 2001). Ademais, CÉSAR & SOARES (2001) ressaltam que introduzir uma única medida de desempenho prévio ou de nível sócio-econômico médio captam pouco da complexidade do efeito de pares. Assim sendo, vários estudos da área foram produzidos somente com o efeito do tamanho da turma. Esses

---

<sup>2</sup> O *tracking* foi definido por ROSENBAUM (2002) como qualquer prática que cria classes homogêneas estratificadas com base no rendimento, por exemplo.

problemas são colocados no sentido de justificar a opção feita, neste estudo, de não se trabalhar com tais interações, mas somente com o efeito do tamanho da turma.

A primeira investigação em larga escala com o objetivo de verificar a efetividade das reduções do tamanho da classe na proficiência dos alunos foi o projeto realizado no Estado do Tennessee nos Estados Unidos durante a segunda metade da década de 80 (HANUSHEK, 2002a, 2003). Segundo este autor, tal projeto, denominado STAR (*Student/Teacher Achievement Ratio*), foi um experimento, em que os alunos foram aleatoriamente selecionados para estudarem em classes grandes (22 a 24 alunos) ou pequenas (14 a 16 alunos), sendo acompanhados do jardim de infância até a terceira série.

A análise do projeto STAR concluiu que alunos em classes menores têm um rendimento melhor no fim do jardim de infância, sendo este melhor rendimento mantido até a terceira série (HANUSHEK, 2002a, 2003). Porém, para o referido autor, o STAR teve implicações políticas relativamente limitadas, fornecendo, por um lado, o potencial impacto de grandes mudanças no tamanho da classe do jardim da infância, e por outro, mostrando que suas evidências não são generalizáveis para outras séries.

Por sua vez, KRUEGER (2003) apresenta uma re-análise dos dados do projeto STAR, sugerindo que os efeitos positivos da redução do tamanho da classe na proficiência dos alunos podem ser pequenos, não generalizáveis entre as séries e obscurecidos por equações mal especificadas ou por pequenas amostras. Entretanto, este autor ressalta que, mesmo esses efeitos sendo sutis, eles podem ser economicamente importantes.

No trabalho intitulado “*The failure of Input-based schooling policies*”, adotando meta-análise, HANUSHEK (2003) fez um sumário da produção escolar nos Estados Unidos, incluindo todas as publicações anteriores a 1995. Estas publicações forneciam estimativas dos efeitos dos recursos escolares no rendimento do aluno e indicavam se essas estimativas eram, ou não, significativas para a proficiência educacional. O sumário continha oitenta e nove estudos individuais, formando a base para a análise, a qual considerou as 376 estimativas separadas das funções de produção.

Então, como os estudos considerados mostravam se tais recursos eram, ou não, significativamente associados ao rendimento escolar, foram contadas as estimativas significativamente positivas e negativas e as não significativas. Esses resultados, em termos relativos, são encontrados na TAB. 2, que mostra a distribuição percentual dos efeitos estimados de recursos-chave no rendimento do aluno, baseando-se em 376 funções de produção ajustadas.



Contrariando a hipótese padrão das iniciativas políticas, para a qual cada um desses recursos deveria ter um efeito positivo no rendimento do estudante, os resultados mostram a grande incerteza de que apenas adicionar qualquer desses recursos mostrados na TAB.2 levaria a uma melhora no rendimento dos alunos (HANUSHEK, 2003).

TABELA 2- Distribuição percentual dos efeitos estimados de recursos-chave em rendimento educacional, baseando-se em 376 funções de produção ajustadas

Recursos	Número de estimativas	Estatisticamente significativas		Estatisticamente não significativas
		Positivas	Negativas	
<b>SALA</b>				
Razão professor-aluno	276	14%	14%	72%
Educação do professor	170	9%	5%	86%
Experiência do professor	206	29%	5%	66%
<b>GASTOS AGREGADOS</b>				
Salário do professor	118	20%	7%	73%
Gasto por aluno	163	27%	7%	66%
<b>OUTROS</b>				
Facilidades	91	9%	5%	86%
Administração	75	12%	5%	83%
Escore de teste do professor	41	37%	10%	53%

FONTE: HANUSHEK (2003)

A principal justificativa de KRUEGER (2003) para criticar os estudos de Hanushek, foi a adoção, em estudos de meta-análise, de um mesmo peso às estimativas. Dessa forma, Hanushek estaria reforçando as questões em torno da decisão do pesquisador ao selecionar estimativas de um particular estudo. Se os estudos, e não as estimativas, apresentassem o mesmo peso na meta-análise, haveria uma diminuição do impacto da decisão do pesquisador na seleção das estimativas (KRUEGER, 2003).

Apesar das críticas, Hanushek tem, freqüentemente, designado pesos para os estudos como a proporção do número de estimativas que se extraiu destes, em outras palavras, cada estimativa tem o mesmo peso. Ele continua a defender esta forma de operacionalização da meta-análise, argumentando que estudos dos quais extraiu mais estimativas são de alta qualidade (KRUEGER, 2002).

Porém, a definição implícita de HANUSHEK (2002b) de estudos de alta qualidade, que tendem a usar dados desagregados, é uma interpretação substantiva questionada por KRUEGER (2002), baseado na inspeção dos estudos que receberam os maiores pesos nas meta-análises de Hanushek.

Enfim, sobre a melhor forma de sumarizar quantitativamente os estudos KRUEGER (2002, p.86) afirma: *I think the least-manipulatable way to quantitatively summarize the studies in the literature is to give each study equal weight.*

Embora Hanushek e Krueger discordem sobre a forma de realização e as conclusões que podem ser retiradas de estudos de meta-análise, eles concordam com a importância de se levar em conta as especificações do modelo, pois as equações adotadas frequentemente apresentam pouca justificção teórica, dificultando identificar se as hipóteses de uma determinada abordagem são, ou não, razoáveis (TOOD & WOLPIN, 2003).

Dado o exposto, pode-se dizer que não existe um consenso sobre o efeito do tamanho da classe na proficiência dos alunos (LAZEAR, 2001; TOOD & WOLPIN, 2003; HANUSHEK, 2003). Neste sentido, LAZEAR (2001) ressalta que a variedade de resultados dos efeitos do tamanho da classe na proficiência deriva da dificuldade de determinar tais efeitos com a utilização de dados sobre o tamanho da classe que sejam verdadeiramente exógenos ao processo escolar, já que, na maioria das vezes, as escolas escolhem o tamanho da turma.

Considerando que o tamanho da classe pode ser escolhido, LAZEAR (2001) afirma que quanto maior o tamanho da classe, menor a atenção despendida aos estudantes. Então, o autor assume que interrupções requerem dos professores uma suspensão do ensino, criando uma externalidade negativa, reduzindo a quantidade de aprendizado em toda a classe. Assim, seria suficiente utilizar poucos professores e uma alta razão estudante/professor, quando os alunos são bem comportados, ou não interrompem a aula. Portanto, na visão de LAZEAR (2001), o tamanho ótimo da classe pode ser maior para estudantes que são bem comportados, porque eles interrompem menos a aula. Nessa situação, as relações entre tamanho da classe e rendimento escolar poderiam ser positivas.

Nessas circunstâncias de endogeneidade (a escola escolhe o tamanho da turma), as escolas teriam incentivo para destinar estudantes em desvantagens ou mal comportados para classes menores, o que pode levar a uma associação espúria entre tamanho da classe e rendimento. LAZEAR (2001) acredita que a redução do tamanho da classe é mais uma questão de alunos em desvantagens ou com necessidades especiais, para os quais a redução do tamanho da classe pode fornecer melhores resultados.

#### 2.4.2-Variáveis explicativas relacionadas ao processo escolar

Conforme dito anteriormente, embora seja reconhecida a importância do processo na aprendizagem por vários autores LEE & BRYK (1989), WILLMS (1992, 2000), WILLMS & SOMERS (1999), HANUSHEK (2002a, 2003), dentre outros, não se pode negar a dificuldade em se obter dados dessa natureza, o que pode, de certa forma, dificultar o ajuste de modelos do tipo insumo-processo-produto.

O relatório SAEB (2004) expressa uma das suas preocupações e ao mesmo tempo uma das possibilidades de se chegar a um processo escolar que conduza a resultados escolares mais bem sucedidos, especialmente por meio de incentivo e implantação de conselhos escolares efetivamente atuantes:

*“Outros aspectos destacados pelas avaliações têm se convertido em objeto de preocupação e fundamentam experiências bem-sucedidas de reestruturação do sistema gerencial das escolas, com o reforço da autonomia escolar e o incentivo à participação da comunidade na escola; de reorganização escolar e de ampliação da autonomia pedagógica, entre outros. Um bom começo para efetivas mudanças no padrão de participação da comunidade é, por exemplo, o incentivo e a implantação dos conselhos escolares que devem atuar de forma ativa e autônoma. Desse modo, qualquer mudança significativa dar-se-á por meio da articulação solidária dos entes federativos, dos poderes constituídos e dos principais atores sociais das unidades escolares.” (SAEB, 2004, p.47)*

Segundo FERRÃO *et al* (2001), a modelagem hierárquica com os dados do SAEB-99 apresenta as seguintes variáveis escolares como significativas: escolas com melhor infraestrutura e escolas com ensino seriado ao invés de ciclos. Contudo, as variáveis utilizadas como critérios para a criação de turmas, processo de recuperação de notas, aquelas relacionadas ao conselho e ao turno integral não apresentaram relação estatisticamente significativa com o rendimento escolar.

Em meio a resultados e afirmações não consensuais, a princípio, acredita-se ser mais sensato considerar a multiplicidade de fatores associados ao rendimento escolar e, portanto, à qualidade do ensino. Entretanto, uma vez que os objetivos principais do trabalho estão direcionados à estimação dos efeitos das variáveis escolares na proficiência e do efeito-escola, pretende-se enfatizar os fatores escolares associados ao rendimento, estimando, como controle o efeito dos recursos dos alunos e de suas famílias.

### **3- METODOLOGIA**

Este capítulo mostra, inicialmente, um histórico sobre o banco de dados e uma breve descrição do processo de amostragem. A seguir, são apresentados os dois métodos utilizados nessa tese para o alcance do objetivo proposto e a verificação das hipóteses formuladas. O primeiro deles é a Teoria da Resposta ao Item que foi adotada, especialmente, para a estimação da variável resposta, permitindo que os desempenhos fossem diretamente comparáveis. O segundo método diz respeito ao ajuste de modelos hierárquicos longitudinais.

#### **3.1- Banco de dados: histórico e breve descrição da amostragem**

Por meio do banco de dados coletado, inicialmente, tinha-se o objetivo de avaliar o impacto das intervenções do Plano de Desenvolvimento da Escola (PDE) sobre os resultados dos sistemas educacionais beneficiários do programa *Fundescola*. Então, o Banco Mundial concedeu um empréstimo (BRA/96/026) ao Ministério da Educação e Cultura (MEC) para que este impacto fosse avaliado. Sendo assim, o *Fundescola* procurou o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), que propôs a realização de uma pesquisa longitudinal para medir o impacto do PDE no rendimento dos alunos.

A pesquisa longitudinal foi denominada “Avaliação do desempenho: fatores associados”, com duração prevista de cinco anos e a realização de seis aplicações, começando em abril de 1999 e terminando em novembro de 2003. Essa pesquisa faria o acompanhamento de alunos desde a 4ª série em 1999 até 2003, quando estes deveriam estar cursando a 8ª série, por meio da aplicação de provas de matemática e português.

Entretanto, muitos alunos se perderam no meio do processo devido à reprovação, evasão, transferência, morte, dentre outros. Por outro lado, durante o processo, outros alunos foram acrescentados à pesquisa, porque todos os alunos das turmas pesquisadas foram avaliados, em virtude da preocupação em minimizar os efeitos da aplicação de provas sobre as rotinas das escolas.

Além dos resultados das provas de português e matemática, o banco de dados fornece informações sócio-econômicas e demográficas sobre o aluno, o professor, o diretor e sobre a infra-estrutura da escola. Sobre o aluno, existe ainda uma ficha escolar, contendo a média no final da série cursada, o número de faltas, o resultado (aprovado ou não).

O questionário sócio-econômico do aluno contém informações sobre fatores individuais e familiares e sobre a relação do aluno com a escola, dentre outras. Basicamente, o questionário do professor contém dados sobre os fatores sócio-econômicos individuais e familiares e informações sobre sua relação com a escola e o aluno. O questionário do diretor é muito semelhante ao do professor, com ênfase nas questões da escola.

Em 2001, o objetivo do projeto foi alterado, porque grande parte das escolas passou a participar do programa PDE. A partir daí, o objetivo principal da pesquisa é estimar os fatores determinantes do desempenho escolar, beneficiando-se da experiência inédita, no Brasil, em pesquisas longitudinais educacionais e da possibilidade de resultados mais confiáveis, uma vez que o aluno é testado em vários períodos.

De início, foi estabelecido que o estudo deveria ser realizado em dois estados de cada uma das regiões brasileiras atendidas pelo programa *Fundescola* (Norte, Nordeste e Centro-Oeste). Embora o programa ainda não houvesse iniciado sua implementação no Nordeste, ficou decidido que essa região seria incluída na pesquisa e seus resultados em abril de 1999 seriam considerados como a referência anterior à entrada no PDE. Assim, os estados escolhidos para representar cada uma das três regiões foram: na Região Norte, os estados de Rondônia (RO) e Pará (PA); na Região Nordeste, Pernambuco (PE) e Sergipe (SE) e na Região Centro-oeste, Mato Grosso do Sul (MS) e Goiás (GO).

Para o estudo dos impactos do PDE no rendimento dos alunos, ANDRADE (1999) definiu como população alvo o conjunto das escolas do PDE com todas as séries do ensino fundamental no período diurno, com pelo menos 200 alunos, localizadas nas microrregiões das capitais e pertencentes às dependências administrativas estaduais ou municipais.

A amostragem foi realizada em estágios. O número total de escolas selecionadas foi fundamentado nas matrículas da 4ª série e na estimativa da média de matrículas nessas séries, baseando-se nos dados do Censo Escolar. Desta forma, foi obtido um número de 160 escolas para a amostra. Para a definição do número total de alunos avaliados, segundo ANDRADE (1999), foram considerados a eventual perda de estudantes ao longo das séries e os custos para a realização de todas as etapas da pesquisa, obtendo-se um total desejado de 12.000 alunos na 4ª série. Maiores detalhes sobre o processo de amostragem podem ser vistos em ANDRADE (1999).

Para a realização do estudo proposto nesta tese, pretende-se utilizar a amostra, considerando apenas os alunos que realizaram mais de duas provas, uma vez que, segundo WILLETT (1997), somente com três ou mais testes, é possível estimar de forma confiável as mudanças individuais.

## **3.2- Principais características da Teoria Clássica do Teste (TCT) e a da Teoria da Resposta ao Item (TRI) na avaliação de testes educacionais**

### **3.2.1- Uma introdução à Teoria Clássica do Teste**

No campo dos testes educacionais, a Teoria Clássica do Teste (TCT) vem perdendo espaço para a Teoria da Resposta ao Item (TRI) que, desde as últimas décadas, tem se destacado como a técnica predominante na avaliação destes testes. Maiores detalhes sobre a Teoria Clássica do Teste podem ser vistos em LORD (1980) e em WEISS & YOES (1991). Contudo, é importante apresentar algumas características da Teoria Clássica, para que se possa notar as principais vantagens em se adotar a Teoria da Resposta ao Item na estimação das habilidades dos alunos.

A Teoria Clássica do Teste tem como medida da proficiência (habilidade) dos alunos a soma dos itens respondidos corretamente no teste (escores totais). Esta soma é chamada número observado de itens respondidos corretamente e, segundo a teoria clássica, representa a proficiência do aluno no teste. Desta forma, os resultados encontrados dependerão, diretamente, de um conjunto particular de itens que compõem o teste, não permitindo a comparação entre examinandos que não foram submetidos às mesmas provas (ANDRADE, 2001; ANDRADE, TAVARES & VALLE, 2000), conforme detalhado nos parágrafos a seguir.

Os parâmetros de interesse na Teoria Clássica do Teste são, essencialmente, a dificuldade e a discriminação do item. As principais limitações da Teoria Clássica estão, em geral, associadas à definição destes parâmetros (WEISS & YOES, 1991), conforme explicado a seguir. Contudo, merece ainda ser ressaltada outra importante limitação da Teoria Clássica, a consideração da prova como elemento central (ANDRADE, TAVARES & VALLE, 2000; HAMBLETON, 1993).

Segundo WEISS & YOES (1991), a Teoria Clássica do Teste é baseada em um modelo linear que estabelece para cada indivíduo uma habilidade - quantidade não observável, chamada escore verdadeiro - estimada diretamente pelo número de itens observado respondidos corretamente. Assim, este modelo é, em geral, representado por:

$$Y = X + E$$

em que  $Y$  é o escore observado;  $X$  é o escore verdadeiro e  $E$  corresponde ao erro.

De acordo com a Teoria Clássica do Teste, a dificuldade do item é, comumente, medida por meio da proporção de respostas corretas ao item ( $p$ ). Por sua vez, a discriminação do item, que mede a capacidade de um item discriminar estudantes com habilidades alta e baixa, pode ser obtida através do coeficiente de correlação bisserial<sup>3</sup>, que expressa a correlação entre um item e o teste, como mostra a seguinte equação (HENRYSSON, 1971):

$$r_{bis} = \frac{M_R - M_W}{S_T} * \frac{p(1-p)}{y} = \frac{M_R - M_T}{S_T} * \frac{p}{y}$$

em que:

$M_R$ = média dos escores obtidos pelos examinandos que acertaram o item;

$M_W$ = média dos escores obtidos pelos examinandos que erraram o item;

$M_T$ = média dos escores obtidos por todos os candidatos;

$S_T$ = desvio-padrão dos pontos obtidos para todos os estudantes;

$p$ =proporção de respostas corretas;

$y$ =valor da função normal padronizada o qual divide a área sob a curva nas proporções  $p$  e  $(1-p)$ .

Diante do exposto, os parâmetros de dificuldade e discriminação do item, definidos conforme a Teoria Clássica do Teste, são dependentes da amostra de estudantes utilizada em um teste, assim como a habilidade dos examinandos depende dos itens que compõem o teste.

Conforme exemplificam WEISS & YOES (1991), se um grupo de itens for administrado a um grupo de examinandos com alta habilidade, os parâmetros de dificuldade e de discriminação dos itens, provavelmente, serão bem diferentes do que se os mesmos itens fossem administrados a um grupo de examinandos com habilidade baixa ou moderada.

Assim, a dificuldade do item será maior, quando a amostra de examinandos tiver habilidade acima da média do que quando a amostra de examinandos apresentar habilidade abaixo da média, enquanto o parâmetro de discriminação tende a ser maior em amostra de examinandos heterogênea do que em amostra homogênea (HAMBLETON, 1993). Por sua vez, a habilidade, sendo baseada no número de itens respondidos corretamente, depende das dificuldades dos itens que compõem o teste (WEISS & YOES,1991).

Assim, em outras palavras, sobre as limitações da aplicação da Teoria Clássica na avaliação de testes educacionais, ANDRADE (2001) afirma que como esta teoria toma por

---

<sup>3</sup> O coeficiente de correlação bisserial é utilizado para medir a correlação entre uma variável contínua e outra dicotômica.

base os resultados expressos por seus escores “brutos”, comparações entre o desempenho dos alunos são possíveis apenas quando todos eles são submetidos à mesma prova. Ademais, estes resultados serão dependentes do grupo de itens que compõem a prova, estando suas interpretações associadas à prova como um todo.

### 3.2.2- Introdução à Teoria da Resposta ao Item

De acordo com BARTHOLOMEW *et al* (2002) existem quatro tipos de análise fatorial, conforme pode ser visualizado no QUADRO 1.

QUADRO 1- Tipos de análise fatorial

Variável Latente	Variável Observada	
	Medida (intervalar ou razão)	Catégorica (nominal ou ordinal)
Medida (intervalar ou razão)	Análise Fatorial	Análise de Traço Latente-TRI
Catégorica (nominal ou ordinal)	Análise de Perfil Latente	Análise de Classe Latente

FONTE: BARTHOLOMEW *et al* (2002)

Sendo assim, a partir dessa classificação proposta por BARTHOLOMEW *et al* (2002), a análise de traço latente, em geral, é realizada com a utilização da Teoria da Resposta ao Item (TRI), tendo variáveis binárias como observadas e buscando estimar um fator latente quantitativo.

Os modelos da Teoria da Resposta ao Item em Educação são, geralmente, ajustados pelo programa *BILOG-MG*, que não apresentam a limitação do programa *LAMI (Latent Model Interface)*, aceitando muito mais de 3000 casos. Contudo, não existe no *BILOG-MG* um teste da qualidade do ajuste da prova como um todo. Assim sendo, somente para cada item é possível verificar o ajuste de seu modelo, por meio do teste da razão de verossimilhança. Considera-se a qualidade do ajuste do teste, especialmente, por meio da retirada de itens que não se enquadram dentro de certos critérios, conforme pode ser visto neste capítulo.

A Teoria da Resposta ao Item, que teve início com os trabalhos de Lord e de Rasch, nos anos 50 e 60, veio complementar a Teoria Clássica na medida em que supera suas



principais limitações, em especial, a dependência dos parâmetros de item em relação à amostra de examinandos e a dependência das habilidades em relação aos itens que compõem o teste (HAMBLETON, 1993).

Vale lembrar que o escore da Teoria Clássica é simplesmente a soma dos itens que o aluno acertou, enquanto o escore da Teoria da Resposta ao Item leva em conta, além do número de itens respondidos corretamente pelo aluno, as características dos itens que ele acertou, garantindo, desta forma, uma medida mais apurada da habilidade ou proficiência do aluno.

A Teoria da Resposta ao Item é definida como um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma resposta certa a um item em função dos parâmetros de item e da habilidade dos respondentes (ANDRADE, TAVARES & VALLE, 2000). De acordo com os mesmos autores, essa relação é expressa pela Curva Característica do Item (CCI), que descreve a relação entre a habilidade e a probabilidade de acerto. Espera-se que quanto maior a habilidade, maior a probabilidade de acerto ao item.

De acordo com a Teoria da Resposta ao Item, os parâmetros de item são independentes da amostra de examinandos utilizada para estimá-los. Segundo afirmam HAMBLETON, SWAMINATHAN & ROGERS (1991), de acordo com a TRI, quando o modelo se ajusta aos dados, a mesma Curva Característica do Item é obtida para um item do teste, independente da distribuição de habilidade do grupo de examinandos usado para estimar os parâmetros de item.

Para um melhor entendimento desta propriedade, os referidos autores fazem uma analogia com o modelo de regressão linear. Considerando uma regressão linear simples, quando o modelo se ajusta aos dados, a mesma curva é obtida dentro de qualquer intervalo restrito para a variável X (ou seja, em qualquer sub-população desta variável explicativa), significando que a inclinação e o intercepto da linha serão os mesmos em qualquer sub-população de X.

Outra vantagem presente nos modelos de resposta ao item é a independência das habilidades em relação aos itens que compõem o teste. De acordo com a Teoria da Resposta ao Item, a habilidade (ou traço latente) do aluno não pode ser mensurada diretamente, sendo modelada em função das características dos itens que compõem o teste. Entretanto, a habilidade não depende de uma amostra particular de itens que compõem o teste, visto que estudantes terão a mesma habilidade entre as várias amostras possíveis de itens que podem compor um determinado teste e, portanto, os estudantes poderão ser comparados mesmo

quando eles não forem submetidos a grupos idênticos de itens no teste (HAMBLETON, 1993).

