

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DE RIBEIRÃO PRETO
DEPARTAMENTO DE ECONOMIA

RODRIGO ARAÚJO FERREIRA

**Desigualdade de desempenho escolar dos alunos do ensino fundamental do
estado de São Paulo: uma análise de decomposição**

Orientador: Prof. Dr. Luis Guilherme Dácar
da Silva Scorzafave

RIBEIRÃO PRETO
2008

Profa. Dra. Suely Vilela
Reitora da Universidade de São Paulo

Prof. Dr. Rudinei Toneto Junior
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão
Preto

Profa. Dra. Maria Christina Siqueira de Souza Campos
Chefe do Departamento de Economia

RODRIGO ARAÚJO FERREIRA

Desigualdade de desempenho escolar dos alunos do ensino fundamental do estado de São Paulo: uma análise de decomposição

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo como requisito para obtenção do título de Mestre em Economia.

Orientador: Prof. Dr. Luis Guilherme Dácar da Silva Scorzafave

RIBEIRÃO PRETO
2008

Ferreira, Rodrigo Araújo

Desigualdade de desempenho escolar dos alunos do ensino fundamental do estado de São Paulo: uma análise de decomposição. Ribeirão Preto, 2008.

78 p. : il. ; 30 cm

Dissertação de Mestrado, apresentada à Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo.

Orientador: Scorzafave, Luis Guilherme Dácar

1. Desempenho Escolar. 2. Escolas Públicas. 3. Economia da Educação. 4. Índice Theil – L.

FOLHA DE APROVAÇÃO

Rodrigo Araújo Ferreira

Desigualdade de desempenho escolar dos alunos do ensino fundamental do estado de São Paulo: uma análise de decomposição

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo como requisito para obtenção do título de Mestre em Economia.

Aprovado em:

Banca Examinadora

Prof. Dr. _____

Instituição _____ Assinatura _____

Prof. Dr. _____

Instituição _____ Assinatura _____

Prof. Dr. _____

Instituição _____ Assinatura _____

RESUMO

Utiliza-se o método de decomposições do índice Theil-L para encontrar determinantes da desigualdade de desempenho escolar das crianças da 4ª série do ensino fundamental do estado de São Paulo. Os dados de proficiência dos alunos em português e matemática são extraídos da Prova Brasil 2005 e combinados com os dados do Censo Escolar 2005. A análise das contribuições brutas e marginais geradas por meio do índice Theil – L mostra que variáveis comumente utilizadas como medidas de background sócio-econômico dos alunos explicam pouco da desigualdade das notas, menos de 2%. Ao mesmo tempo, as contribuições brutas e marginais apontam para a importância significativa das escolas, que explicam 14,58% da desigualdade de desempenho dos alunos em matemática e 13,27% do desempenho dos alunos em português. A única variável que explica significativamente a desigualdade entre as escolas é o município, 14,56% da desigualdade em matemática e 13,98% em português. As demais variáveis quando consideradas isoladamente, respondem por menos de 2,5% da desigualdade. Quando tomadas em grupos, seu poder explicativo aumenta. As variáveis de infraestrutura escolar explicam 18,68% da desigualdade entre as escolas em português e 20% da desigualdade em matemática. As variáveis referentes ao background dos colegas respondem respectivamente por 8,79 e 9,40%. Conclui-se que o impacto das diferentes variáveis no desempenho dos alunos se dá por meio de sua interação mais do que pelo efeito de variáveis específicas. Reitera-se também a importância de se entender o impacto dos insumos escolares para compreender a proficiência dos alunos. O trabalho aponta ainda o efeito que os municípios têm sobre as escolas e oferece base para uma melhor compreensão posterior desse fenômeno. Palavras-chave: Desempenho escolar. Escolas Públicas. Economia da Educação. Índice Theil – L.

ABSTRACT

This research uses the method of Theil-L decomposition to find determinants of the inequality among students' scores of the 4th grade of the fundamental school in the state of São Paulo. The data is provided by a combination of Prova Brasil 2005 standardized tests and the Censo Escolar 2005. The analysis of the gross and marginal contributions measured by the Theil-L index indicates that the variables normally used for students' social and economic background give few explanation about the total inequality measured. The analysis of the decompositions also shows that schools play a main role in the inequalities among students, explaining 14,58% of the students' inequality in math and 13,27% in Portuguese. The only single variable that shows an important impact in the score inequalities among schools is the municipality, 14,56% for math and 13,98% for Portuguese. The other variables have no impact over 2,5% when considered solely. By the other side, when these variables are considered in groups, there is a gain in explanation. The school infra-structure explain 18,68% of the inequality among schools in Portuguese