Além das duas vantagens, anteriormente citadas, obtidas com a utilização da TRI, uma das principais características da TRI, que se configura como uma de suas grandes vantagens sobre a TCT, é considerar como elementos centrais os itens, e não a prova como um todo, conforme afirmam ANDRADE, TAVARES & VALLE (2000, p.3):

*“ Uma das grandes vantagens da TRI sobre a Teoria Clássica é que ela permite a comparação entre populações, desde que submetidas a provas que tenham alguns itens comuns, ou ainda, a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente diferentes. Isto porque uma das principais características da TRI é que ela tem como elementos centrais os itens, e não a prova como um todo.”*

### **3.2.2.1- Os modelos mais comuns para uma população (ou grupo)**

Inicialmente, os modelos da TRI foram baseados na distribuição normal acumulada (ogiva normal), sendo mais tarde substituídos pela função logística, matematicamente mais conveniente para trabalhar, já que não envolve integração. Os modelos de resposta ao item variam em complexidade, dependendo do número de parâmetros considerados. Na prática, os modelos logísticos de um, dois e três parâmetros para itens dicotômicos são os modelos de resposta ao item mais comuns e consideram, respectivamente:

1. Apenas o parâmetro da habilidade necessária para acertar o item  $i$  ( $b_i$ ), uma vez que a discriminação ( $a$ ) de todos os itens no teste é a mesma;
2. Os parâmetros  $b_i$  e  $a_i$ ;
3. Os parâmetros  $b_i$  e  $a_i$  e a probabilidade de acerto ao acaso ( $c$ );

Dos três modelos mais comuns, o modelo logístico de três parâmetros é atualmente o mais utilizado, visto que as hipóteses de ausência de acerto casual entre examinandos de baixa habilidade e de igualdade de discriminação para todos os itens do teste, na prática, não são, em geral, adequadas a dados de testes de múltipla escolha (HAMBLETON, 1993).

O modelo logístico de três parâmetros pode ser descrito conforme a FIG. 4 e a função abaixo:

$$P(X_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}$$

$i=1,2,\dots,I$  e  $j=1,2,\dots,n$

em que:

$P(X_{ij}=1|\theta_j)$  é a probabilidade de um examinando  $j$  com habilidade  $\theta_j$  responder corretamente ao item  $i$ , também chamada de Curva Característica do Item (CCI) ou de Função de Resposta ao Item (FRI);

$X_{ij}$  é uma variável dicotômica que assume o valor 1, quando o examinando  $j$  responde corretamente ao item  $i$  e o valor 0, quando o examinando  $j$  não responde corretamente ao item;

$\theta_j$  é a habilidade do  $j$ -ésimo examinando, a qual não é observável;

$b_i$  é a habilidade necessária para o examinando responder o item  $i$  corretamente com probabilidade  $(1+c_i)/2$ . Em outras palavras, representa o ponto na escala de habilidade em que o examinando tem probabilidade  $(1+c_i)/2$  de responder ao item  $i$  corretamente. Dessa forma, pode-se dizer que quanto maior a habilidade necessária para responder o item  $i$  corretamente ( $b_i$ ), mais difícil é o item.

$a_i$  é o parâmetro de discriminação do item  $i$ , com valor proporcional à inclinação da Curva Característica do Item no ponto  $b_i$ . A inclinação da curva em  $b_i$  é igual a  $0,425 a_i (1+c_i)$ . Estando a habilidade normalmente distribuída, com média zero e desvio padrão igual a um, os valores do parâmetro  $a_i$  geralmente variam entre 0 e 2. Valores baixos de discriminação indicam que o item não consegue separar os alunos em grupos de habilidades distintos. Por outro lado, valores muito altos implicam na ocorrência de dois grupos: aqueles que possuem habilidade abaixo do parâmetro  $b_i$  e os que possuem acima (ANDRADE, 2001).

$c_i$  representa a probabilidade de alunos, com baixa habilidade, responderem corretamente ao item  $i$ , também denominada probabilidade de acerto casual. Quando a alternativa de resposta correta é marcada casualmente, o valor de  $c_i$  deve ser igual a  $1/A$ , em que  $A$  representa o número de alternativas de múltipla escolha do item  $i$ .

$D$  é um fator de escala que assume o valor 1,7, quando se deseja que a função logística forneça resultados semelhantes ao da função de ogiva normal.

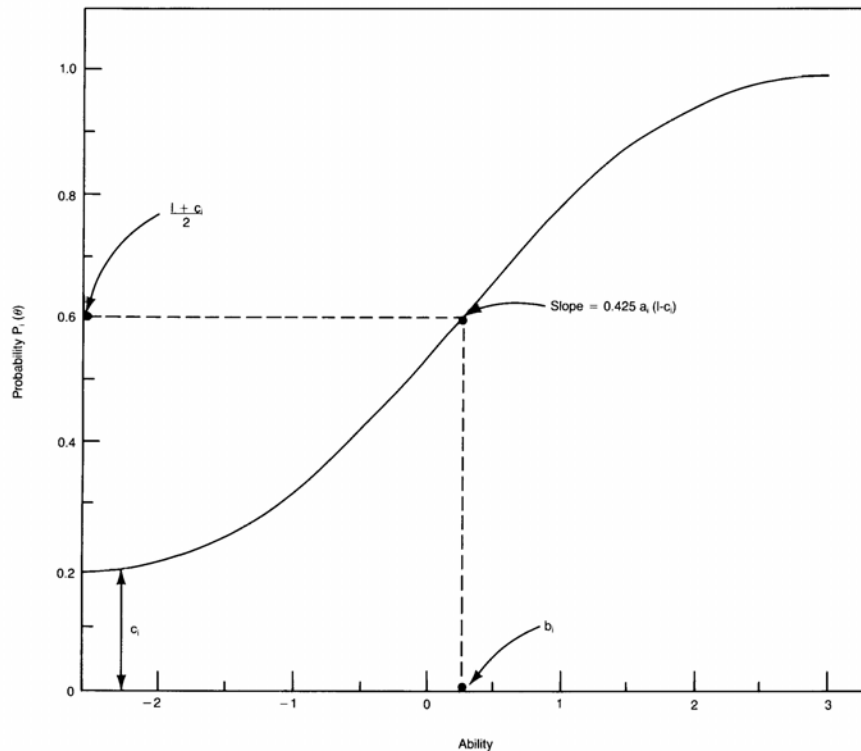


FIGURA 4- Curva Característica do Item para um modelo logístico de três parâmetros.  
 FONTE: HAMBLETON (1993)

### 3.2.2.2- Pressupostos

Os modelos de resposta ao item pressupõem que todos os itens medem uma única habilidade. Apesar do desempenho humano ser multi-determinado, uma vez que mais de uma habilidade participa da execução de qualquer tarefa, para satisfazer o postulado da unidimensionalidade do teste, é suficiente admitir que haja um fator dominante responsável pelo desempenho de todos os itens do teste. Este fator dominante refere-se à habilidade supostamente mensurada pelo teste (LORD, 1980; HAMBLETON, SWAMINATHAN & ROGERS, 1991; HAMBLETON, 1993).

Segundo esses mesmos autores, uma outra suposição destes modelos é a independência local. De acordo com esta suposição, para uma dada habilidade, as respostas de um examinando aos itens do teste são independentes, ou seja, o aluno não aprende com o teste. Formalmente, para um examinando com uma dada habilidade, a probabilidade de um padrão de resposta dicotômico  $x_1, \dots, x_n$  a um conjunto de  $n$  itens é igual aos produtos das probabilidades de resposta do examinando  $j$  a cada item  $i$  individual, conforme mostra a equação a seguir:

$$P_j = (X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | \theta) = \\ = P_j(X_1 = x_1 | \theta)P_j(X_2 = x_2 | \theta) \dots P_j(X_n = x_n | \theta)$$

Desta forma, o desempenho de um examinando em um item não pode afetar sua resposta a qualquer outro item do teste. Assim sendo, a independência local especifica que somente a habilidade do examinando e as características de itens que compõem o teste influenciam o desempenho do examinando.

Entretanto, se as respostas não forem estatisticamente independentes a um dado nível de habilidade (se não houver independência local), isto implicará que alguns examinandos neste nível de habilidade terão escores de testes mais altos que outros examinandos no mesmo nível de habilidade (HAMBLETON, 1993). Consequentemente, mais de uma habilidade será necessária para explicar o desempenho daqueles examinandos, não sendo possível aos itens medir apenas uma dimensão.

Assim sendo, a unidimensionalidade implica a independência local (LORD, 1980; HAMBLETON, SWAMINATHAN & ROGERS, 1991; HAMBLETON, 1993; ANDRADE, 2001). Dito de outra forma, supondo que um grupo de itens meça somente uma habilidade, então, para examinandos com um dado nível desta habilidade, as respostas dos itens são estatisticamente independentes.

### 3.2.2.3- A escala da habilidade

De acordo com a Teoria da Resposta ao Item, a habilidade pode teoricamente variar de  $-\infty$  a  $+\infty$ , diferentemente do que acontece na Teoria Clássica, em que o aluno recebe um escore entre 0 e o número de questões respondidas corretamente.

Como a habilidade na TRI pode assumir qualquer valor real, é necessário estabelecer uma origem e uma unidade de medida para a definição de uma escala (HAMBLETON, SWAMINATHAN & ROGERS, 1991; HAMBLETON, 1993; ANDRADE, 2001; ANDRADE, TAVARES & VALLE, 2000). Sendo a origem e a escala indeterminados, qualquer transformação linear levaria ao mesmo resultado. Desta forma, arbitra-se uma origem e uma escala. Para a origem e a unidade de medida, que representam respectivamente a média e o desvio-padrão das habilidades, geralmente a escala utilizada para a população é  $(0,1)$ .

Apesar da escala  $(0,1)$  ser a mais utilizada, na prática, não faz diferença se outra escala for estabelecida, pois o que importa é a igualdade das probabilidades especificadas pelo

modelo-  $P(X_i=1|\theta) = P(X_i=1|\theta^*)$ . Desta forma, HAMBLETON (1993) e ANDRADE, TAVARES & VALLE (2000) afirmam ser fácil mostrar que transformações lineares de  $\theta$  são aceitáveis, fornecendo ajustes lineares correspondentes aos parâmetros de itens no modelo:

$$\begin{aligned}\theta^* &= \sigma^* \times \theta + \mu^* \\ b^* &= \sigma^* \times b + \mu^* \\ a^* &= a / \sigma^* \\ P(X_i = 1 | \theta) &= P(X_i = 1 | \theta^*) \\ c^* &= c\end{aligned}$$

em que:

$(\mu, \sigma)$  representa a escala anteriormente utilizada;

\* indica valores de acordo com a nova escala.

Por exemplo, na escala (0,1) um aluno com habilidade 1,5 está a 1,5 desvios- padrão acima da média. Se a escala utilizada para o grupo ao qual pertence esse aluno fosse (50,10), o aluno teria uma habilidade de 65 e, conseqüentemente, também estaria a 1,5 desvios- padrão acima da média.

Para um determinado item estimado na escala (0,1) com o parâmetro  $a=0,8$  e o parâmetro  $b=-0,2$ , os valores dos parâmetros  $a$  e  $b$  correspondentes na escala (50,10) são, respectivamente,  $0,08=0,8/10$  e  $48=10(-0,2)+50$ . Se o aluno tiver habilidade igual a um, medida na escala (0,1), sua habilidade será transformada em  $60=10(1)+50$ , quando a escala adotada mudar para (50,10).

Como a probabilidade de um aluno responder corretamente a um determinado item é sempre a mesma, independente da escala utilizada para medir a sua habilidade, é necessário conhecer a escala na qual os itens foram estimados, para que se possa analisá-los adequadamente (ANDRADE, 2001; ANDRADE, TAVARES & VALLE, 2000). Estes autores também ressaltam que o parâmetro de acerto casual também é independente da escala de habilidade utilizada, porque ele estima a probabilidade de acerto para alunos com baixa habilidade, qualquer que seja a escala de medida adotada.

#### **3.2.2.4- A função de informação**

A função de informação do item tem grande importância para o desenvolvimento de testes e para a avaliação da qualidade de itens, uma vez que ela informa a contribuição dos itens na estimação de cada ponto da escala de habilidade.

Segundo VALLE (1999) e ANDRADE, TAVARES & VALLE (2000), esta contribuição depende da discriminação do item (quanto maior o parâmetro de discriminação, maior será a informação), do acerto casual (quanto maior o acerto casual, menor será a informação) e a da dificuldade do item (quanto mais próxima a dificuldade do item estiver da habilidade, maior será a informação). Estas relações, entre a informação e os parâmetros de um determinado item, podem ser ilustradas de acordo com a FIG. 5, que mostra as curvas características e de informação de quatro itens diferentes.

É importante ressaltar que um item apresenta maior discriminação na região em que a inclinação da curva característica do item for máxima, podendo, contudo, fornecer informação bastante limitada, em porções significativas da distribuição de habilidade, onde a inclinação da curva for baixa (HAMBLETON, 1993).

Para o modelo logístico de três parâmetros, segundo HAMBLETON (1993), a curva de informação do item é dada pela seguinte equação:

$$I_i(\theta) = \frac{2,89 a_i^2 (1 - c_i)}{\left[ c_i + e^{1,7 a_i (\theta - b_i)} \right] \left[ 1 + e^{-1,7 a_i (\theta - b_i)} \right]^2}$$

$i = 1, \dots, n$

A função de informação do teste é dada, simplesmente, pelo somatório das funções de informação dos itens no teste:

$$I(\theta) = \sum_i^n I_i(\theta)$$

A informação fornecida pelo teste, dado um nível de habilidade  $\theta_0$ , está inversamente relacionada com o erro-padrão de estimação:

$$SE(\theta_0) = \frac{1}{\sqrt{I(\theta_0)}}$$

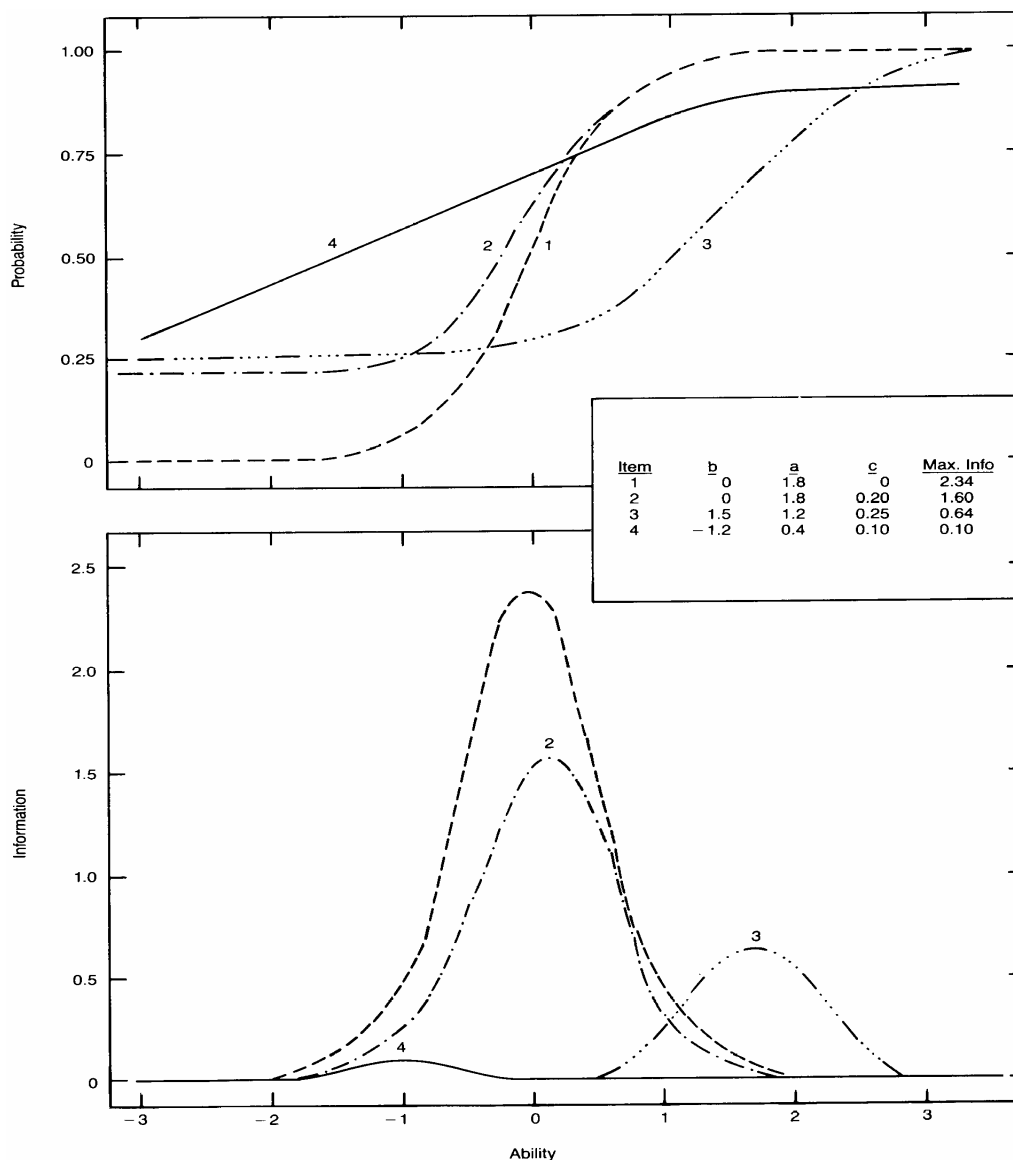


FIGURA 5- Curva Característica de quatro itens e suas funções de informação correspondentes  
 FONTE: HAMBLETON (1993)

### 3.2.2.5- Determinação da qualidade do item

Quanto à definição da qualidade de um item, não existe um critério único. Em geral, para avaliar a qualidade dos itens os pesquisadores consideram, principalmente, os valores referentes às estimativas dos parâmetros  $a_i$  e  $b_i$  e os erros-padrão (EP) destas estimativas.

De acordo com ANDRADE, TAVARES & VALLE (2000), um item pode ser bem classificado, quando ele apresenta, em especial, o parâmetro de discriminação superior a 1 (considerando um modelo em que  $D=1$ ). Contudo, o parâmetro de discriminação não deve ser muito alto, por exemplo, não deve assumir um valor próximo a 2, sob pena de conseguir



discriminar, somente para um grupo restrito de examinandos, quem tem alta ou baixa probabilidade de acertar um item.

Por sua vez, a habilidade necessária para acertar o item  $i$  ( $b_i$ ), tendo sua distribuição padronizada em uma normal com média zero e desvio padrão um, deve variar, preferencialmente, no intervalo  $[-2, 2]$ . Desta forma, os itens devem contemplar parte considerável da escala de habilidade, indo, portanto, ao encontro do que, em geral, é esperado de um teste.

Na prática, os critérios para avaliar a qualidade do itens acabam sendo mais flexíveis. Desta forma, os pesquisadores da área, em geral, consideram os seguintes critérios para a retirada de itens do processo de estimação:

- a) correlação bisserial<sup>4</sup> negativa;
- b) item com 100% de acerto ou de erro;
- c) parâmetro de discriminação menor do que 0,4 (BAKER, 2001);
- d) erros-padrão referentes aos parâmetros  $a_i$  e  $b_i$  maiores que 0,3 (BAKER, 2001).

### 3.2.2.6- Modelo para mais de uma população (ou mais de um grupo)

Os modelos comumente utilizados para mais de uma população (ou grupo) são generalizações dos modelos logísticos unidimensionais de um, dois ou três parâmetros desenvolvidas por ZIMOWSKI, MURAKI, MISLEVY & BOCK (1996) e implementadas no software BILOG-MG<sup>5</sup> (TOIT, 2003).

Para o caso do modelo logístico de três parâmetros, a equação é dada por:

$$P(X_{ijk} = 1 | \theta_{jk}) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_{jk} - b_i)}}$$

$$i=1,2,\dots,I, \quad j=1,2,\dots,n_k \text{ e } k=1,2,\dots,K$$

em que:

---

<sup>4</sup> O coeficiente de correlação bisserial é utilizado para medir a correlação entre uma variável contínua e outra dicotômica.

<sup>5</sup> Este programa permite a estimação conjunta dos itens de todos os grupos, considerando a habilidade de todas as populações numa mesma escala, aquela adotada como referência. Desta forma, todos os escores dos alunos são estimados, podendo ser diretamente comparados.

$X_{ijk}$  é uma variável dicotômica assumindo o valor 1, quando o aluno  $j$  da população  $k$  responde corretamente ao item  $i$ , e o valor 0, quando o aluno não responde corretamente ao item  $i$ ;

$\theta_{jk}$  é a habilidade do  $j$ -ésimo aluno da população  $k$ ;

$P(X_{ijk}=1|\theta_{jk})$  é a probabilidade de um aluno  $j$  da população  $k$ , com habilidade  $\theta_{jk}$  responder corretamente ao item  $i$ ;

Os demais parâmetros foram anteriormente descritos.

Geralmente, alunos pertencentes a diferentes populações não são todos submetidos aos mesmos itens. Contudo, para que se possa comparar essas populações, é necessário que haja alguns itens comuns (ou seja, os mesmos itens) entre os testes aplicados a elas. Desta forma,  $I$  corresponde ao número total de itens distintos apresentados às populações.

Com a implementação computacional deste modelo, é possível comparar alunos de grupos distintos, submetidos a testes diferentes com itens comuns, de uma forma mais eficiente pelo processo de equalização a priori, diminuindo possíveis erros de modelagem advindos da adoção de metodologias anteriormente utilizadas (ANDRADE, TAVARES & VALLE, 2000), conforme pode ser visto, a seguir, na seção de processos de equalização.

### **3.3- Os principais processos de equalização de acordo com a Teoria da Resposta ao Item**

#### **3.3.1- Definição de equalização e de sua importância em estudos longitudinais**

De início, é importante definir o termo equalizar como o ato de tornar comparável ou de equiparar. No caso da TRI, equalizar significa estimar, na mesma métrica, parâmetros de itens pertencentes a provas diferentes ou habilidades de alunos de grupos distintos. Em outras palavras, equalizar é estimar parâmetros de itens e habilidades numa escala comum, tornando os itens e as habilidades comparáveis (HAMBLETON, SWAMINATHAN & ROGERS, 1991; KOLEN & BRENNAN, 1995; ANDRADE, TAVARES & VALLE, 2000).

Quando diferentes testes são estimados separadamente<sup>6</sup>, não se pode garantir que os parâmetros dos itens destes testes estejam na mesma escala, porque a posição da escala

---

<sup>6</sup> Em geral, assumindo como verdadeiro o modelo proposto e partindo das respostas dadas pelos alunos de um ou mais grupos, estima-se os parâmetros de itens e/ou habilidades por meio do método de máxima verossimilhança ou de métodos bayesianos (ANDRADE, 2001). Ambos os métodos exigem procedimentos iterativos que

depende do grupo de respondentes considerado. Sendo assim, esta questão torna-se especialmente importante em estudos longitudinais, porque deles se espera um nível mais elevado de habilidade para os alunos das últimas séries, justificando, desta forma, a importância de se equalizar as proficiências estimadas em estudos como estes.

### **3.3.2- A equalização horizontal (via população) e a equalização vertical (via itens comuns)**

De uma forma sintética, pode-se dizer que existem dois tipos de equalização: a equalização via população, também conhecida por equalização horizontal, e a equalização via itens comuns, a denominada equalização vertical (WEISS & YOES, 1991; ANDRADE, TAVARES & VALLE, 2000).

A equalização via população, ou equalização horizontal, envolve equalizar escores obtidos em testes com aproximadamente o mesmo nível de dificuldade. Por exemplo, quando mais de uma versão do teste é aplicada a alunos de um mesmo nível, ou seja, de uma mesma série. Neste tipo de equalização, quando um único grupo de estudantes é submetido a provas distintas, para se garantir que os parâmetros de itens estão na mesma escala, basta estimar conjuntamente estes itens.

Por outro lado, a equalização via itens comuns, ou a equalização vertical, permite que se coloquem escores de testes de diferentes níveis de dificuldade em uma mesma escala. Por exemplo, escores obtidos por alunos de diferentes níveis educacionais, respondendo a testes diferentes, mas com a presença de itens comuns. Neste tipo de equalização, os itens comuns servem de ligação entre as populações envolvidas, garantindo que seus parâmetros estejam numa mesma escala. Assim sendo, a equalização vertical é o processo adequado para o desenvolvimento de estudo longitudinal do desempenho escolar.

Segundo ANDRADE, TAVARES & VALLE (2000), existem dois tipos de equalização via itens comuns (ou vertical), *a posteriori* e *a priori*. A equalização *a posteriori*, como o próprio nome sugere, é realizada após a estimação, separadamente, de dois conjuntos de itens que foram submetidos a duas populações de interesse. Como as estimativas dos parâmetros dos itens não dependem do grupo de examinandos, tendo estimativas em duas diferentes escalas, pode-se estabelecer algum tipo de relação entre os itens comuns que permita colocar os parâmetros de um dos conjuntos de itens na escala do outro.

---

envolvem cálculos complexos e programas computacionais específicos. Conforme *default* do BILOG, os parâmetros dos itens são estimados por meio do método de máxima verossimilhança marginal e as habilidades, por meio do método do valor esperado EAP–*Expected a posteriori estimator* (MISLEVY & BOCK, 1990).

Assim sendo, autores como HAMBLETON, SWAMINATHAN & ROGERS (1991), KOLEN & BRENNAN (1995) e ANDRADE, TAVARES & VALLE (2000) afirmam que, se o pressuposto da invariância dos parâmetros dos itens for válido e o modelo se ajustar aos dados, os parâmetros de discriminação e de dificuldade de um item comum a dois grupos de respondentes devem satisfazer, senão por flutuações amostrais, as seguintes relações lineares:

$$b_{G1} = \alpha b_{G2} + \beta$$

$$a_{G1} = \frac{1}{\alpha} a_{G2}$$

em que  $b_{G1}$  e  $b_{G2}$  correspondem aos valores dos parâmetros de dificuldade dos grupos 1 e 2, respectivamente;  $a_{G1}$  e  $a_{G2}$  representam os valores dos parâmetros de discriminação desses grupos. Desta forma, os coeficientes  $\alpha$  e  $\beta$  são determinados a partir das equações anteriormente apresentadas e as estimativas dos parâmetros dos itens do grupo 2 podem ser colocadas na mesma escala das estimativas do grupo 1. Conforme as relações lineares acima sugerem, a solução mais comum seria adotar o método de regressão linear simples para determinar tais coeficientes. Entretanto, pesquisadores da área criticam o fato desse método não ser simétrico, ou seja, resultados de uma regressão de  $x$  por  $y$  são diferentes dos resultados de uma regressão de  $y$  por  $x$ .

Vários são os métodos de equalização *a posteriori*, que se baseiam em relações lineares entre os parâmetros de um mesmo item medidos em escalas diferentes e são simétricos, como por exemplo, o conhecido método Média-Desvio (HAMBLETON, SWAMINATHAN & ROGERS, 1991; KOLEN & BRENNAN, 1995; ANDRADE, TAVARES & VALLE, 2000).

Segundo os mesmos autores, o método Média-Desvio (*Mean-Sigma*) é baseado nas seguintes relações:

$$\alpha = \frac{S_{G1}}{S_{G2}}$$

$$\beta = M_{G1} - \alpha M_{G2}$$

em que  $S_{G1}$  e  $S_{G2}$  representam os desvios-padrão amostrais das estimativas dos parâmetros de dificuldade dos itens comuns nos grupos 1 e 2, respectivamente;  $M_{G1}$  e  $M_{G2}$  correspondem às médias amostrais desses itens comuns. Assim sendo, as habilidades dos alunos do grupo 2

podem ser colocadas na mesma escala das habilidades dos alunos do grupo 1 por meio da relação:

$$\theta^1_{G2} = \alpha\theta_{G2} + \beta$$

em que  $\theta^1_{G2}$  é uma estimativa de habilidade  $\theta_{G2}$  representada na escala do grupo 1.

Por sua vez, a equalização *a priori* é realizada durante o processo de estimação ou calibração dos itens, sendo mais eficaz do que a equalização *a posteriori*, já que possibilita estimações com menores erros (ANDRADE, TAVARES & VALLE, 2000). Ressaltando, ainda, as vantagens do processo de equalização *a priori*, os mesmos autores explicam que como esta equalização é realizada automaticamente no próprio processo de calibração, não se está mais submetido a possíveis erros de diferenças nas estimativas dos parâmetros em função do método de equalização adotado.

O processo de equalização *a posteriori* anteriormente adotado (como, por exemplo, Média-Média e Média-Desvio) envolve várias etapas de estimação dos parâmetros dos itens e habilidades, que são estimados separadamente para cada uma das populações. Posteriormente, por meio de transformações lineares dos itens comuns todos parâmetros são equalizados. Assim sendo, acredita-se que as várias estimações envolvidas produzem mais erros no processo, do que aqueles produzidos pela estimação *a priori*. Ademais, conforme já mencionado, com a utilização da equalização *a priori* não se está mais submetido a diferentes resultados nas estimativas obtidos em função do método de equalização *a posteriori* adotado.

Complementarmente, na presença de muitas populações (por exemplo  $K \geq 5$ ), a realização de equalização *a posteriori* produz erros relativos à regressão, associados a cada equalização entre duas populações (ANDRADE, TAVARES & VALLE, 2000). Conseqüentemente, estes erros vão sendo acumulados até que se consiga equalizar todas as populações estudadas, podendo levar a uma má estimação dos parâmetros.

Por meio de simulações, ANDRADE (2001) mostra que as equalizações *a priori* exigem um menor número de itens comuns do que as equalizações *a posteriori*, para a obtenção de resultados similares. Em geral, alguns autores têm sugerido pelo menos 6 itens comuns entre 2 provas de 30 itens, quando a equalização é feita durante a calibração, como por exemplo, ANDRADE (2001) e ANDRADE, TAVARES & VALLE (2000).

Diante do exposto sobre as vantagens da realização da equalização *a priori*, entende-se como mais adequado a utilização deste método no estudo de dados longitudinais com a presença de itens comuns. Assim, considerando as habilidades numa mesma escala de

referência (abril de 1999), os parâmetros de itens podem ser estimados conjuntamente, permitindo que as habilidades estimadas sejam diretamente comparadas, uma vez que elas foram estimadas sob uma mesma escala.

### **3.4- Modelos para o desempenho escolar**

A utilização do modelo hierárquico longitudinal é principalmente justificada pela estrutura de dependência presente nos dados educacionais longitudinais. Esta dependência foi descrita por WILLETT (1997) enfatizando dois níveis:

- a) nível intra-alunos, em que se foca no individual, tentando entender as mudanças individuais ocorridas ao longo do estudo, considerando que a dependência entre as observações é devido ao fato do mesmo indivíduo ser medido várias vezes ao longo do tempo;
- b) nível inter-aluno, permitindo a formulação de questões sobre quais as mudanças individuais no tempo diferem entre os alunos em virtude de suas características fixas no tempo.

O mesmo autor esclarece que estes dois níveis, intra e inter alunos, constituem uma hierarquia natural, em que as características intra-alunos estão aninhadas nas características inter-alunos, fornecendo um importante arcabouço teórico para medidas de mudança.

Complementarmente, um terceiro nível que pode ser acrescentado a essa hierarquia seria o nível da escola, uma vez que os alunos se encontram aninhados nelas. Desta forma, seria possível indagar quais as características escolares interferem no desempenho dos alunos.

Ademais, uma das mais importantes justificativas e/ou vantagens em se utilizar modelos hierárquicos longitudinais na estimação dos fatores associados ao desempenho escolar é a possibilidade de obter estimativas mais confiáveis das mudanças individuais, especialmente, quando se trabalha com mais de duas observações (WILLETT, 1994, 1997; SINGER & WILLETT, 2003).

Segundo WILLETT (1994, 1997), o processo contínuo de desenvolvimento humano pode ser facilmente e eficientemente documentado simplesmente medindo cada pessoa repetitivamente sobre extensos períodos de tempo. O referido autor menciona a controvérsia existente entre os pesquisadores sobre a confiabilidade de estimativas provenientes de modelos que utilizaram a diferença entre escores, ou seja, apenas duas observações.

Se o atributo de interesse é a mudança sobre um longo período de tempo, talvez três ou quatro medidas de cada pessoa possam ser suficientes para captar a forma e a direção da

mudança, se a trajetória de mudança individual não for muito complexa (WILLETT, 1989a, 1989b, 1994, 1997, WILLETT & BUB, 2004).

### 3.4.1- Modelos hierárquicos

Antes de apresentar os modelos hierárquicos longitudinais, considera-se necessário a inclusão da descrição de modelos hierárquicos mais simples, modelos não-longitudinais. Neste tipo de modelo, os dados apresentam uma estrutura de dependência, na qual os alunos estão aninhados em escolas, por exemplo. Assim, o nível 1 considera as características individuais e familiares dos alunos, enquanto no nível 2 encontram-se as variáveis escolares.

Esta existência de agrupamentos por unidade de análise permite uma utilização mais eficiente dos dados, evitando erros decorrentes do uso de metodologia inadequada, como erros de estimação dos parâmetros e erros na estimação da variância (RAUDENBUSH & BRYK, 2002; GOLDSTEIN, 1995).

Assim, a utilização de modelos hierárquicos, em dados com esta estrutura de dependência, permite obter melhores estimativas para as unidades específicas de análise, testar hipóteses relativas a efeitos entre níveis e identificar a partição da variância em componentes (RAUDENBUSH & BRYK, 2002; GOLDSTEIN, 1995).

Existem várias especificações para os modelos hierárquicos não-longitudinais. Contudo, a seguir, é apresentada a mais simples, por meio da qual é possível fazer as demais especificações. Apesar de ser considerado o mais simples, o modelo de análise de variância com um fator aleatório fornece informações preliminares úteis sobre quanto da variação nos resultados é intra-escolar ou inter-escolar.

Este modelo pode ser descrito pelas seguintes funções (RAUDENBUSH & BRYK, 2002; GOLDSTEIN, 1995):

#### Nível 1: Alunos

$$Y_{ij} = \beta_{0j} + r_{ij}$$

$i=1,2,\dots,n_j$  alunos na escola  $j$ ;

$j=1,2,\dots, J$  escolas.

em que:

$Y_{ij}$  representa o desempenho do aluno  $i$  na escola  $j$ ;  $\beta_{0j}$  é o desempenho médio, isto é, o desempenho esperado da escola  $j$ ;  $r_{ij}$  é o erro aleatório de nível 1, considerado independente e com distribuição  $\sim N(0, \sigma^2)$ ;  $\sigma^2$  é considerado a variância a nível do aluno.

### Nível 2: Escolas

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

em que:

$\beta_{0j}$  (o desempenho médio da escola  $j$ ) é representado como uma função de  $\gamma_{00}$  (da média, ou seja, do valor esperado, de todas as escolas consideradas no estudo) mais  $u_{0j}$  (erro aleatório de nível 2, considerado independente e com distribuição  $\sim N(0, \tau_{00})$ );  $\tau_{00}$  é a variância a nível da escola;  $\text{cov}(r_{ij}, u_{0j}) = 0$ , a covariância entre os erros de nível 1 e 2 é igual a zero.

A correlação intra-classe medida pela proporção da variância no desempenho dos alunos que é devida à variação entre as escolas é uma estatística essencial para a estimação da importância da diferença entre as escolas no que diz respeito ao desempenho escolar. Esta estatística é calculada através do modelo de análise de variância com um fator, que não apresenta variáveis explicativas em nenhum dos seus níveis.

De acordo com RAUDENBUSH & BRYK (2002), a proporção da variância do desempenho escolar que é devida à diferença entre as escolas é dada por:

$$\frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2}$$

### 3.4.2- Modelos hierárquicos longitudinais

De início, vale ressaltar que a utilização de dados longitudinais tem muito a contribuir com as avaliações da qualidade do ensino, permitindo a análise da trajetória da proficiência dos alunos no tempo e a estimação adequada da mudança (SINGER & WILLETT, 2003) e do efeito-escola (FERRÃO & FERNANDES, 2003).

Nos modelos hierárquicos longitudinais, o efeito-escola pode estar associado ao *status* inicial de desempenho, à taxa de crescimento instantânea no desempenho ou à curvatura



(aceleração) do crescimento (GOLDSTEIN, 1995; RAUDENBUSH & BRYK, 2002; RASBASH *et al*, 2004).

É importante ressaltar que, em geral, a curva de aprendizagem apresenta um formato não linear, no qual se verifica um aprendizado a taxas decrescentes ao longo do tempo, de forma que o desempenho dos alunos tende a uma certa estabilidade no final do processo.

Sobre os níveis geralmente abordados com a utilização de modelos hierárquicos longitudinais, LEE (2004) afirma que os escores dos alunos e as variáveis relacionadas ao tempo são incluídas no nível 1, respectivamente, como variável resposta e variáveis explicativas. No nível 2, as características pessoais ou familiares, fixas no tempo, são incluídas como variáveis explicativas. Por fim, o nível 3 contempla as variáveis explicativas escolares associadas ao desempenho do aluno.

Os programas geralmente utilizados são o *HLM (Hierarquical Linear Model)* e o *MLwin*, respectivamente desenvolvidos por RAUDENBUSH *et al* (2000) e RASBASH *et al* (2004). A escolha do programa utilizado depende da sua adequação ao trabalhar com as características do banco de dados.

O *HLM* apesar de ser um programa mais didático na apresentação dos resultados, não trabalha na presença de dados perdidos nos níveis 2 e 3. Assim, como o *HLM* só trabalha com dados perdidos no nível 1, o *MLwin* v. 2.02 foi utilizado no ajuste dos modelos dessa tese, uma vez que, este programa ajusta modelos na presença de dados perdidos em todos os níveis considerados, ignorando esses dados automaticamente (RASBASH *et al*, 2004).

O *MLwin* v.2.02 adota como método de estimação o *IGLS* (Mínimos quadrados generalizados iterativos). Este método é o mais adequado ao ajuste de modelos hierárquicos de mudança, porque além de permitir resíduos autocorrelacionados e heterocedásticos, admite um indicativo de ajuste adequado, se o processo converge nos limites estabelecidos no *default* e com poucas interações (SINGER & WILLETT, 2003; JONES, 2004).

#### **3.4.2.1- Modelo incondicional**

O primeiro modelo apresentado é o modelo incondicional. De acordo com a nomenclatura adotada por RAUDENBUSH & BRYK (2002), o modelo incondicional é aquele que não apresenta variável explicativa nos níveis 2 e 3, somente incluindo no nível 1, as variáveis idade e/ou idade ao quadrado.

Este modelo permite o cálculo de estatísticas, por meio das quais é possível determinar a percentagem da variabilidade do *status* inicial, da taxa instantânea de crescimento e da

curvatura do crescimento, que são devidas à variabilidade entre as escolas. Desta forma, a análise longitudinal deve começar pela estimação do modelo incondicional, que permite estimar adequadamente o efeito-escola relacionado ao *status* inicial, à taxa instantânea de crescimento do desempenho e à curvatura desse crescimento.

A especificação teórica do modelo incondicional descrita a seguir é uma adaptação da especificação contida no livro de RAUDENBUSH & BRYK (2002), seguindo a mesma nomenclatura desse, para um modelo sem variáveis explicativas:

### **Especificação teórica do modelo incondicional**

#### **Nível 1: Intra-aluno**

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(idad\_ci)_{tij} + \pi_{2ij}(idadci2)_{tij} + e_{tij}$$

$$\begin{aligned} t &= 0, 1, \dots, t_{ij} \\ i &= 1, 2, \dots, N_j \\ j &= 1, 2, \dots, J \end{aligned}$$

em que:

$Y_{tij}$  é o desempenho do aluno  $i$  na escola  $j$  no tempo  $t$ ;

$idad\_ci$  representa a idade do aluno  $i$  na escola  $j$  no tempo  $t$ , centralizada na idade ideal na quarta série (10 anos);

$idadci2$  é igual a  $idad\_ci * idad\_ci$ ;

$\pi_{0ij}$  é o *status* inicial do aluno  $i$  da escola  $j$ , isto é, o desempenho esperado para esse aluno no início da pesquisa;

$\pi_{1ij}$  é a taxa de crescimento instantânea para o aluno  $ij$  na sua idade em  $t$ ;

$\pi_{2ij}$  é a curvatura (ou aceleração) em cada trajetória de crescimento do desempenho escolar;

$e_{tij}$  é o erro aleatório de nível 1, considerado independente e com distribuição  $\sim N(0, \sigma^2)$ ;

A combinação das taxas  $(\pi_{1ij}, \pi_{2ij})$  vai ao encontro de uma estrutura esperada para uma curva de aprendizagem, se  $\pi_{1ij} > 0$  e  $\pi_{2ij} < 0$  – taxas de crescimento cada vez menores, tendendo à estabilidade no final do processo. Em geral, a taxa de crescimento de qualquer idade particular é a derivada primeira do modelo de crescimento:

$$\text{taxa de crescimento na idade } a \text{ no tempo } t = \pi_{1ij} + 2\pi_{2ij}(idade\_ci)$$

### Nível 2: Inter-aluno

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

$$\pi_{2ij} = \beta_{20j} + r_{2ij}$$

em que:

$\beta_{00j}$  representa o *status* inicial médio, ou seja, esperado na escola  $j$ ;

$r_{0ij}$  é erro aleatório de nível 2, considerado independente e com distribuição  $\sim N(0, \tau_{\pi 00})$ ;

$\tau_{\pi 00}$  representa a variância dos *status* iniciais entre os alunos da escola  $j$ ;

$\beta_{10j}$  é a taxa instantânea esperada de aprendizado na escola  $j$ ;

$r_{1ij}$  é erro aleatório de nível 2, considerado independente e com distribuição  $\sim N(0, \tau_{\pi 11})$ ;

$\tau_{\pi 11}$  representa a variância das taxas instantâneas de crescimento entre o aprendizado dos alunos da escola  $j$ ;

$\beta_{20j}$  é a aceleração esperada do aprendizado na escola  $j$ ;

$r_{2ij}$  é erro aleatório de nível 2, considerado independente e com distribuição  $\sim N(0, \tau_{\pi 22})$ ;

$\tau_{\pi 22}$  representa a variância das curvaturas do crescimento entre o aprendizado dos alunos da escola  $j$ ;

### Nível 3: Escola

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}$$

$$\beta_{20j} = \gamma_{200} + u_{20j}$$

em que:

$\gamma_{000}$  é o *status* inicial médio (esperado) de todas as escolas pesquisadas;

$u_{00j}$  é erro aleatório de nível 3, considerado independente e com distribuição  $\sim N(0, \tau_{\beta 00})$ ;

$\tau_{\beta 00}$  representa a variância dos *status* iniciais entre as escolas;

$\gamma_{100}$  é a taxa instantânea média (esperada) de todas as escolas pesquisadas;

$u_{10j}$  é erro aleatório de nível 3, considerado independente e com distribuição  $\sim N(0, \tau_{\beta 11})$ ;

$\tau_{\beta 11}$  representa a variância das taxas instantâneas de crescimento entre as escolas;

$\gamma_{200}$  é a curvatura média (esperada) de todas as escolas pesquisada;

$u_{20j}$  é erro aleatório de nível 3, considerado independente e com distribuição  $\sim N(0, \tau_{\beta 22})$ ;

$\tau_{\beta 22}$  representa a variância das curvaturas de crescimento entre as escolas;

A matriz variância-covariância do nível 2,  $T_{\pi}$ , e a matriz variância-covariância do nível 3,  $T_{\beta}$ , são:

$$\begin{aligned} \text{Nível 2} - T_{\pi} &= \begin{bmatrix} \tau_{\pi 00} & & \\ \tau_{\pi 01} & \tau_{\pi 11} & \\ \tau_{\pi 02} & \tau_{\pi 12} & \tau_{\pi 22} \end{bmatrix} \\ \text{Nível 3} - T_{\beta} &= \begin{bmatrix} \tau_{\beta 00} & & \\ \tau_{\beta 01} & \tau_{\beta 11} & \\ \tau_{\beta 02} & \tau_{\beta 12} & \tau_{\beta 22} \end{bmatrix} \end{aligned}$$

Baseando-se nas estimativas dos componentes da variância, é possível computar a porcentagem da variação no desempenho dos alunos que é devida a variação entre as escolas, para o *status* inicial, a taxa instantânea de crescimento e a aceleração desse crescimento.

Assim, os modelos incondicionais permitem estimar o efeito-escola, possibilitando confirmar a importância da presença do nível 3 no modelo, ou seja, a importância da variabilidade entre as escolas na explicação da variabilidade dos coeficientes associados ao *status* inicial, à taxa instantânea de crescimento e à curvatura desse crescimento.

Formalmente, esse modelo hierárquico longitudinal incondicional apresenta três tipos de efeito-escola, respectivamente, relacionados ao intercepto (*status* inicial), ao termo linear (taxa instantânea de crescimento) e ao termo não-linear (curvatura do crescimento), conforme descritos nas seguintes equações (RAUDENBUSH & BRYK, 2002):

1. a porcentagem da variação do *status* inicial explicada pela variabilidade entre as escolas:

$$\frac{\hat{\tau}_{\beta 00}}{\hat{\tau}_{\beta 00} + \hat{\tau}_{\pi 00}}$$

em que  $\hat{\tau}_{\beta 00}$  é a estimativa da variância do *status* inicial entre as escolas;  $\hat{\tau}_{\pi 00}$  é a estimativa da variância do *status* inicial dentro das escolas;

2. a porcentagem da variação da taxa instantânea de crescimento explicada pela variabilidade entre as escolas:

$$\frac{\hat{\tau}_{\beta 11}}{\hat{\tau}_{\beta 11} + \hat{\tau}_{\pi 11}}$$

em que  $\hat{\tau}_{\beta 11}$  é a estimativa da variância da taxa instantânea de crescimento entre as escolas;  $\hat{\tau}_{\pi 11}$  é a estimativa da variância da taxa instantânea de crescimento dentro das escolas;

3. a porcentagem da variação da curvatura do crescimento explicada pela variabilidade entre as escolas:

$$\frac{\hat{\tau}_{\beta 22}}{\hat{\tau}_{\beta 22} + \hat{\tau}_{\pi 22}}$$

em que  $\hat{\tau}_{\beta 22}$  é a estimativa da variância da curvatura do crescimento entre as escolas;  $\hat{\tau}_{\pi 22}$  é a estimativa da variância da aceleração do crescimento dentro das escolas.

### 3.4.2.2- Modelo condicional

Em geral, os modelos condicionais apresentam variáveis explicativas nos níveis 2 e/ou 3, podendo ainda considerar variáveis explicativas no nível 1, além daquelas relativas à idade do aluno. A especificação teórica de modelo condicional mostrada abaixo considera a variável explicativa trabalha no nível 1 do modelo, ou seja, no nível referente às características intra-aluno.

#### Especificação teórica de modelos condicionais

##### Nível 1: Intra-aluno

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(idad\_ci)_{tij} + \pi_{2ij}(idadci2)_{tij} + \pi_{3ij}(trab)_{tij} + e_{tij}$$

$$\begin{aligned}
t &= 0, 1, \dots, t_{ij} \\
i &= 1, 2, \dots, N_j \\
j &= 1, 2, \dots, J
\end{aligned}$$

em que:

$Y_{ij}$  é o desempenho do aluno  $i$  na escola  $j$  no tempo  $t$ ;

$idad\_ci$  representa a idade do aluno  $i$  na escola  $j$  no tempo  $t$ , centralizada na idade ideal na quarta série (10 anos);

$idadci2$  é igual a  $idad\_ci * idad\_ci$ ;

$trab$  corresponde à variável indicadora se o aluno trabalha;

$\pi_{0ij}$  é o *status* inicial do aluno  $i$  da escola  $j$ , controlando pelas  $q$ -ésimas variáveis explicativas do aluno fixas em relação ao tempo ( $X_{qij}$ ) e pelas  $s$ -ésimas variáveis explicativas da escola ( $W_{sj}$ );

$\pi_{1ij}$  é a taxa instantânea de crescimento para o aluno  $ij$  na sua idade em  $t$ , controlando pelas  $q$ -ésimas variáveis explicativas do aluno fixas em relação ao tempo ( $X_{qij}$ ) e pelas  $s$ -ésimas variáveis explicativas da escola ( $W_{sj}$ );

$\pi_{2ij}$  é a curvatura (ou aceleração) em cada trajetória de crescimento do desempenho escolar, controlando pelas  $q$ -ésimas variáveis explicativas do aluno fixas em relação ao tempo ( $X_{qij}$ ) e pelas  $s$ -ésimas variáveis explicativas da escola ( $W_{sj}$ );

$\pi_{3ij}$  é o efeito do trabalho na trajetória do aluno  $ij$ ;

$e_{tij}$  é o erro aleatório de nível 1, considerado independente e com distribuição  $\sim N(0, \sigma^2)$ ;

## Nível 2: Inter-aluno

$$\pi_{p ij} = \beta_{p0j} + \sum_{q=1}^Q \beta_{q0j} X_{qij} + r_{p ij}$$

$$p = 0, 1, 2.$$

em que:

$\pi_{p ij}$  é o  $p$ -ésimo parâmetro do modelo de crescimento;

$\beta_{p0j}$  representa o valor médio (esperado) para o  $p$ -ésimo parâmetro do modelo de crescimento, controlando pelas  $s$ -ésimas variáveis explicativas da escola ( $W_{sj}$ );

$X_{qij}$  representa a  $q$ -ésima variável explicativa, correspondente a características do aluno fixas em relação ao tempo (por exemplo, sexo);

$\beta_{q0j}$  representa o efeito da  $q$ -ésima variável  $X_{qij}$  no  $p$ -ésimo parâmetro do modelo de crescimento;

$r_{pij}$  é o  $p$ -ésimo erro aleatório de nível 2; Como  $p$  varia de 0 a 2, os erros aleatórios de nível 2 podem ser interpretados conforme modelo incondicional.

### Nível 3: Escola

$$\beta_{p0j} = \gamma_{p00} + \sum_{s=1}^S y_{s0j} W_{sj} + u_{p0j}$$

$$p = 0,1,2.$$

em que:

$\gamma_{p00}$  representa o valor médio (esperado) do  $p$ -ésimo parâmetro da escola  $j$ ;

$W_{sj}$  representa a  $s$ -ésima variável explicativa de nível 3, correspondendo a características da escola  $j$ ;

$y_{s0j}$  é o efeito da  $s$ -ésima variável ( $W_{sj}$ ) no  $p$ -ésimo parâmetro da escola  $j$ ;

$u_{p0j}$  é o  $p$ -ésimo erro aleatório de nível 3. Como  $p$  varia de 0 a 2, os erros aleatórios de nível 3 podem ser interpretados conforme modelo incondicional;

As matrizes de variância-covariância para os níveis 2 e 3 são:

$$\text{Nível 2 - } Tr = \begin{bmatrix} \sigma^2_{r0} & & \\ \sigma_{r01} & \sigma^2_{r1} & \\ \sigma_{r02} & \sigma_{r12} & \sigma^2_{r2} \end{bmatrix}$$

$$\text{Nível 3 - } Tu = \begin{bmatrix} \sigma^2_{u0} & & \\ \sigma_{u01} & \sigma^2_{u1} & \\ \sigma_{u02} & \sigma_{u12} & \sigma^2_{u2} \end{bmatrix}$$

## **4- APRESENTAÇÃO E ANÁLISE DOS RESULTADOS**

### **4.1- Procedimentos adotados na calibração e na equalização dos itens de português e de matemática**

Os procedimentos utilizados para a calibração dos itens de português e matemática estão descritos nas TAB. 3 e 5, respectivamente, assim como os motivos pelos quais alguns itens foram retirados até a obtenção da calibração final. Vale ressaltar que em ambas as disciplinas os testes continham quarenta itens, com exceção do período abril de 1999, no qual os testes contavam com somente trinta e seis itens.

O processo de calibração dos itens foi realizado em vários passos, estratégia utilizada para manter o maior número possível de itens nas provas. A qualidade dos itens foi avaliada considerando-se, principalmente, os valores referentes às estimativas dos parâmetros de discriminação e de dificuldade e, ainda, os erros-padrão (EP) destas estimativas.

Em resumo, os critérios considerados para a retirada dos itens foram: a) correlação bisserial negativa; b) item com 100% de acerto; c) parâmetro de discriminação menor do que 0,4; d) erros-padrão referentes aos parâmetros de discriminação ou de dificuldade maiores que 0,3 (BAKER, 2001).

Depois do primeiro passo, ou seja, após a primeira calibração sem a retirada de itens, os parâmetros dos itens foram sendo novamente re-estimados por meio do programa *BILOG-MG*, que estima conjuntamente os itens de todos os períodos e coloca a habilidade de todas as séries numa mesma escala, fazendo com que todos os escores dos alunos sejam comparáveis, por meio do processo de equalização.

Os primeiros resultados apresentados abaixo são da disciplina de português. Antes, porém, vale ressaltar que as provas de abril 1999, nas duas disciplinas, foram consideradas como escala de referência para os processos de calibração e equalização.

Na TAB. 4, que apresenta um esquema longitudinal com a numeração dos itens das provas de português de abril de 1999 a novembro de 2003, as questões estão numeradas por ordem e ano, sendo que o primeiro dígito diz respeito ao ano e os dois últimos, à posição do item na prova. Os itens comuns se mantêm com o mesmo número que recebem na primeira prova em que aparecem, mesmo quando se repetem em provas seguintes. Por exemplo, o item 307 aparece pela primeira vez na posição 7 do teste de novembro de 2000. Este item se repete na posição de número 16 do teste de novembro de 2001 (no lugar do item 416). Novamente, o



item 307 aparece na posição 20 no teste de 2002 e, finalmente, na posição 32 no teste realizado em 2003.

TABELA 3- Procedimentos para a calibração e retirada de itens, referentes às provas de português nos períodos de abril de 1999 a novembro de 2003

Passo	Item retirado	Motivo
1	nenhum	O processo de estimação não convergiu
2	101 106 107 209 215 617	Todos com parâmetro de discriminação menor do que 0.400
3	203 213	EP > 0.300 para o parâmetro a
4	508	100% de acerto

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

É importante lembrar que seria possível tratar de formas diferentes itens comuns que foram aplicados em mais do que duas avaliações, diante da possibilidade do aluno memorizar a resposta correta. Como é o caso do item comum 307, mencionado anteriormente, que fez parte dos testes de português a partir de 2000, totalizando quatro participações nas provas do período abril de 1999 a novembro de 2003.

Uma das possíveis formas de se tratar o item 307 seria considerá-lo somente nas avaliações de 2000 e 2001, eliminando-o das avaliações de 2002 e 2003. Entretanto, a opção escolhida foi a de manter o item 307 no processo de estimação, porque não se esperava que o resultado sofresse muitas mudanças, caso este item fosse retirado da aplicação de 2002 e/ou 2003.

TABELA 4- Esquema longitudinal dos itens das provas de português de abril/1999 a novembro/2003

Itens	Abr/99	Nov/99	2000	2001	2002	2003
1	101	201	124	401	501	601
2	102	124	302	402	502	323
3	103	203	126	403	330	603
4	104	204	304	315	418	604
5	105	132	222	405	505	526
6	106	206	306	406	506	606
7	107	207	307	407	507	607
8	108	208	224	240	508	429
9	109	209	228	409	509	609
10	110	210	310	410	510	409
11	111	211	131	411	511	411
12	112	212	312	119	401	612
13	113	213	313	413	513	613
14	114	214	314	414	514	614
15	115	215	315	415	515	615
16	116	109	123	307	516	616
17	117	126	317	417	409	617
18	118	119	122	418	518	618
19	119	219	319	419	519	619
20	120	220	320	317	307	620
21	121	221	240	421	521	621
22	122	222	119	422	522	622
23	123	223	323	423	523	623
24	124	224	324	122	524	624
25	125	225	325	425	525	625
26	126	226	326	426	526	626
27	127	123	327	427	527	529
28	128	228	328	337	429	331
29	129	229	329	429	529	629
30	130	230	330	430	530	523
31	131	231	331	431	531	631
32	132	232	332	304	532	307
33	133	110	333	433	533	633
34	134	234	334	434	331	634
35	135	235	335	323	335	317
36	136	118	336	436	411	636
37		237	337	437	317	637
38		122	338	438	538	638
39		125	339	439	539	639
40		240	340	330	540	640

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Vale lembrar que os mesmos procedimentos utilizados na calibração e retirada de itens referentes aos testes de português foram adotados na calibração, na retirada e na equalização

dos itens de matemática. Conforme mostram as TAB. 5 e 6 somente um item comum (305) foi eliminado da calibração final, sendo que estava presente nas aplicações de 2000 e 2001.

TABELA 5- Procedimentos para a calibração e retirada de itens, referentes às provas de matemática nos períodos de abril de 1999 a novembro de 2003

Passo	Item retirado	Motivo
1	nenhum	O processo de estimação não convergiu
2	113	Correlação bisserial negativa
	429	100% de acerto
	617	Correlação bisserial negativa
	631	100% de acerto
	637	Correlação bisserial negativa
3	305	Todos com parâmetro de discriminação menor do que 0.400
	306	
	320	
	333	
	417	
	601	
	620	
4	107	EP > 0.300 para os parâmetros a ou b
	413	EP ≈ 0.25 para o parâmetro a
	206	EP > 0.15 para o parâmetro b
5	211	Todos com parâmetro de discriminação menor do que 0.400
	415	
	606	
6	426	a < 0.400

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR(2005)

TABELA 6- Esquema longitudinal dos itens das provas de matemática de abril/1999 a novembro/2003

Itens	Abr/99	Nov/99	2000	2001	2002	2003
1	101	201	109	401	501	601
2	102	109	302	305	502	602
3	103	203	303	403	503	603
4	104	204	204	213	504	511
5	105	110	305	309	422	501
6	106	206	306	406	506	606
7	107	207	218	119	507	607
8	108	208	308	408	508	608
9	109	209	309	409	432	432
10	110	210	219	303	406	534
11	111	211	311	324	511	515
12	112	212	312	312	303	537
13	113	213	223	413	513	613
14	114	214	119	210	514	532
15	115	215	315	415	515	303
16	116	216	316	334	420	616
17	117	134	317	417	517	617
18	118	218	318	418	518	618
19	119	219	230	419	519	619
20	120	220	320	420	434	620
21	121	221	321	219	521	621
22	122	222	128	422	522	622
23	123	223	323	423	523	623
24	124	119	324	424	312	526
25	125	225	114	425	403	527
26	126	226	326	426	526	626
27	127	227	327	427	527	627
28	128	228	124	428	528	628
29	129	229	329	429	423	629
30	130	230	330	430	530	630
31	131	128	331	431	531	631
32	132	104	332	432	532	632
33	133	233	333	433	309	633
34	134	234	334	434	534	634
35	135	122	335	435	535	635
36	136	236	336	436	536	636
37		114	337	437	537	637
38		124	338	438	538	638
39		126	339	439	539	639
40		240	340	440	540	640

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Por sua vez, a avaliação das provas, considerando todos os períodos, foi feita por meio da distribuição dos valores dos parâmetros dos itens, em especial, o da discriminação, além da interpretação da curva de informação do teste, tanto para a disciplina de português quanto para a de matemática. Vale lembrar que os testes de português e de matemática são compostos por quarenta itens nas rodadas de novembro de 1999 a novembro de 2003. Em abril de 1999, os testes continham apenas trinta e seis itens. Entretanto, após a equalização, os itens comuns passaram a ser considerados uma única vez, enquanto outros tiveram que ser retirados dos testes por apresentarem, em especial, baixo valor para o parâmetro de discriminação. Dessa forma, o teste de português contém, ao todo, cento e sessenta e sete itens válidos, e o de matemática, cento e setenta e sete.

A TAB. 7 mostra a distribuição dos valores dos parâmetros de discriminação os testes de português e de matemática.

**TABELA 7- Distribuição dos valores do parâmetro de discriminação nos testes de português e de matemática**

Discriminação	Português	Matemática
$0.4 \leq a < 0.5$	9 5,08%	14 8,38%
$0.5 \leq a < 0.75$	46 25,99%	37 22,16%
$\geq 0.75$	122 68,93%	116 69,46%
Total	177 100%	167 100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Quanto aos resultados de português, a TAB. 7 mostra que os testes contêm, ao todo, cento e setenta e sete itens diferentes. Destes, cerca de 5% (5,08%) apresentaram parâmetro de discriminação com valores entre 0,4 e 0,5. Para aproximadamente um quarto (25,99%) dos itens de português, os valores dos parâmetros de discriminação foram estimados entre 0,5 e 0,75. A grande maioria dos itens, cerca de 70% (68,93%) obteve parâmetros de discriminação com valor igual ou superior a 0,75.

Quanto aos resultados de matemática, a TAB. 7 apresenta um percentual de aproximadamente 8% (8,38%) dos itens aplicados nas provas de matemática apresentou discriminação entre 0,4 e 0,5. Para trinta e sete itens, o valor do parâmetro de discriminação ficou entre 0,5 e 0,75, o que corresponde a aproximadamente 22% (22,16%) do total de itens aplicados. Um percentual em torno de 70% (69,46%) dos itens apresentou parâmetro igual ou

superior a 0,75. Diante do exposto, de uma forma geral, pode-se dizer que a maioria dos itens apresenta discriminação desejada ( $\geq 0.75$ ), tanto em matemática quanto em português (cerca de 70% nas duas disciplinas).

## 4.2- Análise Descritiva

### 4.2.1- A reorganização do banco de dados e o tratamento das variáveis para o ajuste dos modelos

Um problema com a numeração dos alunos foi identificado, quando se observou para um mesmo aluno números identificadores diferentes ao longo dos períodos de realização da pesquisa. Quanto à identificação das escolas, foram constatados alunos pertencentes a uma determinada escola na prova, enquanto estes mesmos alunos declararam estarem em uma outra escola no questionário sócio-econômico. Questões semelhantes acontecem com o turno e a turma preenchidos na ficha A e as mesmas variáveis declaradas nas provas. Em anexo, encontra-se a descrição do processo utilizado para a correção dos identificadores de alunos e das escolas, aplicado em todo o banco de dados.

Conforme mencionado no capítulo anterior, neste estudo foram considerados os alunos que realizaram mais de dois testes. A TAB. 8 mostra a distribuição desses alunos, de acordo com o número de realizações dos testes por disciplina. Tanto em português quanto em matemática, em torno de 42% dos alunos participaram de três testes, enquanto o percentual de alunos que realizaram os seis testes foi de aproximadamente 17%.

TABELA 8- Frequência de alunos por número de participação nas rodadas dos testes de português e de matemática

Número de rodadas na pesquisa	Português		Matemática	
	Frequência			
	absoluta	relativa	absoluta	relativa
3	4857	42,17	4899	42,29
4	3164	27,47	3167	27,34
5	1554	13,49	1569	13,54
6	1942	16,86	1950	16,83
Total	11517	100,00	11585	100,00

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Para as disciplinas de português e de matemática, a distribuição dos alunos por período de realização da pesquisa é apresentada na TAB. 9. De acordo com essa tabela, o percentual de alunos por período variou entre aproximadamente 13% em novembro de 2003 e em torno de 20% em novembro de 2000. Conforme a TAB. 10, pode-se observar que o número de escolas nas três primeiras rodadas da pesquisa foi de aproximadamente 157, enquanto que nas últimas três rodadas a quantidade de escolas aumentou em pelo menos 10.

TABELA 9- Número de alunos por período e por disciplina

Período	Português		Matemática	
	Frequência			
	absoluta	relativa	absoluta	relativa
abril/1999	7291	15,63	7327	15,62
nov/1999	6941	14,88	7007	14,94
nov/2000	9449	20,26	9504	20,26
nov/2001	9262	19,85	9288	19,80
nov/2002	7749	16,61	7795	16,62
nov/2003	5957	12,77	5989	12,77

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

TABELA 10- Número de escolas por período de realização das provas

Período	Número de Escolas
abril/1999	157
novembro/1999	157
novembro/2000	156
novembro/2001	171
novembro/2002	170
novembro/2003	167

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Quanto ao tratamento das variáveis explicativas, vale salientar que foram realizadas tentativas de estimação de fatores de infra-estrutura relacionados à escola durante todos os períodos pesquisados. Tomando por base as questões presentes em todos os períodos da pesquisa e aquelas consideradas em BARBOSA & FERNADES (2001) e SOARES, CÉSAR & MAMBRINI (2001), os fatores associados à infra-estrutura das escolas foram divididos em quatro, de acordo com as seguintes variáveis:

1. índice de serviços da escola: há atendimento médico, há atendimento odontológico, há transporte, há merenda;
2. índice de instalações das salas de aula: salas são iluminadas, salas são arejadas, volume de ruídos é prejudicial;
3. índice de mobiliário das salas de aula: há carteiras para todos os alunos; professor dispõe de mesa; professor dispõe de armário;
4. índice de conservação das salas de aula: carteiras estão em boas condições; quadro de giz está em boas condições de uso; mesa e/ou armário estão em boas condições.

Dessa forma, os índices foram estimados por análise fatorial para dados binários, com a utilização do programa *LAMI (Latent Model Interface)*. Esse programa ajusta os fatores de acordo com os modelos de resposta ao item<sup>7</sup>, apenas para menos de 3000 casos. Contudo, calcula a carga padronizada - a associação entre cada variável e seu fator latente correspondente - e as medidas da qualidade do ajuste da análise fatorial.

A análise fatorial permite a estimação de um fator latente, ou seja, não observável, que se manifesta por meio de um conjunto de variáveis observadas. Por estas variáveis serem a manifestação de um mesmo fator, apresentam alta correlação. O pressuposto da análise fatorial estabelece que a covariância entre as variáveis observadas é devida ao relacionamento de cada variável observada e a variável latente (BARTHOLOMEW *et al*, 2002). Assim, esta variável é dita ser a “verdadeira” fonte da covariância originalmente observada.

A TAB. 11 mostra a associação entre as variáveis consideradas e seu fator latente correspondente. Observa-se, então, que algumas variáveis apresentam pequena associação com o seu fator correspondente, conforme, por exemplo, o atendimento médico em 2001 e a existência de transporte em 1999, 2000 e 2001. Dessa forma, só foi possível estimar o índice de conservação das escolas, o único a apresentar um modelo ajustável, com pelo menos três variáveis para todos os períodos da pesquisa. Além de suas variáveis correspondentes apresentarem altas cargas padronizadas, a qualidade do ajuste do modelo desse índice mostrou-se satisfatória, de acordo com o método dos resíduos utilizado por BARTHOLOMEW *et al* (2002).

---

<sup>7</sup> Vale ressaltar que a Teoria da Resposta ao Item é uma das mais importantes abordagens para análise fatorial de dados binários (BARTHOLOMEW *et al*, 2002).



TABELA 11- Valores das cargas padronizadas na tentativa inicial de estimação dos índices de infra-estrutura das escolas

Índices de Infra-estrutura	1999	2000	2001	2002	2003
Índice de serviços da escola					
há atendimento médico	0,99	0,85	0,13	0,90	0,92
há atendimento odontológico	0,82	0,99	0,99	0,99	0,87
há merenda	0,49	0,98	0,64	0,98	0,66
há transporte	0,06	0,33	0,21	0,59	0,81
Índice de instalações das salas de aula					
salas são iluminadas	0,97	0,88	0,99	0,87	0,94
salas são arejadas	0,99	0,99	0,95	0,99	0,99
volume de ruídos é prejudicial	-0,64	-0,17	-0,61	-0,43	-0,75
Índice de mobiliário das salas de aula					
há carteira para todos os alunos	0,28	0,13	0,76	-0,31	0,43
professor dispõe de mesa	0,81	0,76	0,97	0,82	0,99
professor dispõe de armário	0,91	0,97	0,67	0,89	0,28
Índice de conservação das salas de aula					
carteiras estão em boas condições	0,86	0,80	0,74	0,87	0,81
quadro de giz em boas condições	0,90	0,63	0,90	0,89	0,83
mesa e/ou armário em boas condições	0,82	0,90	0,95	0,76	0,97

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Além do fator de infra-estrutura relacionado à escola foi necessário criar a variável tamanho médio da turma por série e por escola. A utilização da variável tamanho médio das turmas por escola, e não a nível de turma, pode ser justificada perante o resultado de escolas com poucas turmas, conforme pode ser observado nas TAB. 12 e 13. Estas tabelas mostram que o número de professores por escola é pequeno, de forma que em 50% das escolas, as séries apresentam 2, 3 ou 4 professores, indicando a existência de escolas com poucas turmas. Outra justificativa para a criação da variável tamanho médio da turma por escola foi a não coincidência entre a turma presente nas provas, nos questionários e na ficha A do aluno. Então, o pressuposto adotado foi o de que os dados referentes à turma provenientes da ficha A são os mais confiáveis, uma vez que não foram preenchidos pelos alunos e consideram aqueles que faltaram à aula no dia da prova. Uma outra variável criada foi o percentual de professores com curso superior por série e por escola.

TABELA 12 - Número de professores por série na disciplina de português

Português	quarta	quinta	sexta	Sétima	oitava
Número de válidos	158	156	172	170	168
professores <i>missing</i>	19	21	5	7	9
25	2	2	2	2	1
Percentis 50	2	3	3	2	2
75	3	4	4	3	3

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

TABELA 13 - Número de professores por série na disciplina de matemática

Matemática	quarta	Quinta	sexta	Sétima	oitava
Número de válidos	158	156	172	169	168
professores <i>missing</i>	19	21	5	8	9
25	2	3	2	2	1
Percentis 50	2	4	3	2	2
75	3	5	4	3	3

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Por fim, vale ressaltar que, no início do processo de ajuste dos modelos, as variáveis indicadoras que apresentavam resultados muito próximos foram agregadas quando possível, diminuindo a quantidade de tais variáveis nos modelos, na tentativa de torná-los mais parcimoniosos.

#### 4.2.2- Estatísticas descritivas das variáveis resposta e explicativas consideradas nos modelos

O QUADRO 2 mostra as variáveis resposta e explicativas consideradas no ajuste dos modelos hierárquicos longitudinais, de acordo com o nível e a categoria correspondentes.

QUADRO 2- Variáveis resposta e explicativas consideradas no ajuste dos modelos hierárquicos longitudinais

Níveis	Variáveis	Categorias
Nível 1	português = (score equalizado x 10) + 50	
	matemática =(score equalizado x 10) + 50	
Nível 2	idade centralizada na idade ideal da 4ª série	
	idade centralizada na idade ideal da 4ª série ao quadrado	
	trabalha	0-não 1-sim
	Sexo	0-masculino 1- feminino
	repetiu de ano	0-não 1-sim
	mãe nunca freqüentou a escola	
	mãe não completou a 4ª série do ensino fundamental (*)	0-não
	mãe não completou a 8ª do ensino fundamental	1-sim
	Mãe completou pelo menos o ensino fundamental	
	Nível 3	tamanho médio das turmas por escola e por períodos
índice de infra-estrutura de conservação das salas de aula		
percentual de professores com curso superior por escola		
conselho da escola não se reuniu		0-não
conselho da escola se reuniu uma vez(*)		1-sim
conselho da escola se reuniu duas ou mais vezes		
escola não adota sistema de recuperação da aprendizagem		
escola adota sistema bimestral de recuperação da aprendizagem(*)		0-não 1-sim
escola adota sistema semestral ou anual de recuperação		
escolas pertencentes aos estados do Mato grosso do Sul e Goiás		
escolas pertencentes ao estado de Pernambuco(*)	0-não	
escolas pertencentes aos estados de Rondônia, Pará e Sergipe	1-sim	

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

NOTA: (\*) indicadora omitida

Considerando o teste de português na TAB. 14, o desempenho médio dos alunos no início da pesquisa foi de 51,06. Na última rodada da pesquisa em 2003, o desempenho médio em português atingiu 60,74. No caso da disciplina de matemática, os alunos apresentaram um desempenho médio de 51,12 em abril de 1999, enquanto que no último período da pesquisa este desempenho foi de 61,74.

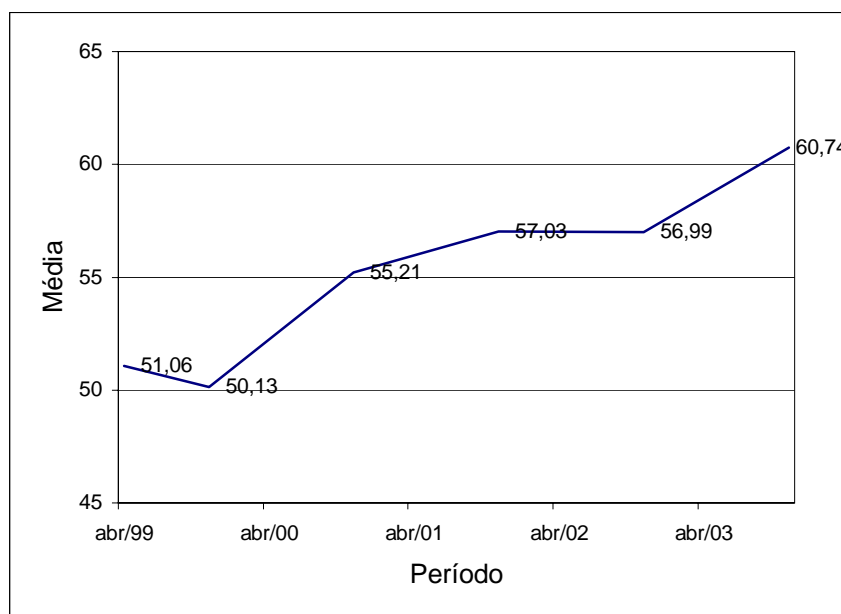
TABELA 14- Média e desvio-padrão dos testes de português e de matemática por período

Período	Português		Matemática	
	Média	Desvio-padrão	Média	Desvio-padrão
abril/1999	51,06	8,78	51,12	8,62
nov/1999	50,13	9,94	52,85	8,44
nov/2000	55,21	9,26	56,62	8,05
nov/2001	57,03	8,95	60,96	7,78
nov/2002	56,99	10,36	62,04	9,28
nov/2003	60,74	10,80	61,74	11,20

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

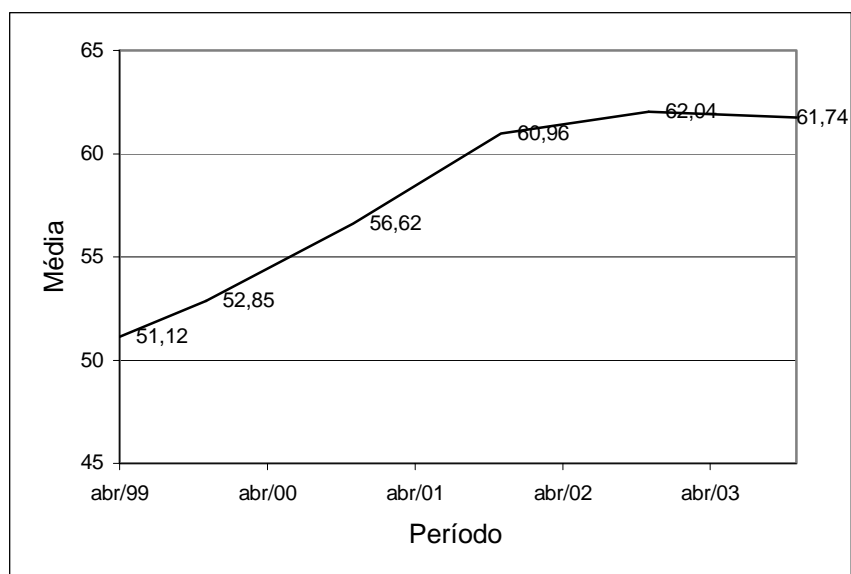
Por meio do GRAF. 4, observa-se que a curva das médias dos testes de português por período apresenta um formato próximo ao esperado pela curva de aprendizagem, especialmente entre os períodos de nov/99 a nov/02. Por sua vez, a curva das médias dos testes de matemática por período, conforme GRAF. 5, também apresenta um formato aproximado ao quadrático para os desempenhos dos alunos ao longo dos períodos.

GRÁFICO 4- Média dos testes de português por período



FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

GRÁFICO 5- Média dos testes de matemática por período



FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Pode-se observar nas tabelas 15 e 16, respectivamente, as médias de desempenho nas disciplinas de português e matemática por período e estado. Vale destacar que o estado de Pernambuco apresentou as menores médias de desempenho em ambas as disciplinas. Por outro lado, nota-se que comparando essas tabelas com a TAB. 14, o desempenho médio nos estados do Mato Grosso do Sul e de Goiás superou o desempenho médio geral em todos os períodos.

TABELA 15- Média e desvio-padrão dos testes de português por período e estado

Períodos	RO	PA	PE	SE	MS	GO
abr/1999	50,46	51,35	48,83	50,86	51,21	53,65
	(8,60)	(8,29)	(9,13)	(8,95)	(8,73)	(8,31)
nov/1999	49,55	49,65	47,36	50,90	51,58	52,58
	(9,40)	(9,33)	(10,02)	(9,49)	(10,11)	(9,97)
nov/2000	54,88	54,48	53,03	56,26	56,83	57,24
	(9,02)	(9,07)	(9,34)	(8,71)	(9,43)	(8,82)
nov/2001	56,27	57,19	55,03	57,20	58,29	58,46
	(9,00)	(8,56)	(9,07)	(8,76)	(9,06)	(8,77)
nov/2002	56,41	56,07	53,63	56,15	61,02	59,26
	(10,29)	(9,71)	(10,19)	(10,22)	(10,16)	(9,93)
nov/2003	58,48	59,45	58,23	59,11	65,74	62,75
	(11,42)	(10,25)	(10,94)	(11,28)	(9,49)	(10,12)

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

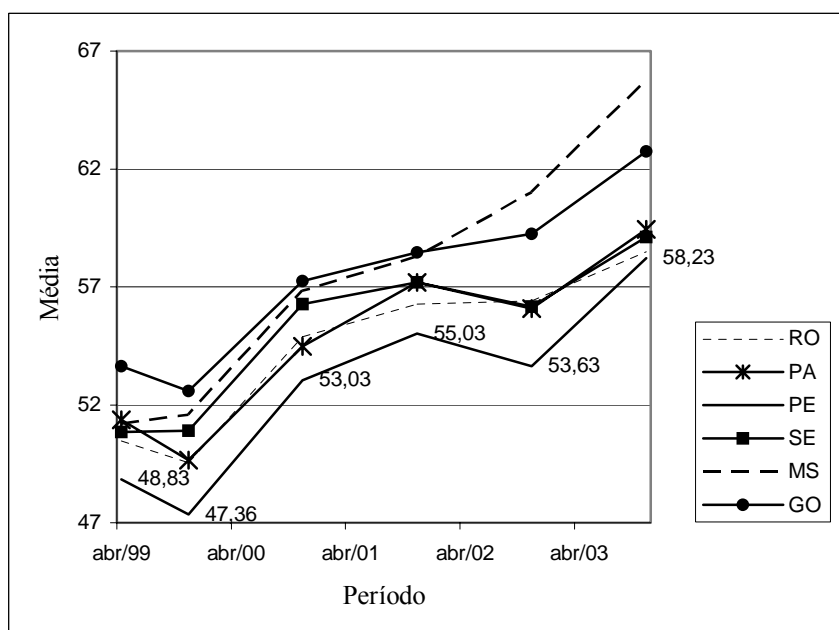
TABELA 16- Média e desvio-padrão dos testes de matemática por período e estado

Períodos	RO	PA	PE	SE	MS	GO
abr/1999	50,01	50,34	48,69	51,66	52,15	54,33
	(8,34)	(7,80)	(8,52)	(8,32)	(9,08)	(8,37)
nov/1999	51,69	52,31	49,97	52,71	54,33	56,15
	(8,35)	(7,50)	(8,50)	(8,05)	(8,74)	(7,86)
nov/2000	56,01	55,47	54,42	56,48	58,56	59,62
	(7,82)	(7,62)	(8,03)	(8,18)	(7,86)	(7,57)
nov/2001	60,50	60,37	58,59	60,60	63,03	63,12
	(7,53)	(7,19)	(7,71)	(7,95)	(8,01)	(7,38)
nov/2002	61,16	61,06	58,82	61,87	65,13	64,87
	(9,23)	(8,58)	(9,30)	(8,68)	(9,03)	(8,88)
nov/2003	59,64	60,00	59,08	60,61	66,63	64,33
	(10,37)	(10,44)	(11,27)	(11,39)	(10,64)	(10,98)

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

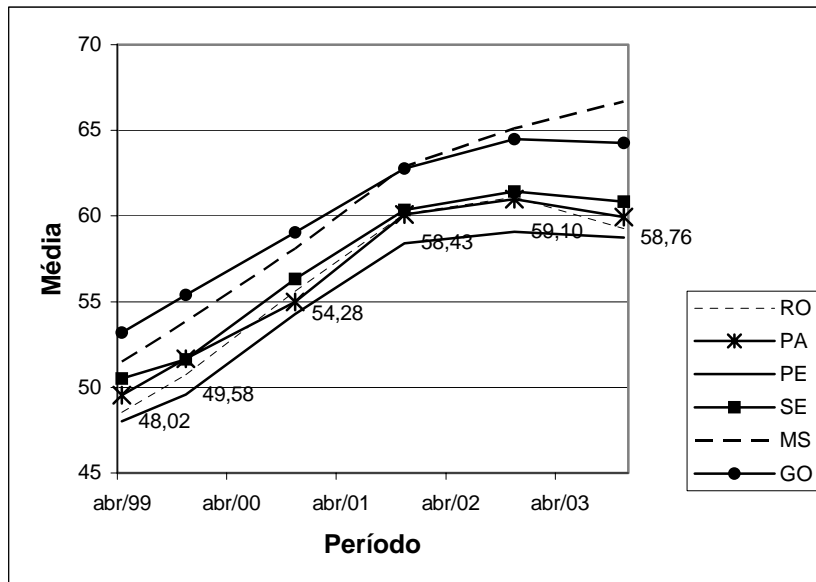
Complementarmente, nos GRAF. 6 e 7, podem ser vistas as diferenças existentes entre as média dos testes, respectivamente, de português e matemática por período e estado. Para ambas as disciplinas, observa-se uma proximidade no nível das curvas de aprendizagem do estado de Mato Grosso do Sul e de Goiás. Vale destacar que o formato da curva de aprendizagem de MS parece se distinguir dos demais formatos, especialmente, pelo aumento apresentado no desempenho entre 2001 e 2002. A posição no desempenho dos estados de Rondônia, Pará e Sergipe é muito próxima, tanto no que diz respeito ao nível quanto ao formato de suas curvas. Por sua vez, o estado de Pernambuco, apesar de apresentar formato da curva semelhante aos estados de posição intermediária (RO, PA e SE), encontra-se em nível inferior de desempenho em ambas as disciplinas.

GRÁFICO 6- Média dos testes de português por período e estado



FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

GRÁFICO 7- Média dos testes de matemática por período e estado



FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

De acordo com as tabelas 17 e 18, nota-se que, durante o período pesquisado, o percentual de alunos que trabalha variou entre 12% e 16,8%. Por meio destas tabelas, observa-se ainda, nos três primeiros períodos da pesquisa, um percentual de alunos que trabalham por volta de 16%. Este percentual apresentou uma diminuição entre os anos de

2000 e 2002 para um valor de aproximadamente 12%. Contudo, no último período da pesquisa, o percentual de alunos que trabalham voltou a aumentar, chegando a aproximadamente 15%.

TABELA 17- Percentual de alunos que trabalham por período para a disciplina de português

Trabalha	Período					
	abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
não	5417	5538	7551	8012	6689	4970
	83,6%	83,6%	83,2%	88,0%	87,1%	84,7%
sim	1066	1083	1529	1090	994	900
	16,4%	16,4%	16,8%	12,0%	12,9%	15,3%
total	6483	6621	9080	9102	7683	5870
	100%	100%	100%	100%	100%	100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

TABELA 18- Percentual de alunos que trabalham por período para a disciplina de matemática

Trabalha	Período					
	abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
não	5445	5596	7592	8007	6702	4978
	83,5%	83,6%	83,2%	88,0%	87,1%	84,6%
sim	1078	1098	1532	1088	995	908
	16,5%	16,4%	16,8%	12,0%	12,9%	15,4%
total	6523	6694	9124	9095	7697	5886
	100%	100%	100%	100%	100%	100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

As tabelas 19 e 20 apresentam, respectivamente para as disciplinas de português e matemática, a distribuição dos alunos por sexo entre abril de 1999 e novembro de 2003. Durante este período, considerando ambas as disciplinas, o percentual de mulheres na amostra estudada foi superior ao dos homens em aproximadamente 3% nos dois primeiros períodos da pesquisa e em torno de 11% no último período.



TABELA 19- Distribuição dos alunos por sexo e por período para a disciplina de português

Sexo	Período					
	abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
masculino	3529	3374	4489	4203	3477	2442
	48,4%	48,6%	47,5%	45,4%	44,9%	44,2%
feminino	3762	3567	4956	5053	4265	3089
	51,6%	51,4%	52,5%	54,6%	55,1%	55,8%
total	7291	6941	9445	9256	7742	5531
	100%	100%	100%	100%	100%	100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

TABELA 20- Distribuição dos alunos por sexo e por período para a disciplina de matemática

Sexo	Período					
	abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
masculino	3551	3397	4514	4218	3499	2454
	48,5%	48,5%	47,5%	45,4%	44,9%	44,1%
feminino	3776	3610	4986	5064	4289	3105
	51,5%	51,5%	52,5%	54,6%	55,1%	55,9%
total	7327	7007	9500	9282	7788	5559
	100%	100%	100%	100%	100%	100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Em ambas as disciplinas, conforme as tabelas 21 e 22, o percentual de alunos que já repetiram de ano foi de aproximadamente 45% nos dois primeiros períodos da pesquisa correspondentes à quarta série do ensino fundamental. Este percentual sofreu uma queda, atingindo um valor em torno de 38% nos dois últimos períodos da pesquisa.

TABELA 21- Percentual de alunos que repetiram de ano por período para a disciplina de português

Repetiu	Período					
	abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
não	3390	3455	5039	5182	4723	3630
	54,7%	54,6%	56,0%	56,9%	62,0%	61,7%
sim	2803	2868	3955	3931	2897	2253
	45,3%	45,4%	44,0%	43,1%	38,0%	38,3%
total	6193	6323	8994	9113	7620	5883
	100%	100%	100%	100%	100%	100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

TABELA 22- Percentual de alunos que repetiram de ano por período para a disciplina de matemática

Repetiu	Período					
	abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
não	3415	3498	5052	5171	4733	3636
	54,8%	54,7%	55,9%	56,8%	62,0%	61,6%
sim	2820	2897	3986	3937	2899	2264
	45,2%	45,3%	44,1%	43,2%	38,0%	38,4%
total	6235	6395	9038	9108	7632	5900
	100%	100%	100%	100%	100%	100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Quanto ao nível 3, a TAB. 23 apresenta os percentuais referentes aos índices de conservação por valor ajustado e por período de realização da pesquisa. Como só existe questionário da escola em novembro de 1999, os índices de infra-estrutura de abril de 1999 foram considerados os mesmos estimados para novembro de 1999. Por meio desta tabela, observa-se uma diminuição do percentual do índice de conservação nos valores mínimo (0) e intermediários (1 e 2) a partir de nov/00. Para o percentual de escolas com infra-estrutura mínima (0), vale destacar a queda observada entre nov/99 (6,1%) e nov/00 (2,0%). Por outro lado, o percentual de escolas com índice de conservação máximo (3) sofre um aumento entre nov/99 (50,4%) e nov/01 (67,3%), estabilizando-se num nível mais elevado em nov/02 (65,2%) e nov/03 (66,7%).

TABELA 23- Percentual dos valores dos índices de infra-estrutura de conservação por período nas escolas para as disciplinas de português e de matemática

Infra-estrutura	Período					
	abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
0	8	8	3	4	4	5
	6,1%	6,1%	2,0%	2,6%	2,5%	3,2%
1	18	18	19	14	14	12
	13,7%	13,7%	12,9%	9,2%	8,7%	7,7%
2	39	39	37	32	38	35
	29,8%	29,8%	25,2%	20,9%	23,6%	22,4%
3	66	66	88	103	105	104
	50,4%	50,4%	59,9%	67,3%	65,2%	66,7%
total	131	131	147	153	161	156
	100%	100%	100%	100%	100%	100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

De acordo com a TAB. 24, para os anos de 1999 e 2002, menos de 10% das escolas apresentaram tamanhos médios de turma com até 20 alunos. Nos demais períodos da pesquisa esse percentual supera os 10%, chegando a atingir aproximadamente 16% em 2003. Um percentual em torno de pelo menos 40% das escolas possuem turmas com tamanhos médios de turma variando entre 26 e 35 alunos. Os menores percentuais de escolas com tamanhos médios de turma superiores a 35 alunos foram os observados nos anos de 2000 (16,0%) e de 2001 (13,5%). Por meio da TAB. 25, observa-se um tamanho médio das turmas por escola e por período de aproximadamente 30 alunos e, um desvio-padrão de aproximadamente 7, com algumas variações por período.

TABELA 24- Percentual de escolas, de acordo com o tamanho médio das turmas por período

Tamanho	Período					
	abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
≤20	13	13	22	18	14	26
	8,4%	8,4%	14,1%	10,5%	8,3%	15,6%
21 a 25	27	27	37	35	23	26
	17,4%	17,4%	23,7%	20,5%	13,7%	15,6%
26 a 30	44	44	47	63	45	33
	28,4%	28,4%	30,1%	36,8%	26,8%	19,8%
31 a 35	33	33	25	32	44	38
	21,3%	21,3%	16,0%	18,7%	26,2%	22,7%
>35	38	38	25	23	42	44
	24,5%	24,5%	16,0%	13,5%	25,00%	26,3%
total	155	155	156	171	168	167
	100%	100%	100%	100%	100%	100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

TABELA 25- Média e desvio padrão do tamanho médio das turmas por período

Variável	Estatísticas	Período					
		abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
Tamanho médio das turmas	média	30,43	30,43	28,08	28,14	30,55	29,46
	desvio-padrão	(7,37)	(7,37)	(7,40)	(6,17)	(7,64)	(8,66)

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Ainda com relação às variáveis do nível 3, podem ser observadas as tabelas 26 e 27. Essas tabelas informam, respectivamente, o percentual de escolas por tipo de sistema de recuperação adotado e pelo número de vezes em que o conselho de escola se reuniu.

Por meio da TAB. 26, observa-se um aumento entre nov/00 (1,4%) e nov/02 (88,2%) do percentual de escolas que não adotam sistema de recuperação, em contraste com os demais períodos, em que este percentual não ultrapassava 7%. Por sua vez, o percentual de escolas que adotam o sistema bimestral apresenta queda entre nov/00 (63,6%) e nov/02 (5,0%), voltando a se recuperar, parcial em nov/03 (49,7%). Analogamente, pode ser analisado o percentual de escolas que adotam o sistema semestral ou anual. Esse percentual sofre uma queda entre nov/00 (35,0%) e nov/02 (6,8%), alcançando em nov/03 (43,7%), nível que chega a ser superior àqueles do início do processo (entre 30,2 e 35,0%).

TABELA 26- Percentual do tipo de sistema de recuperação das escolas por período

Sistema de Recuperação	Período					
	abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
não adota	2	2	2	93	142	10
	1,7%	1,7%	1,4%	67,4%	88,2%	6,6%
bimestral	81	81	89	16	8	75
	68,1%	68,1%	63,6%	11,6%	5,0%	49,7%
semestral ou anual	36	36	49	29	11	66
	30,2%	30,2%	35,0%	21,0%	6,8%	43,7%
total	119	119	140	138	161	151
	100%	100%	100%	100%	100%	100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

Na TAB. 27, vale destacar a diminuição apresentada, em nov/03, do percentual de escolas em que o conselho se reuniu duas ou mais vezes (88,9%). O percentual de escolas em que o conselho não se reuniu foi zero nos períodos de nov/00 e nov/02. Ademais, este percentual não chega a alcançar nem 2% em nov/03 (1,4%). Nas escolas em que o conselho se reuniu uma vez, destaca-se a queda entre os períodos nov/00 (8,1%) e nov/01 (2,2%), seguida do aumento neste percentual entre nov/01(2,2%) e nov/03(9,7%).

TABELA 27- Percentual do número de vezes em que os conselhos de escolas se reuniram por período

Número de reuniões	Período					
	abr/99	nov/99	nov/00	nov/01	nov/02	nov/03
nenhuma	1	1	0	1	0	2
	1,0%	1,0%	0,0%	0,7%	0,0%	1,4%
uma	4	4	9	3	9	14
	4,1%	4,1%	8,1%	2,2%	6,1%	9,7%
duas ou mais	92	92	102	131	138	129
	94,9%	94,9%	91,9%	97,1%	93,9%	88,9%
total	97	97	111	135	147	145
	100%	100%	100%	100%	100%	100%

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

### 4.3- Ajuste de modelos hierárquicos longitudinais para o desempenho educacional de português e de matemática

#### 4.3.1- Passo a passo do ajuste dos modelos

De início, parte-se do pressuposto de que os desempenhos dos alunos têm estrutura quadrática, em ambas as disciplinas, o que vai ao encontro, conforme dito anteriormente, com os estudos de WILLETT (1989a, 1989b, 1997). Para o referido autor, se o atributo de interesse é a mudança sobre um longo período de tempo, talvez três ou quatro medidas de cada pessoa possam ser suficientes para captar a forma e a direção da mudança, se a trajetória de mudança individual não for muito complexa. Desta forma, o primeiro modelo ajustado para a disciplina de português foi o seguinte (este mesmo modelo foi analogamente estimado, como modelo base, para matemática):

#### Nível 1: Intra-aluno

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(idad\_ci)_{tij} + \pi_{2ij}(idadci2)_{tij} + e_{tij}$$

$$t=0,1,\dots,t_{ij}$$

$$i=1,2,\dots,N_j$$

$$j=1,2,\dots,J$$

em que:

$Y_{tij}$  é o desempenho do aluno  $i$  na escola  $j$  no tempo  $t$ ;

$idad\_ci$  representa a idade do aluno  $i$  na escola  $j$  no tempo  $t$ , centralizada na idade ideal na quarta série (10 anos);

$idadci2$  é igual a  $idad\_ci * idad\_ci$ ;

$\pi_{0ij}$  é o *status* inicial do aluno  $i$  da escola  $j$ , isto é, o desempenho esperado para esse aluno no início da pesquisa;

$\pi_{1ij}$  é a taxa de crescimento instantânea para o aluno  $ij$  na sua idade em  $t$ ;

$\pi_{2ij}$  é a curvatura (ou aceleração) em cada trajetória de crescimento do desempenho escolar;

$e_{tij}$  é o erro aleatório de nível 1, considerado independente e com distribuição  $\sim N(0, \sigma^2)$ ;

A combinação das taxas  $(\pi_{1ij}, \pi_{2ij})$  vai ao encontro de uma estrutura esperada para uma curva de aprendizagem, se  $\pi_{1ij} > 0$  e  $\pi_{2ij} < 0$  – taxas de crescimento cada vez menores, tendendo à estabilidade no final do processo. Em geral, a taxa de crescimento de qualquer idade particular é a derivada primeira do modelo de crescimento:

taxa de crescimento na idade  $a$  no tempo  $t = \pi_{1ij} + 2\pi_{2ij}(idade\_ci)$

### **Nível 2: Inter-aluno**

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

$$\pi_{2ij} = \beta_{20j} + r_{2ij}$$

em que:

$\beta_{00j}$  representa o *status* inicial médio, ou seja, esperado na escola  $j$ ;

$r_{0ij}$  é erro aleatório de nível 2, considerado independente e com distribuição  $\sim N(0, \tau_{\pi 00})$ ;

$\tau_{\pi 00}$  representa a variância dos *status* iniciais entre os alunos da escola  $j$ ;

$\beta_{10j}$  é a taxa instantânea esperada de aprendizado na escola  $j$ ;

$r_{1ij}$  é erro aleatório de nível 2, considerado independente e com distribuição  $\sim N(0, \tau_{\pi 11})$ ;

$\tau_{\pi 11}$  representa a variância das taxas instantâneas de crescimento entre o aprendizado dos alunos da escola  $j$ ;

$\beta_{20j}$  é a aceleração esperada do aprendizado na escola  $j$ ;

$r_{2ij}$  é erro aleatório de nível 2, considerado independente e com distribuição  $\sim N(0, \tau_{\pi 22})$ ;

$\tau_{\pi 22}$  representa a variância das curvaturas do crescimento entre o aprendizado dos alunos da escola  $j$ ;

### Nível 3: Escola

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}$$

$$\beta_{20j} = \gamma_{200} + u_{20j}$$

em que:

$\gamma_{000}$  é o *status* inicial médio (esperado) de todas as escolas pesquisadas;

$u_{00j}$  é erro aleatório de nível 3, considerado independente e com distribuição  $\sim N(0, \tau_{\beta 00})$ ;

$\tau_{\beta 00}$  representa a variância dos *status* iniciais entre as escolas;

$\gamma_{100}$  é a taxa instantânea média (esperada) de todas as escolas pesquisadas;

$u_{10j}$  é erro aleatório de nível 3, considerado independente e com distribuição  $\sim N(0, \tau_{\beta 11})$ ;

$\tau_{\beta 11}$  representa a variância das taxas instantâneas de crescimento entre as escolas;

$\gamma_{200}$  é a curvatura média (esperada) de todas as escolas pesquisada;

$u_{20j}$  é erro aleatório de nível 3, considerado independente e com distribuição  $\sim N(0, \tau_{\beta 22})$ ;

$\tau_{\beta 22}$  representa a variância das curvaturas de crescimento entre as escolas;

Contudo, em ambas as disciplinas, o  $r_{2ij}$  e o  $u_{20j}$  não foram significativos, de forma que o modelo incondicional ou modelo base ajustados para as disciplinas de português e matemática foi o seguinte:

### Nível 1: Intra-aluno

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(idad\_ci)_{tij} + \pi_{2ij}(idadci2)_{tij} + e_{tij}$$

$t=0,1,\dots,t_{ij}$   
 $i=1,2,\dots,N_j$   
 $j=1,2,\dots,J$

### Nível 2: Inter-aluno

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

$$\pi_{2ij} = \beta_{20j}$$

### Nível 3: Escola

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}$$

$$\beta_{20j} = \gamma_{200}$$

A partir deste modelo base (modelo 1), cada nível teve suas variáveis rodadas separadamente e, posteriormente, as variáveis significativas foram unidas num modelo ajustado ao seu nível correspondente. Assim, foram ajustados os modelos, separadamente:

1. a nível escola (modelo 2);
2. a nível de características fixas do aluno (modelo 3);
3. a nível de características variáveis do aluno ao longo do tempo (modelo 4);
4. variáveis explicativas a nível da escola e a nível das características fixas do aluno, aquelas que foram significativas do modelo 2 e as do modelo 3 (modelo 5);
5. variáveis significativas do modelo 5 mais a variável trabalha (modelo 6);

No caso de  $x-1$  variáveis indicadoras, em que  $x$  representa o número de categorias da variável original, cada grupo dessas  $x-1$  variáveis indicadoras foram testadas juntamente, de acordo com o nível do modelo a que se referiam. Além disso, quando os resultados das variáveis indicadoras, correspondentes a determinada variável categórica, eram semelhantes, a variável categórica original foi re-agrupada em menos variáveis indicadoras.

Sobre a questão de períodos igualmente espaçados, SINGER & WILLET (2003) afirmam que esses períodos têm um certo apelo, porque dados igualmente espaçados oferecem balanço e simetria. Contudo, segundo os referidos autores, independentemente dos espaços entre os períodos serem iguais, o mais importante é coletar dados suficientes para fornecer uma visão razoável de cada trajetória de crescimento individual.

Então, para a definição do calendário letivo, considerou-se que o ano letivo contém 10 meses, levando-se em conta as férias. Como a pesquisa tem sua primeira rodada em primeiro de abril de 1999, esta será tomada como linha de base. A segunda rodada foi realizada em primeiro de novembro de 1999, oito meses após a primeira. E todas as rodadas seguintes foram realizadas com um intervalo de um ano. Para identificar cada período de acordo com os intervalos entre as rodadas e usando como base abril de 1999 e o calendário com 10 meses, as rodadas foram classificadas da seguinte forma:

1. Abr/99 = 0 – base;



2. Nov/99 = 0,8 – referente aos 08 meses entre as rodadas;
3. Nov/ 00 = 1,8 – referente aos 10 meses letivos entre as rodadas;
4. Nov/01 = 2,8 – referente aos 10 meses letivos entre as rodadas;
5. Nov/02 = 3,8 – referente aos 10 meses letivos entre as rodadas;
6. Nov/03 = 4,8 – referente aos 10 meses letivos entre as rodadas.

A seguir, a TAB. 28 mostra os resultados do ajuste do modelo incondicional ou modelo base (modelo 1) e para as disciplinas de português e de matemática.

#### **4.3.2- Ajuste do modelo incondicional e estimação do efeito-escola por disciplina**

Conforme o modelo incondicional da TAB. 28, o *status* inicial esperado para os alunos em abril de 1999, na disciplina de português, é de aproximadamente 50 pontos. A taxa instantânea esperada de crescimento indica que para cada unidade de acréscimo nesta taxa, de um período para o seguinte, os alunos têm um crescimento esperado de aproximadamente 2,6 (2,596) no seu desempenho em português. Contudo, considerando que o crescimento no desempenho não é linear, estimou-se a curvatura desse crescimento em  $-0,181$ .

Para a disciplina de matemática, o *status* inicial esperado para os alunos em abril de 1999 é de aproximadamente 50 (49,935) pontos. A taxa instantânea esperada de crescimento indica que para cada unidade de acréscimo nesta taxa, de um período para o seguinte, os alunos têm um crescimento esperado de aproximadamente 4,1 (4,071) no seu desempenho em matemática. Sabendo que o crescimento no desempenho não é linear, estimou-se, para a disciplina de matemática uma curvatura de  $-0,333$ .

TABELA 28- Resultados do ajuste do modelo incondicional por disciplina

Parâmetros	Português		Matemática	
	Estimativa	(Erro-padrão)	Estimativa	(Erro-padrão)
Parte fixa				
Intercepto	49,993	(0,300)	49,935	(0,346)
Idade centralizada na idade ideal da 4ª série	2,596	(0,089)	4,071	(0,077)
Idade centralizada na idade ideal da 4ª série ao quadrado	-0,181	(0,008)	-0,333	(0,008)
Parte Aleatória				
	Entre escolas			
$u_{0k}$	12,357	(1,588)	17,672	(2,150)
$u_{1k}$	0,734	(0,094)	0,435	(0,060)
	Inter-alunos			
$r_{0jk}$	58,915	(1,413)	57,892	(1,346)
$r_{1jk}$	1,211	(0,081)	1,497	(0,078)
	Intra-alunos			
$e_{0ijk}$	33,267	(0,277)	29,187	(0,243)
	Segundos de convergência	7 segundos	6 segundos	

FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

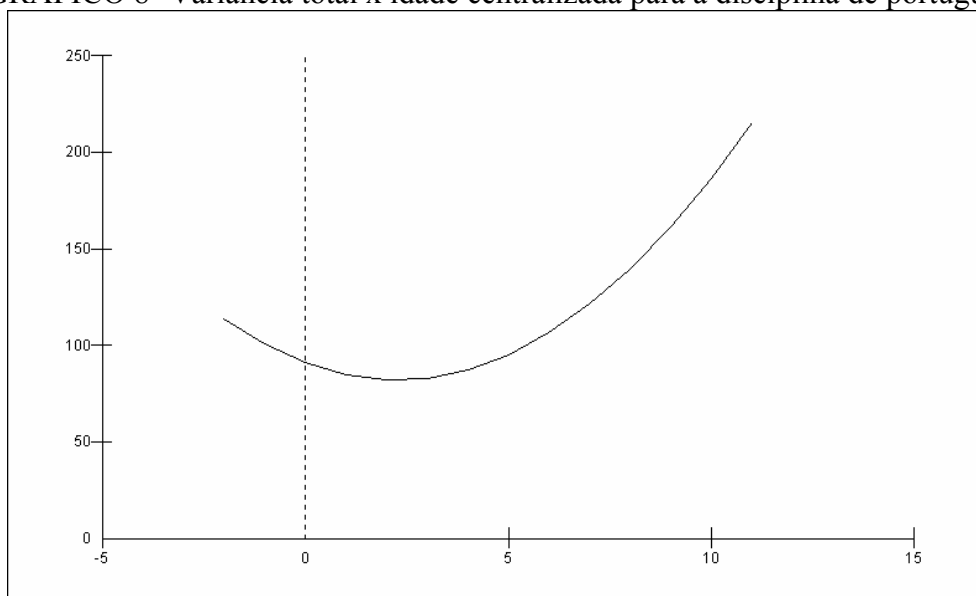
Quanto a parte aleatória do modelo, pode-se dizer que todos os três níveis (intra-aluno, inter-aluno e da escola) são significativos na explicação do desempenho escolar. A partir dessa estatísticas, é calculado o efeito-escola de acordo com os modelos hierárquicos longitudinais como pode ser visto a seguir.

Pode-se adiantar que em todos os modelos ajustados (modelos de 1 a 6), os três níveis foram significativos na explicação do desempenho escolar. Vale ressaltar, ainda, que foram poucos os segundos de convergência para os ajuste de todos os modelos em ambas as disciplinas, um indicativo, conforme JONES (2004), de ajustes adequados.

Ademais, para verificar o ajuste de modelos de crescimento quadrático incondicionais, RASBASH *et al* (2004) propõem a utilização do gráfico variância total estimada como uma função da idade centralizada. Segundo os referidos autores, o formato de um modelo incondicional de crescimento quadrático com ajuste adequado é aquele em que a variância total aumenta a taxas crescentes com o aumento da idade centralizada. Os GRAF. 8 e 9 apresentam a variância total estimada como uma função da idade, respectivamente, para as disciplinas de português e de matemática. Observa-se, em geral, um aumento a taxas

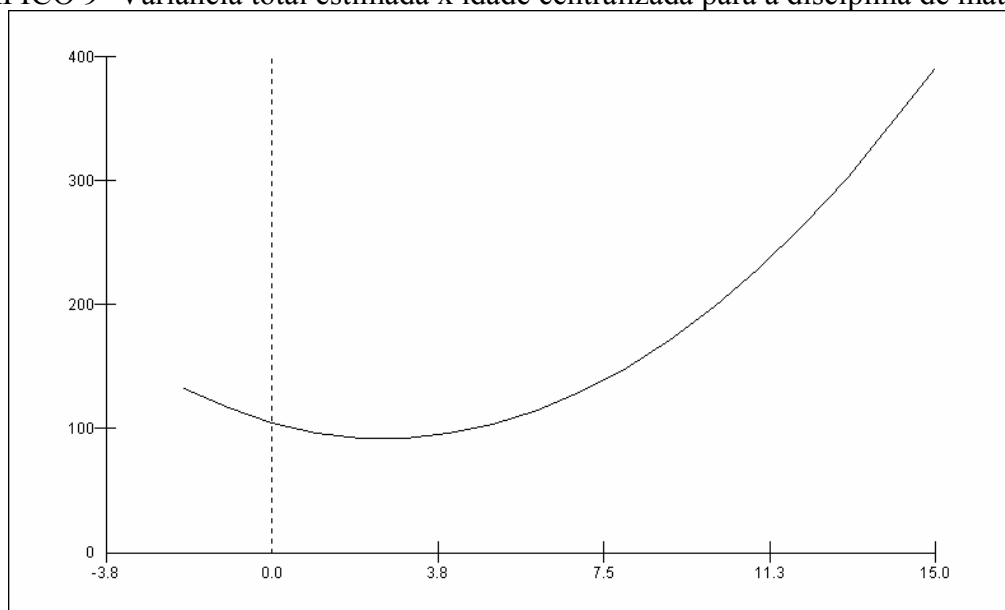
crecentes da variância total com o aumento da idade centralizada, indicando um ajuste adequado para os modelos incondicionais de português e de matemática.

GRÁFICO 8- Variância total x idade centralizada para a disciplina de português



FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

GRÁFICO 9- Variância total estimada x idade centralizada para a disciplina de matemática



FONTE: Elaboração própria com a utilização dos dados UFMG/CEDEPLAR (2005)

#### 4.3.2.1- Efeito-escola de acordo com os modelos hierárquicos longitudinais

##### a) Efeito-escola para a disciplina de português no período de abril/1999 a novembro/2003

A variabilidade entre as escolas explica 17,38% da variação no *status* inicial dos alunos em português:

$$\frac{\hat{\tau}_{\beta 00}}{\hat{\tau}_{\beta 00} + \hat{\tau}_{\pi 00}} \times 100 = \frac{12,357}{12,357 + 58,915} \times 100 = 17,38\%$$

A variabilidade entre as escolas explica 37,74% da variação na taxa instantânea de crescimento do desempenho dos alunos em português:

$$\frac{\hat{\tau}_{\beta 11}}{\hat{\tau}_{\beta 11} + \hat{\tau}_{\pi 11}} \times 100 = \frac{0,734}{0,734 + 1,211} \times 100 = 37,74\%$$

##### b) Efeito-escola para a disciplina de matemática no período de abril/1999 a novembro/2003

A variabilidade entre as escolas explica 23,39% da variação no *status* inicial dos alunos em matemática:

$$\frac{\hat{\tau}_{\beta 00}}{\hat{\tau}_{\beta 00} + \hat{\tau}_{\pi 00}} \times 100 = \frac{17,672}{17,672 + 57,892} \times 100 = 23,39\%$$

A variabilidade entre as escolas explica 22,52% da variação na taxa instantânea de crescimento do desempenho dos alunos em matemática:

$$\frac{\hat{\tau}_{\beta 11}}{\hat{\tau}_{\beta 11} + \hat{\tau}_{\pi 11}} \times 100 = \frac{0,435}{0,435 + 1,497} \times 100 = 22,52\%$$

### 4.3.3- Ajuste de modelos condicionais para as disciplinas de português e matemática

As TAB. 29 e 30 apresentam, respectivamente, os resultados do ajuste dos modelos hierárquicos longitudinais para a disciplina de português e de matemática. Por meio dessas tabelas, observa-se os resultados do ajuste dos modelos 2, 3, 4 e 5, assim como, os resultados do modelo 6, que considera variáveis explicativas dos níveis 2 e 3 somente no intercepto.

Por sua vez, a TAB. 31 mostra, para ambas as disciplinas, os resultados do modelo 6, reproduzidos das TAB. 29 e 30, e do modelo 7, no qual as variáveis explicativas significativas do intercepto no modelo 6 foram acrescentadas à taxa instantânea de crescimento.

TABELA 29- Resultados do ajuste de modelos condicionais para a disciplina de português

Português	Modelo 2 <sup>(1)</sup>	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Parâmetro	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)
Parte Fixa <sup>(2)</sup>					
Trabalha	-	-	-0,913 (0,106)	-	-0,688 (0,116)
Intercepto	47,247 (0,543)	49,768 (0,320)	50,289 (0,300)	46,777 (0,548)	46,898 (0,548)
feminino	-	3,093 (0,155)	-	3,121 (0,156)	3,082 (0,156)
repetiu	-	-2,932 (0,128)	-	-2,868 (0,129)	-2,841 (0,130)
mãe nunca freqüentou a escola	-	não significativa	-	-	-
mãe completou 4 <sup>a</sup> , mas não completou 8 <sup>a</sup> série do ensino fundamental	-	0,460 (0,145)	-	0,556 (0,146)	0,540 (0,147)
mãe tem pelo menos o ensino fundamental	-	0,701 (0,154)	-	0,780 (0,155)	0,778 (0,156)
percentual de prof.s com curso superior	0,027 (0,001)	-	-	0,027 (0,001)	0,027 (0,001)
não existe sistema de recuperação	não significativa	-	-	-	-
sistema semestral ou anual de recuperação	não significativa	-	-	-	-
conselho de escola não se reuniu	não significativa	-	-	-	-
conselho de escola se reuniu duas ou mais vezes	não significativa	-	-	-	-
tamanho médio das turmas por escola	-0,031 (0,007)	-	-	-0,035 (0,007)	-0,034 (0,007)
índice de infra-estrutura de conservação das salas de aula	não significativa	-	-	-	-
Mato Grosso do Sul ou Goiás	5,089 (0,566)	-	-	5,179 (0,534)	5,231 (0,532)
Rondônia, Pará ou Sergipe	2,014 (0,575)	-	-	2,258 (0,541)	2,272 (0,539)
Idade centralizada na idade ideal da 4 <sup>a</sup> série	2,304 (0,096)	2,657 (0,092)	2,541 (0,089)	2,397 (0,098)	2,362 (0,099)

TABELA 29- Resultados do ajuste de modelos condicionais para a disciplina de português  
(continuação)

Português	Modelo 2 <sup>(1)</sup>	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Parâmetro	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)
Idade centralizada na idade ideal da 4ª série ao quadrado	-0,168 (0,009)	-0,195 (0,009)	-0,176 (0,009)	-0,182 (0,010)	-0,176 (0,010)
Parte Aleatória					
Entre escolas					
$u_{0k}$	12,227 (1,598)	10,284 (1,396)	12,217 (1,580)	10,688 (1,462)	10,645 (1,458)
$u_{1k}$	0,885 (0,112)	0,684 (0,091)	0,705 (0,091)	0,806 (0,106)	0,811 (0,107)
Inter-alunos					
$r_{0jk}$	57,404 (1,440)	49,816 (1,453)	58,093 (1,440)	49,122 (1,487)	48,554 (1,485)
$r_{1jk}$	1,241 (0,085)	1,223 (0,088)	1,167 (0,083)	1,226 (0,092)	1,193 (0,092)
Intra-alunos					
$e_{0ijk}$	32,542 (0,282)	33,340 (0,305)	33,097 (0,283)	32,468 (0,307)	32,398 (0,309)
Segundos de convergência	≈ 7 segundos	≈ 7 segundos	≈ 7 segundos	≈ 7 segundos	≈ 7 segundos

FONTE: Elaboração própria com a utilização dos dados do UFMG/CEDEPLAR (2005)

NOTA: (1) O programa utilizado para o ajuste dos modelos dessa tese, *MLwin* v. 2.02, trabalha na presença de dados perdidos (*missings*), ignorando-os automaticamente.

(2) Variáveis indicadoras omitidas: a) mãe começou a estudar, mas não completou a série do ensino fundamental; b) conselho de escola se reuniu uma vez; c) a escola adota sistema bimestral de recuperação da aprendizagem; d) escolas pertencentes ao estado de Pernambuco.

TABELA 30 - Resultados do ajuste de modelos condicionais para a disciplina de matemática

Matemática	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Parâmetro	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)
Parte Fixa					
trabalha	-	-	-0,445 (0,100)	-	-0,304 (0,119)
Intercepto	49,165 (0,579)	51,958 (0,342)	50,227 (0,342)	50,913 (0,567)	51,004 (0,567)
feminino	-	-1,246 (0,143)	-	-1,307 (0,144)	-1,360 (0,145)
repetiu	-	-3,370 (0,119)	-	-3,090 (0,127)	-3,089 (0,127)
mãe nunca freqüentou a escola	-	não significativa	-	-	-
mãe completou 4ª, mas não completou 8ª série do ensino fundamental	-	0,936 (0,136)	-	0,907 (0,143)	0,933 (0,144)
mãe tem pelo menos o ensino fundamental	-	1,347 (0,145)	-	1,327 (0,150)	1,338 (0,152)
percentual de prof.s com curso superior	0,017 (0,001)	-	-	0,013 (0,001)	0,013 (0,001)
não existe sistema de recuperação	2,835 (0,091)	-	-	2,132 (0,087)	2,131 (0,088)

sistema semestral ou anual de recuperação	não significativa	-	-	-	-
conselho de escola não se reuniu	não significativa	-	-	-	-
conselho de escola se reuniu duas ou mais vezes	não significativa	-	-	-	-
tamanho médio das turmas por escola	-0,039 (0,007)	-	-	-0,037 (0,008)	-0,037 (0,008)
índice de infra-estrutura de conservação das salas de aula	não significativa	-	-	-	-
Mato Grosso do Sul ou Goiás	5,411 (0,616)	-	-	5,043 (0,568)	5,036 (0,567)
Rondônia, Pará ou Sergipe	1,469 (0,628)	-	-	1,458 (0,577)	1,471 (0,576)
Idade centralizada na idade ideal da 4ª série	2,835 (0,091)	4,116 (0,081)	4,016 (0,078)	3,109 (0,095)	3,089 (0,095)
Idade centralizada na idade ideal da 4ª série ao quadrado	-0,229 (0,009)	-0,345 (0,009)	-0,331 (0,008)	-0,256 (0,010)	-0,253 (0,010)
Parte Aleatória					
Entre escolas					
$u_{0k}$	12,966 (1,663)	13,450 (1,725)	17,002 (2,085)	10,209 (1,399)	10,115 (1,388)
$u_{1k}$	0,575 (0,079)	0,414 (0,060)	0,431 (0,060)	0,505 (0,073)	0,503 (0,073)
Inter-alunos					
$r_{0jk}$	45,346 (1,289)	47,269 (1,351)	57,630 (1,378)	39,114 (1,355)	38,881 (1,363)
$r_{1jk}$	1,345 (0,087)	1,506 (0,085)	1,525 (0,081)	1,297 (0,093)	1,284 (0,093)
Intra-alunos					
$e_{0ijk}$	28,612 (0,274)	29,256 (0,268)	28,787 (0,247)	28,692 (0,301)	28,744 (0,304)
Segundos de convergência	≈ 6segundos	≈ 6segundos	≈ 7 segundos	≈ 6segundos	≈ 6segundos

FONTE: Elaboração própria com a utilização dos dados do UFMG/CEDEPLAR (2005)

Por meio da TAB. 31, observa-se que, na disciplina de português, se o aluno for do sexo feminino, controlando pelas demais variáveis explicativas, ele terá um aumento esperado de 2,351 em seu status inicial e de 0,247 em sua taxa instantânea de crescimento.

Em português, o fato do aluno já ter repetido de ano, controlando pelas demais variáveis explicativas, faz com que haja uma diminuição esperada de 3,897 em seu status inicial. Por outro lado, sua taxa instantânea de crescimento apresentará um aumento esperado de 0,300.

Controlando pelas demais variáveis explicativas, se a mãe tiver completado a 4ª série do ensino fundamental, mas não tiver conseguido completar a 8ª série, o status inicial em português de seu filho terá um crescimento esperado de 0,552. Contudo, a taxa instantânea de crescimento do desempenho de seu filho não será alterada significativamente.

Resultado semelhante foi obtido com os filhos das mães que chegaram a completar pelo menos o ensino fundamental. Neste caso, o status inicial terá um aumento esperado de 0,781, enquanto a taxa instantânea de crescimento não se altera significativamente.

O aumento de 1 ponto no percentual de professores com curso superior em uma escola, controlando pelas demais variáveis explicativas, produz um aumento esperado de 0,017 no status inicial de seus alunos e de 0,003 na taxa instantânea de crescimento desses alunos.

TABELA 31- Resultados do ajuste de modelos condicionais mais complexos para as disciplinas de português e de matemática

Parâmetro	Português		Matemática	
	Modelo 6	Modelo 7	Modelo 6	Modelo 7
	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)
<b>Parte Fixa</b>				
Trabalha	-0,688 (0,116)	-0,674 (0,116)	-0,304 (0,119)	-0,312 (0,119)
Intercepto	46,898 (0,548)	48,749 (0,576)	51,004 (0,567)	50,380 (0,583)
Feminino	3,082 (0,156)	2,351 (0,222)	-1,360 (0,145)	-0,965 (0,217)
Repetiu	-2,841 (0,130)	-3,897 (0,229)	-3,089 (0,127)	-4,227 (0,234)
mãe não frequentou a escola	não significativa	-	não significativa	-
mãe completou 4ª, mas não completou 8ª série do ensino fundamental	0,540 (0,147)	0,552 (0,147)	0,933 (0,144)	0,927 (0,144)
mãe tem pelo menos o ensino fundamental	0,778 (0,156)	0,781 (0,156)	1,338 (0,152)	1,341 (0,151)
percentual de prof.s com curso superior	0,027 (0,001)	0,017 (0,002)	0,013 (0,001)	0,021 (0,002)
não existe sistema de recuperação	não significativa	-	2,131 (0,088)	2,408 (0,217)
sistema semestral ou anual de recuperação	não significativa	-	não significativa	-
conselho de escola não se reuniu	não significativa	-	-	-
conselho de escola se reuniu duas ou mais vezes	não significativa	-	não significativa	-
tamanho médio das turmas por escola	-0,034 (0,007)	-0,032 (0,007)	-0,037 (0,008)	-0,034 (0,008)
índice de infra-estrutura de conservação das salas de aula	-	-	-	-
Mato Grosso do Sul ou Goiás	5,231 (0,532)	3,149 (0,641)	5,036 (0,567)	5,042 (0,561)
Rondônia, Pará ou Sergipe	2,272 (0,539)	2,195 (0,537)	1,471 (0,576)	1,511 (0,570)
Idade centralizada na idade ideal da 4ª série	2,362 (0,099)	1,859 (0,118)	3,089 (0,095)	3,327 (0,105)



TABELA 31- Resultados do ajuste de modelos condicionais mais complexos para as disciplinas de português e de matemática (continuação)

Parâmetro	Português		Matemática	
	Modelo 6	Modelo 7	Modelo 6	Modelo 7
	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)	Estimativa (erro-padrão)
feminino	-	0,247 (0,054)	-	-0,134 (0,055)
repetiu	-	0,300 (0,055)	-	0,336 (0,057)
mãe não freqüentou a escola	não significativa	-	não significativa	-
mãe completou 4ª, mas não completou 8ª série do ensino fundamental	-	não significativa	-	não significativa
mãe tem pelo menos o ensino fundamental	-	não significativa	-	não significativa
percentual de prof.s com curso superior	-	0,003 (0,001)	-	-0,003 (0,001)
não existe sistema de recuperação	-	-	-	-0,116 (0,055)
sistema semestral ou anual de recuperação	não significativa	-	não significativa	-
conselho de escola não se reuniu	não significativa	-	não significativa	-
conselho de escola se reuniu duas ou mais vezes	não significativa	-	não significativa	-
tamanho médio das turmas por escola	-	não significativa	-	não significativa
Mato Grosso do Sul ou Goiás	-	0,774 (0,143)	-	não significativa
Rondônia, Pará ou Sergipe	-	não significativa	-	não significativa
Idade centralizada na idade ideal da 4ª série ao quadrado	-0,176 (0,010)	-0,211 (0,011)	-0,253 (0,010)	-0,268 (0,012)
Parte Aleatória				
Entre escolas				
$u_{0k}$	10,645 (1,458)	9,203 (1,290)	10,115 (1,388)	9,783 (1,351)
$u_{1k}$	0,811 (0,107)	0,641 (0,087)	0,503 (0,073)	0,511 (0,074)
Inter-alunos				
$r_{0jk}$	48,554 (1,485)	48,212 (1,477)	38,881 (1,363)	38,419 (1,351)
$r_{1jk}$	1,193 (0,092)	1,176 (0,091)	1,284 (0,093)	1,270 (0,093)
Intra-alunos				
$e_{0ijk}$	32,398 (0,309)	32,348 (0,308)	28,744 (0,304)	28,697 (0,303)
Segundos de convergência	≈ 7 segundos	≈ 7 segundos	≈ 6segundos	≈ 6segundos

FONTE: Elaboração própria com a utilização dos dados do UFMG/CEDEPLAR (2005)

Se o tamanho médio das turmas por escola aumenta em um aluno, controlando pelas demais variáveis explicativas, haverá uma diminuição esperada de 0,032 no status inicial dos alunos desta escola. Entretanto, o tamanho médio das turmas por escola não apresenta efeito significativo na taxa instantânea de crescimento desses alunos.

O aluno de escola pertencente aos estados do Mato Grosso do Sul e de Goiás, controlando pelas demais variáveis explicativas, apresenta um aumento esperado de 3,149 em seu status inicial de português e de 0,774 em sua taxa instantânea de crescimento.

Por sua vez, o aluno de escola localizada nos estados de Rondônia, Pará e Sergipe apesar de apresentar um aumento esperado de 2,195 no status inicial de português, não terá sua taxa instantânea de crescimento alterada significativamente.

Por fim, vale ressaltar a presença da variável trabalha no nível intra-aluno. Controlando por idade e pelo quadrado da idade do aluno, se o aluno trabalha, seu desempenho esperado em português sofrerá uma queda de 0,674.

No caso da disciplina de matemática, de acordo com a TAB. 31, se o aluno for do sexo feminino, controlando pelas demais variáveis explicativas, ele terá uma diminuição esperada de 0,965 em seu status inicial e de 0,134 em sua taxa instantânea de crescimento.

O fato do aluno já ter repetido de ano em matemática, controlando pelas demais variáveis explicativas, faz com que haja uma diminuição esperada de 4,227 em seu status inicial. Por outro lado, sua taxa instantânea de crescimento apresentará um aumento esperado de 0,336.

Controlando pelas demais variáveis explicativas, se a mãe tiver completado a 4ª série do ensino fundamental, mas não tiver conseguido completar a 8ª série, o status inicial em matemática de seu filho terá um crescimento esperado de 0,927. Contudo, a taxa instantânea de crescimento do desempenho de seu filho não será alterada significativamente.

Resultado semelhante foi obtido com os filhos das mães que chegaram a completar pelo menos o ensino fundamental. Neste caso, o status inicial em matemática terá um aumento esperado de 1,341, enquanto a taxa instantânea de crescimento não se altera significativamente.

O aumento de 1 ponto no percentual de professores com curso superior em uma escola, controlando pelas demais variáveis explicativas, produz um aumento esperado de 0,021 no status inicial de seus alunos e uma diminuição esperada de 0,003 na taxa instantânea de crescimento desses alunos na disciplina de matemática.

Se o tamanho médio das turmas por escola aumenta em um aluno, controlando pelas demais variáveis explicativas, haverá uma diminuição esperada de 0,034 no status inicial dos

alunos desta escola na disciplina de matemática. Entretanto, o tamanho médio das turmas por escola não apresenta efeito significativo na taxa instantânea de crescimento desses alunos.

O aluno de escola pertencente aos estados do Mato Grosso do Sul e de Goiás, controlando pelas demais variáveis explicativas, apresenta um aumento esperado de 5,042 em seu status inicial na disciplina de matemática, embora a sua taxa instantânea de crescimento não seja alterada significativamente.

Por sua vez, o aluno de escola localizada nos estados de Rondônia, Pará e Sergipe apesar de apresentar um aumento esperado de 1,511 no status inicial na disciplina de matemática, não terá sua taxa instantânea de crescimento alterada significativamente.

O aluno de escola em que não apresenta sistema de recuperação tem um aumento esperado em seu status inicial na disciplina de matemática de 2,408 pontos, enquanto sua taxa instantânea de crescimento sofre uma queda de 0,116.

Quanto ao nível intra-aluno, se o aluno trabalhar, controlando por sua idade e pelo quadrado de sua idade, seu desempenho esperado na disciplina de matemática cairá em aproximadamente 0,312 pontos.

Ainda de acordo com a TAB. 31, é possível observar que as variáveis indicadoras “*repetiu*” e “*escolas pertencentes aos estados do MS ou de GO*” parecem ter uma contribuição maior tanto na explicação do *status* inicial, quanto na explicação da taxa instantânea de crescimento. Especialmente, na disciplina de português, tendo controlado pelas variáveis explicativas, a taxa instantânea de crescimento foi estimada em 1,859. Vale destacar que se a escola pertencer aos estados de MS ou de GO, controlando pelas demais variáveis explicativas, essa taxa tem um aumento de 0,774, o que corresponde a aproximadamente 40% de 1,859.

## 5- CONSIDERAÇÕES FINAIS

### 5.1- Contribuições da tese

Nesta tese, o banco de dados longitudinais utilizado tem a vantagem de ser o primeiro banco de dados longitudinal brasileiro na área de avaliação educacional, podendo-se afirmar que este estudo será um dos primeiros, no Brasil, a apresentar estimativas mais adequadas do efeito-escola em contextos brasileiros com os piores rendimentos educacionais, a saber: Nordeste, Norte e Centro-Oeste. Vale lembrar que os dados longitudinais são essenciais para as avaliações da qualidade do ensino, uma vez que permitem a análise da trajetória da proficiência dos alunos ao longo do tempo e, portanto, a análise do crescimento (da tendência e da mudança) do desempenho no tempo.

A importância desta investigação pode ser justificada pela diminuição da pressão por acesso à escola em decorrência da redução na participação relativa da população em idade de cursar o ensino fundamental, fazendo com que o foco das atenções esteja direcionado para a qualidade da educação e, conseqüentemente, para a necessidade de um re-direcionamento das políticas públicas educacionais.

A primeira contribuição dessa tese diz respeito à importância da escola na explicação dos desempenhos escolares com a utilização de dados longitudinais. Conforme os modelos hierárquicos longitudinais, aproximadamente 17% (17,38%) da variação do *status* inicial dos alunos, em português, é explicada pela variabilidade do desempenho entre as escolas e, em torno de 38% (37,74%) da variação da taxa instantânea de crescimento é explicada pela variabilidade entre as escolas. Em matemática, aproximadamente 23% (23,39%) da variação do *status* inicial dos alunos é explicada pela variabilidade do desempenho entre as escolas e, em torno de 23% (22,52%) da variação da taxa instantânea de crescimento é explicada pela variabilidade entre as escolas. Perante estes resultados, pode-se dizer que a variabilidade no desempenho das escolas explica parte significativa da variabilidade do desempenho dos alunos.

Uma outra contribuição é a constatação da importância dos fatores individuais e familiares, apesar das dificuldades como o custo, a operacionalização e o controle de políticas públicas para a diminuição da dependência do resultado educacional dos alunos em relação a seus recursos familiares. Desta forma, embora a ênfase desta tese seja o efeito dos fatores escolares na proficiência educacional, faz-se necessário também mencionar os efeitos dos fatores individuais e familiares, de acordo com a TAB. 32, que apresenta um resumo dos

principais resultados do modelo final para ambas as disciplinas, facilitando o entendimento das considerações finais.

TABELA 32- Resumo dos principais resultados do modelo final para as disciplinas de português e de matemática

Principais resultados	Português		Matemática	
	<i>Status</i> inicial	Taxa instantânea de crescimento	<i>Status</i> inicial	Taxa instantânea de crescimento
Intercepto/taxa instantânea de crescimento	48,749	1,859	50,380	3,327
feminino	2,351	0,247	-0,965	-0,134
repetiu	-3,897	0,300	-4,227	0,336
mãe completou a 4 <sup>a</sup> , mas não completou a 8 <sup>a</sup> série do ensino fundamental	0,552	não significativa	0,927	não significativa
mãe tem pelo menos o ensino fundamental	0,781	não significativa	1,341	não significativa
Percentual de professores com curso superior	0,017	0,003	0,021	-0,003
Escola não tem sistema de recuperação	-	-	2,408	-0,116
Tamanho médio das turmas por escola e por série	-0,032	não significativa	-0,034	não significativa
Escolas pertencentes aos estados do MS ou de GO	3,149	0,774	5,042	não significativa
Escolas pertencentes aos estados de RO, PA ou SE	2,195	não significativa	1,511	não significativa

FONTE: TABELA 31

Neste sentido, em ambas as disciplinas, a educação da mãe é uma variável importante para um melhor desempenho escolar, essencialmente no que diz respeito ao *status* inicial de seus filhos, uma vez que a taxa instantânea de crescimento não foi alterada significativamente pelo nível educacional das mães. Se o nível de escolaridade das mães aumenta, seus filhos apresentam melhores desempenhos em ambas as disciplinas. Em geral, os resultados sobre o nível educacional das mães, nesta tese, estão de acordo com as conclusões de CAMARGO & BARROS (1991). Segundo os referidos autores, pais com menor nível educacional tendem a não perceberem claramente a importância da educação para os rendimentos futuros de seus filhos. O baixo nível educacional e a maior pobreza desencorajam esses pais a investirem na educação, devido ao elevado custo de oportunidade da retirada dos seus filhos do mercado de

trabalho. Este fato demonstra a grande estratificação educacional existente no Brasil, estando de acordo com os estudos de SILVA & HASENBALG (2001).

Se o aluno já repetiu de ano, seu escore é inferior aos alunos que nunca repetiram de ano, ainda que o aluno tenha um aumento esperado em sua taxa instantânea de crescimento. Este resultado indica que a repetência, apesar de parecer importante para o desempenho do aluno em termos de aprovação (SAEB, 2004), ainda não consegue fazer com que os repetentes tenham o mesmo desempenho dos não repetentes (CEDEPLAR, 2005; FRANCO *et al.*, 2002).

Em ambas as disciplinas, o fato de o aluno trabalhar faz com que ele tenha um desempenho educacional inferior, indo ao encontro do que afirmam CÉSAR & SOARES (2001), KASSOUF (2002) e SILVA & KASSOUF (2002). Daí, a importância de projetos de transferência de renda como o bolsa-família, por exemplo.

Por fim, mais uma contribuição é ressaltar a importância de políticas públicas que parecem mais eficazes ao combate à dependência da educação dos alunos de seus fatores familiares. Acredita-se que tais políticas, em geral, deveriam estar direcionadas a fatores escolares, como por exemplo, ao nível de escolaridade dos professores, ao tamanho da turma e ao enfoque regional.

Uma faceta desta contribuição está relacionada ao melhor desempenho dos alunos frequentadores das escolas em que o percentual de professores com curso superior é maior. Em português, o *status* inicial e a taxa instantânea de crescimento apresentam aumentos significativos, enquanto em matemática a taxa instantânea de crescimento tem sentido oposto. Apesar dessa diferença entre as disciplinas e da pequena contribuição do percentual de professores na taxa instantânea de crescimento, acredita-se na importância da quantidade de recursos aplicados na educação com vistas à elaboração e à viabilização de programas voltados para o aumento do nível educacional dos professores.

Outra faceta desta contribuição é a verificação da associação negativa entre o tamanho médio das turmas por escola e o status inicial dos alunos em ambas as disciplinas, embora o tamanho médio das turmas por escola não tenha sido significativo na explicação da taxa instantânea de crescimento. Este resultado corrobora com o que KRUEGER (2003) defende: as reduções do tamanho da turma tende a produzir um efeito positivo na proficiência, porque o tamanho da turma é o principal determinante dos gastos em educação. Assim sendo, a quantidade de recursos escolares tem importância para o desempenho escolar, uma vez que menores turmas significam maior número de professores e, portanto, maiores gastos em educação. Neste ponto, vale lembrar que, de acordo com KRUEGER (2003), os efeitos

positivos da redução do tamanho da classe na proficiência dos alunos podem ser pequenos, não generalizáveis entre as séries e obscurecidos por equações mal especificadas. Contudo, este autor ressalta que, mesmo esses efeitos sendo sutis, eles podem ser economicamente importantes.

A última faceta desta contribuição diz respeito ao melhor desempenho educacional em função das regiões. Isto indica que a política educacional parece ser a mesma, onde se necessitaria de políticas educacionais diferenciadas por estado ou regiões, uma vez que, segundo BUCHMANN & HANNUM (2001), os recursos básicos da escola têm mais importância na proficiência dos alunos em contextos onde há maiores desigualdades nos recursos educacionais.

Os resultados mostraram que tanto o status inicial, quanto a taxa instantânea de crescimento na disciplina de português aumentam, se o aluno pertence a uma escola localizada nos estados do Mato Grosso do Sul ou Goiás. Contudo, na disciplina de matemática, somente o status inicial aumenta, se o aluno pertence a uma escola localizada nestes estados. Caso o aluno estude em uma escola pertencente aos estados de Rondônia, Pará ou Sergipe, somente o status inicial do aluno aumentará em ambas as disciplinas, já que as taxas instantâneas de crescimento não sofrerão alterações significativas.

Desta forma, acredita-se que as políticas educacionais precisam, de início, conhecer as diferentes realidades e necessidades regionais, a fim de maximizar o sistema de distribuição de recursos escolares, na tentativa de tornar cada vez mais próximos os desempenhos educacionais dos alunos dos estados Pernambuco, de Rondônia, Pará, e Sergipe daqueles encontrados nos estados de Mato Grosso do Sul e de Goiás.

Em síntese, esta tese foi uma das primeiras a utilizar dados longitudinais educacionais do Brasil e uma das primeiras a estimar adequadamente o efeito-escola em contextos brasileiros, uma vez que o efeito-escola tem um caráter longitudinal. Conforme o efeito-escola estimado de acordo com os modelos hierárquicos longitudinais, a escola apresenta uma importância fundamental no desempenho educacional dos alunos. Contudo, não se pode negar a contribuição dos fatores familiares no resultado educacional de seus filhos. Sendo assim, um melhor desempenho educacional parece ser um *mix* de recursos familiares e escolares. Em termos de políticas públicas, acredita-se na maior eficácia daquelas voltadas à escola, levando em conta a quantidade e a distribuição dos recursos disponíveis, em função das diferentes realidades e necessidades regionais.

## 5.2- Limitações da tese e possibilidades de estudos futuros

Ao mesmo tempo em que os dados longitudinais utilizados, nesta tese, apresentam um grande diferencial em relação aos bancos de dados transversais existentes na área de educação, oferecem, por outro lado, limitações que dificultaram a análise do crescimento da aprendizagem e a identificação de fatores associados. Essas limitações, em geral, estão associadas à construção e à digitação dos questionários.

Quanto às limitações da construção do questionário, pode-se citar a falta de padronização dos mesmos, de maneira que as perguntas diferem de um ano para o outro, especialmente em 2001, ano em que estavam presentes perguntas sobre o processo escolar, ausentes nos demais períodos da pesquisa.

Outro ponto a ser colocando quanto à limitação da construção do questionário diz respeito à presença de questões mal formuladas, como, por exemplo, “Você leu alguma coisa neste ano?”. Também é digno de menção as alternativas de resposta muito agrupadas, por exemplo, aquelas sobre a posse de bens (nenhum, um, dois ou mais) dificultando a estimação de fatores latentes como o nível sócio-econômico.

A construção dos questionários, em relação aos recursos escolares, tem como limitação contemplar questões por meio das quais não foi possível estimar mais do que um fator latente a nível escolar.

No início do processo de digitação, não houve um controle adequado, sendo que alternativas de respostas não existentes contavam como opção de resposta. Para contornar esse problema, foi realizada uma “limpeza” nos questionários dos anos de 1999, 2000 e 2001, retirando alternativas não existentes como resposta; Contudo, foram detectados ainda problemas nos identificadores de alunos (no banco de dados “Cadastro”) e de escolas, o que se buscou solucionar, conforme descrito em anexo.

As datas de nascimentos constantes nos questionários e no cadastro não coincidem, sendo praticamente inviável a verificação dessa digitação, contando apenas com parte das imagens dos questionários e com problemas na abertura destas. Então, a opção escolhida foi trabalhar com a data de nascimento do cadastro, que se pressupunha mais confiável.

Outra limitação da digitação dos questionários é a falta de coincidência entre as variáveis turno e turma nos questionários e na ficha do aluno. O pressuposto adotado foi o de que as variáveis da ficha são mais confiáveis, uma vez que não foram preenchidas pelos alunos.



Perante essas limitações, foi possível ajustar somente um fator de infra-estrutura de conservação das salas de aula, uma vez que não se encontravam para todos os períodos as mesmas variáveis observadas para a construção de um fator latente.

Ademais, diante da falta de coincidência entre as variáveis turno e turma nos questionários e na ficha, assim como, do reduzido número de professores por série, optou-se por trabalhar em nível de escola e não de turma, criando a variável tamanho médio das turmas por série e por escola.

Além das limitações citadas, pode-se pensar ainda numa limitação metodológica, uma vez que a equalização dos dados com a utilização da Teoria da Resposta ao Item considera o mesmo aluno em séries diferentes como alunos diferentes, não captando, portanto, a dependência existente entre as observações de um mesmo indivíduo no tempo.

Uma outra limitação refere-se ao ajuste dos modelos com especificação teórica do modelo condicional mais simples do que aquela mostrada no capítulo da metodologia. Apenas o intercepto foi controlado por variáveis explicativas fixas do aluno em relação ao tempo e por variáveis explicativas das escolas. Houve a tentativa de controlar a taxa instantânea e a curvatura do crescimento por variáveis explicativas dos níveis 2 e 3, entretanto não foram obtidos bons resultados, de forma que a opção foi partir para o ajuste de um modelo mais parcimonioso, controlando apenas o intercepto.

Perante o exposto, com relação a possibilidades de estudos futuros, pode-se mencionar a adoção de variáveis explicativas do Censo Escolar em nível da escola na tentativa de se obter melhores resultados para modelos mais complexos.

Uma outra possibilidade de estudo futuro diz respeito à adoção de métodos, que consideram a dependência das observações individuais ao longo do tempo na estimação de escores equalizados, como o método testado por simulação no trabalho realizado por ANDRADE & TAVARES (2005).

## 6- 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.

ANDRADE, D.F.. *Plano da Amostra*. Amostragem realizada para a pesquisa Avaliação do Desempenho: fatores associados, 1999.

ANDRADE, D. F.. Comparando desempenhos de grupos de alunos por intermédio da teoria da resposta ao item. *Estudos em Avaliação Educacional*, 23, 31-69, 2001.

ANDRADE, D.F; TAVARES, H.R.; VALLE, R.C. Teoria da resposta ao item: conceitos e aplicações. Caxambu: Associação Brasileira de Estatística, 2000.

ANDRADE, D. F; TAVARES, H.R. Item response theory for longitudinal data: population parameter estimation. *Journal of Multivariate Analysis*, 2005. Disponível em [www.elsevier.com](http://www.elsevier.com). Acesso em 25 de jul. 2005.

ARAÚJO, C.H.; LUZIO, N. Educação e desigualdades regionais, 2004. Disponível em <http://www.inep.gov.br/>. Acesso em 02 de set. 2004.

BAKER, F.B. *The Basics of Item Response Theory*. USA: ERIC Clearinghouse on Assessment and Evaluation, 2001.

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.

BARTHOLOMEW, D.J.; STEELE F.; MOUSTAKI, I.; GALBRAITH, J.I. . *The analysis and interpretation of multivariate data for social scientists*. Boca Raton: Chapman & Hall/CRC, 2002.

BUCHMANN, C.; HANNUM, E. Education and stratification in developing countries: a review of theories and research. *Annual Review of Sociology*, v. 102, p.27-77, 2001.

CAMARGO, J.; BARROS, R.. As causas da pobreza no Brasil: porca miséria! In: IPEA. *Perspectivas da economia brasileira*. Brasília: IPEA, Cap. 27, p.525-544, 1991

CARVALHO, J.A .M., WONG, L.R. A window of opportunity: some demographic and socioeconomic implications of the rapid fertility decline in Brazil. Belo Horizonte: FACE/CEDEPLAR, 1995 (*Texto para discussão 91*).

CEDEPLAR- Centro de Desenvolvimento e Planejamento Regional. *Avaliação de Desempenho- fatores associados*. Relatório entregue ao INEP, 2005.

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

CERQUEIRA, C.A. Tipologia e características dos estabelecimento escolares brasileiros. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2004

COLEMAN, J.S., CAMPBELL, E.Q., HOBSON, C.J., McPARTLAND, J., MOOD, A M., WEINFELD, F.D.; YORK, R.L. 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*, v. 113, p.64-98, Feb.2003

FERRÃO, M. A.; BELTRÃO, K. I.; FERNANDES, C.; SANTOS, D.; SUÁREZ, M; ANDRADE, A. C. O SAEB- Sistema Nacional de Avaliação da Educação Básica: objetivos, características e contribuições na investigação da escola eficaz. In: *Revista Brasileira de Estudos de População*. Campinas: Associação Brasileira de Estudos Populacionais, v. 18, n.1/2, p. 111-130, jan./dez. 2001.

FERRÃO, M. A.; FERNANDES, C. O efeito-escola e a mudança- Dá para mudar? Evidências da Investigação Brasileira. *Revista Eletrônica Iberoamericana sobre Calidad em Educación*, v.1, n. 1, 2003

FRANCO, C.; BONAMINO, A. Iniciativas recentes de avaliação da qualidade da educação no Brasil. In: FRANCO, C. (org.) *Avaliação, ciclos e promoção na Educação*. Porto Alegre: ARTMED, 2001.

FRANCO, C.; MANDARINO, M.; ORTIGÃO, M. I. O projeto pedagógico e os resultados escolares. *Pesquisa e planejamento econômico*. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada, V. 32, n. 3, p.477-4493, dez. 2002.

GOLDSTEIN, H. *Multilevel Statistical Models*. 2<sup>nd</sup> edition. London: Edward Arnold, 1995.

HAMBLETON, R.K. Principles and selected applications of item response theory. In: LINN, R. L. *Educational Measurement*. 3ed. Phoenix: American Council on Education, Oryx Press, p.147-200, 1993.

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*. Federal Reserve Bank of New York, vol.4, (1), March, p. 11-28, 1998a.

HANUSHEK, E. A. . *The Evidence on class size*. Occasional Paper Number 98-1. W. Allen Wallis Instituto of Political Economy. University of Rochester, February, 1998b.

HANUSHEK, E. A.; KAIN, J.F.; MARKMAN, J.M.; RIVKIN, S.G. Does peer ability affect student achievement?, National Bureau of Economic Research , 2001.

HANUSHEK, E. A. Publicly provided education. In: AUERBACH , A. J., FELDSTEIN, M. *Handbook of Public Economics*. Stanford: Elsevier Science B.V., v. 4, p. 2046-2107, 2002a.

HANUSHEK, E. A Evidence, politics and the class size debate. In: LAWRENCE, M; ROTHSTEIN, R., eds., *The class size debate*. Economic policy Instituto. Washington, 2002b.

HANUSHEK, E. A. The failure of input-based schooling policies. *The Economic Journal*, v. 113, p.64-98, Feb.2003.

HENRYSSON, S. Gathering, Analyzing, and Using Data on Test Items. In: Thorndike R. L. *Educational Measurement*. Second Edition. Washington: American Council on Education, 1971.

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Avaliação internacional mostra desempenho de alunos de 41 países*, 2003. Disponível em <http://www.inep.gov.br>. Acesso em 15 de jul. 2004.

JONES, A. A review of random effects models in Mlwin (version 2.0). School of Environmental Sciences, University of East Anglia, UK., 2004. Disponível em <http://www.multilevel.ioe.ac.uk> . Acesso em 8 abril de 2005.

KASSOUF, A. L. O efeito do trabalho infantil para o rendimento dos jovens, controlando o *background* familiar. In: *XIII Encontro da Associação Brasileira de Estudos Populacionais*. Ouro Preto: ABEP, 2002.

KOLEN, M. J., BRENNAN, R.L. *Test Equating – Methods and Practices*. New York: Springer, 1995.

KREGER, A. B. Reassessing the view that american schools are broken. *Economic Policy Review*. Federal Bank of New York, March, vol.4 (1), p.29-46, 1998.

KREGER, A. B. A response to Eric Hanushek’s “Evidence, politics and the class size debate”. In: LAWRENCE, M; ROTHSTEIN, R. eds. *The class size debate*. Economic policy Instituto. Washington , 2002.

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

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

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

LEE, V.E . *O melhor método para abordar questões de pesquisa multiníveis*. Tradução da Palestra ministrada na Faculdade de Educação da Universidade Federal de Minas Gerais, 2004.

LORD, F. M. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale: Lawrence Erlbaum Associates, Inc., 1980.

LÜDKE, M. Evoluções em avaliação. In: FRANCO, C. (org.) *Avaliação, ciclos e promoção na Educação*. Porto Alegre: ARTMED, 2001.

MISLEVY, R. J. & BOCK, R. D. *BILOG: Item Analysis and Test Scoring with Binary Logistic Models*. Chicago: Scientific Software, Inc., 1990.

MURILLO, F. J. Los modelos jerárquicos lineares aplicados a la investigación sobre eficacia escolar. *Revista de investigación educativa*, 17 (2), 453-460, 1999.

PISA- Programa Internacional de Avaliação de Estudantes. PISA 2000: Relatório Nacional. Brasília, dez.2001. Disponível em <http://www.inep.gov.br>. Acesso em 15 de jul. 2004.

RASBASH, J.; STEELE, F.; BROWNE, W.; PROSSER, B. *A User's Guide to Mlwin*. Institute of Education. London: University of London, 2004.

RAUDENBUSH, S.W.; BRYK, A S.; CHEONG, Y.F.; CONGDON, R. T. Jr. *HLM5: Hierarchical linear and nonlinear modeling*. Lincolnwood: SSI, 1992.

RAUDENBUSH, S.W.& BRYK, A S. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Second Edition. London: Sage Publications, 2002.

RIANI, J.L.R. Determinantes do resultado educacional no Brasil: família, perfil escolar dos municípios e dividendo demográfico numa abordagem hierárquica e espacial. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2005

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.

ROSENBAUM, J.E. Two revolutions in educational attainment research over the past 30 years: their impact on public understanding and social policy. Institute for Policy Research. Working paper 02-21, 2002.

SAEB- Sistema Nacional de Avaliação da Educação Básica. *Relatório SAEB99*. Disponível em <http://www.inep.gov.br>. Acesso em 12 de mai. 2004

SAEB- Sistema Nacional de Avaliação da Educação Básica. *Relatório SAEB 2001-Matemática*. Brasília, 2002. Disponível em <http://www.inep.gov.br>. Acesso em 12 de mai. 2004

SAEB- Sistema Nacional de Avaliação da Educação Básica. *Resultados do SAEB 2003*. Brasília, 2004. Disponível em <http://www.inep.gov.br>. Acesso em 23 de set.. 2004.

SILVA, N. V.; HASENBALG. *Recursos familiares e transições educacionais*. Versão preliminar apresentada no Workshop de Demografia da Educação. Associação Brasileira de Estudos Populacionais, ABEP, Salvador, jun. 2001.

SILVA, N.D.V.; KASSOUF, A. L.. O trabalho e a escolaridade dos brasileiros jovens. In: *XIII Encontro da Associação Brasileira de Estudos Populacionais*. Ouro Preto: ABEP, 2002.

SINGER, J.D.; WILLETT, J. B. *Applied Longitudinal Data Analysis: Modeling change and event occurrence*. New York: OXFORD, 2003.

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.

TOIT, MATHILDA du, editor. *ITR from SSI: BILOG-MG, MULTIOLOG, PARSCALE, TESTFACT*. Lincolnwood: SSI, 2003.

TOOD, P. E., WOLPIN, K.I. On the specification and estimation of the production function for cognitive achievement . *The Economic Journal*, v. 113, p.F3-F33, Feb.2003.

TORRECILLA, F. J.M. La red iberoamericana de investigación sobre eficacia escolar y mejora de la escuela: un ejemplo de colaboración internacional en investigación educativa. *Revista Española de Educación Comparada*, 6, 67-85, 2000.

VALLE, R. C. *Teoria da resposta ao item*. Dissertação (Mestrado em Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1999.

ZIMOWSKI, M. F.. MURAKI, E. MISLEVY, R. J. & BOCK, R. D. *BILOG-MG: Multiple-Group IRT Analysis and Test Maintenance for Binary Items*. Chicago: Scientific Software, Inc., 1996.

WEISS, D.J. & YOES, M. E. Item Response Theory. In: Hambleton, R. K. & Zaal, J. N. *Advances in Educational and Psychological Testing: Theory and Applications*. Boston: Kluwer Academic Publishers, 1991.

WILLETT, J. B. Questions and Answers in the Measurement of Change. In: ROTHKOPF, E. Z. (Ed.). *Review of Research in Education*, 15, 345-422, 1989a. Disponível em <http://www.gseacademic.harvard.edu/~willetjo/paperdownloads.htm>. Acesso em fev.2005.



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*, 49, 587-602, 1989b. Disponível em <http://www.gseacademic.harvard.edu/~willetjo/paperdownloads.htm>. Acesso em fev.2005.

WILLETT, J.B. Measurement of change. In: HUSEN, T. & POSTLEHWAITE, T. N. (Eds). *International Encyclopedia of Education*, 2<sup>nd</sup> edition. Oxford, UK: Elsevier Science Press, 671-678, 1994. Disponível em <http://www.gseacademic.harvard.edu/~willetjo/paperdownloads.htm>. Acesso em fev.2005.

WILLETT, J.B. Measuring change: what individual growth modeling buys you. In: AMSEL, E. & RENNINGER, K.A. (Eds). *Change and development: issues of theory, method and application*. Mahwah, NJ: Lawrence Erlbaum Associates, Chapter 11, 213-243, 1997.

WILLETT, J.B.; BUB, K.L. *Latent growth curve analysis*. Harvard University Graduate School of Education, p.1-23, 2004. Disponível em <http://www.gseacademic.harvard.edu/~willetjo/paperdownloads.htm>. Acesso em fev.2005.

WILLMS, J.D. Monitoring school performance: a non-technical guide for educational administrators. Lewes: Falmer press, 1992.

WILLMS, J. D.; SOMERS, M-A. *Schooling outcomes in Latin America*. A report for UNESCO, 1999.

WILLMS, J. D. Monitoring school performance for “standards-based reform”. *Evaluation and Research in Education*, v.14, n. 3 e 4, p. 237-253, 2000.

## ANEXO I

### Correção dos identificadores dos alunos

Conforme mencionado anteriormente, o banco de dados apresenta erros em variáveis importantes tais como identificação do aluno, identificação da escola, etc. Como a re-digitação dos dados é inviável, a única solução possível seria a correção destas variáveis. Entretanto, algumas variáveis são muito difíceis de serem corrigidas, pois elas possuem registros não coincidentes em três bancos de dados diferentes (prova, questionário e ficha A), dificultando, a criação de um critério aceitável para a correção dos dados. Como exemplo, pode-se citar as diferenças entre as variáveis turno e turma presentes nos bancos de dados das provas, dos questionários e das fichas A. Neste caso, o pressuposto é de que os dados referentes a turno e turma mais confiáveis são os provenientes da ficha A, uma vez que esta não é preenchida pelo aluno.

Entretanto, no caso da variável identificador da escola, foi possível estabelecer um critério de escolha. Como o seu identificador consta em três bancos de dados (do cadastro, da prova e do questionário) e coincide em dois deles (do cadastro e da prova), optou-se por trabalhar com identificador de escola do banco de dados da prova.

Vale ressaltar que uma das variáveis mais importantes – identificador do aluno - pode ser corrigida, visto que no banco de dados “Cadastro do aluno” consta o nome do aluno e sua data de nascimento, possibilitando detectar alunos com mais de um número identificador.

Assim sendo, procedeu-se à correção dos identificadores dos alunos. Primeiramente, com o auxílio de programas de computador, detectou-se duplicações no cadastro, gerando três arquivos com as seguintes características:

1. Alunos com identificadores diferentes, mas com nome, data de nascimento e escolas iguais – arquivo 1;
2. Alunos com identificadores diferentes, porém apresentando data de nascimento e escolas iguais e nomes semelhantes – arquivo 2;
3. Alunos com identificadores diferentes, mas com nomes e escolas iguais e datas semelhantes, ou não – arquivo 3.

A seguir, utilizando os arquivos gerados anteriormente, foram criados três programas em *SPSS* (correção 1, correção 2 e correção 3) para a correção dos identificadores nos bancos de dados dos alunos por ano. Assim, em cada banco de dados foram executados três programas, sendo que estes programas foram criados com o auxílio de inspeção visual.

Assim, para a realização da correção de números identificadores para o mesmo aluno foi utilizado o banco do cadastro dos alunos, que contém as seguintes variáveis: número identificador do aluno, nome do aluno, data de nascimento e escola.

A primeira forma de detectar tais alunos no banco foi por meio de um programa, denominado correção 1, para identificar aqueles com mesmo nome, mesma data de nascimento e mesma escola. Assim foram identificados, por meio do cadastro, 1.389 alunos com identificadores diferentes a serem corrigidos, conforme TAB. 1.1.

TABELA 1.1- Número de alunos com identificador diferente e número de identificadores (casos) diferentes a serem corrigidos

Estatísticas	Arquivo 1	Arquivo 2	Arquivo 3
	Nomes, datas e escolas iguais	Datas e escolas iguais Nomes semelhantes	Nomes e escolas iguais Datas semelhantes
<b>Nº de alunos</b>	1.389	676	489
Percentual	3,6%	1,8%	1,3%
<b>Total</b>	<b>38.317</b>	<b>38.317</b>	<b>38.317</b>
<b>Nº de casos</b>	2.967	1.485	999
Percentual	7,7%	3,9%	2,6%
<b>Total</b>	<b>38.317</b>	<b>38.317</b>	<b>38.317</b>

FONTE: Elaboração própria com a utilização dos dados Cedeplar/Inep (2005)

Um segundo programa, chamado correção 2, foi adotado para identificar alunos com datas de nascimento iguais e que estudavam na mesma escola. Para a criação desse programa, foi necessária uma inspeção visual para detectar as seguintes situações:

1. alunos com nomes praticamente idênticos, diferenciados apenas por acento, til, cedilha, de, pequenos erros de digitação com trocar s por z, etc;
2. alunos com um sobrenome diferente em dois nomes, com a mesma data de nascimento e na mesma escola;
3. alunos com nomes praticamente idênticos, na mesma escola, mas com data de nascimento diferentes;
4. nome raro e um sobrenome diferente, com a mesma data de nascimento e escola;
5. nome raro, somente uma data de nascimento preenchida, na mesma escola;
6. sobrenome abreviado, mesma data de nascimento, mesma escola;
7. sobrenome abreviado, somente uma data de nascimento preenchida, mesma escola;
8. alunos com um sobrenome diferente em três nomes;
9. sobrenome abreviado, somente uma data de nascimento preenchida, mesma escola;
10. nomes incompletos, como dois sobrenomes a menos em quatro nomes;

11. nomes não muito freqüentes ou extensos, mas praticamente idênticos ou abreviado, datas de nascimento diferentes, mesma escola.

Desta forma, por meio do programa correção 2 foram identificados 676(1,8%) dos alunos para serem corrigidos.

Um terceiro programa, denominado correção 3, foi utilizado para detectar alunos com nomes exatamente iguais, datas de nascimento diferentes e estudando na mesma escola, identificando-se duplicidade ou até mesmo três identificadores para um mesmo aluno quando:

1. uma das datas de nascimento não foi preenchida;
2. somente o dia do nascimento não coincide;
3. troca entre dia e mês de nascimento;
4. dia e ano de nascimento diferentes;
5. mês e ano de nascimento diferentes;
6. datas de nascimento como 1903, etc.

Voltando à TAB. 1.1, obteve-se um total de 489 (1,3%) alunos com identificadores diferentes que podem ser corrigidos pelo programa correção 3. Por meio desta tabela, somando-se os percentuais dos alunos que aparecem no cadastro com mais de um identificador, tem-se que 6,7% dos alunos presentes no cadastro apresentam identificadores diferentes, que podem ser corrigidos com base nos três programas citados acima.

Ademais, de acordo com a TAB. 1.2, após a execução do programa correção 1, por meio de uma inspeção visual, foram identificados dois alunos duplicados em 2000, quinze em 2002 e oito em 2003. Quanto aos resultados obtidos após a execução do programa correção 2 foram detectados um aluno duplicado em abril de 1999, um em 2000, dez em 2002 e dois em 2003. Depois da execução do programa correção 3, encontra-se um aluno duplicado em 2000, quatro em 2001, nove em 2002 e cinco em 2003.

Todos esses alunos perfazem um total de cinquenta e oito (0,15%), sendo que quarenta desses não participaram da equalização prévia realizada com os dados antes das correções realizadas, o que reforça a hipótese de que pelo menos esses 40 alunos foram duplicados na produção das etiquetas para a coleta de dados da Ficha A na sétima e na oitava séries e registrados com outro identificador no cadastro.

Somando-se o percentual de alunos com mais de um identificador no banco de dados cadastro, corrigidos pelos programas de correção 1,2 ou 3 (6,7%) com o percentual de alunos com identificadores diferentes por detecção visual (0,15%), em torno de 6,85% dos alunos no banco de dados cadastro apresentaram mais de um identificador.

Contudo, o banco de dados referente aos questionários e testes dos alunos não foi corrigido em 6,85%, porque muitos desses alunos só estavam com identificadores diferentes no banco de dados do cadastro, conforme pode ser visto na TAB. 1.3 através das pequenas diferenças entre o número de alunos antes e após as correções.

TABELA 1.2- Número de alunos duplicados ainda encontrados no banco de dados de alunos após as correções

Período	Número de alunos duplicados			
	Correção 1	Correção 2	Correção 3	Total
Abril/1999	0	1	0	1
Novembro/1999	0	0	0	0
Novembro/2000	2	1	1	4
Novembro/2001	0	0	4	4
Novembro/2002	15	10	9	34
Novembro/2003	8	2	5	15
<b>Total</b>	<b>25</b>	<b>14</b>	<b>19</b>	<b>58</b>

FONTE: Elaboração própria com a utilização dos dados Cedeplar/Inep (2005)

TABELA 1.3- Distribuição dos alunos por período antes e após a correção, utilizando os bancos de dados dos alunos

Período	Questionário sócio-econômico	Matemática	Português
Abril/1999	(a)10482	(a)12641	(a)12624
	(b)10482	(b)12640	(b)12623
Novembro/1999	(a)10482	(a)10686	(a)10583
	(b)10482	(b)10686	(b)10583
Novembro/2000	(a)15476	(a)15494	(a)15463
	(b)15474	(b)15492	(b)15461
Novembro/2001	(a)15268	(a)15280	(a)15221
	(b)15264	(b)15276	(b)15217
Novembro/2002	(a)12159	(a)12191	(a)12146
	(b)12157	(b)12189	(b)12144
Novembro/2003	(a)10054	(a)10101	(a)10085
	(b)10048	(b)10095	(b)10079

FONTE: Elaboração própria com a utilização dos dados Cedeplar/Inep (2005)

NOTAS: (a) antes de correções; (b) após as correções

## **Correção dos números identificadores de escolas**

O banco de dados referente ao aluno contém 3 identificadores de escola, que estavam presentes no cadastro do aluno, nos testes do aluno e nos questionários sócio-econômicos respondido pelos estudantes. Após a correção dos identificadores dos alunos, procede-se a correção dos identificadores de escola. Inicialmente, os arquivos foram divididos por ano e por disciplina. A seguir, da variável “chave” foi extraído o identificador de escola da prova. Para verificar a coincidência entre os identificadores de escola presentes nos bancos de dados por período, fez-se a diferença entre o idescola do cadastro e o idescola da prova. Esses resultados mostram que existe total coincidência entre os identificadores de escola presentes no cadastro e aqueles presentes nos testes de matemática e de português. Entretanto, o mesmo não acontece entre os identificadores de escola do cadastro e dos questionários sócio-econômicos. Como os questionários são aplicados em data próxima à data de aplicação das provas, cujos identificadores de escola são iguais aos do cadastro, é bastante plausível supor que as não coincidências entre os identificadores de prova e questionário são erros de digitação na variável identificador da escola no questionário. Assim sendo, os identificadores de escola dos questionários foram substituídos pelos identificadores presentes nas provas.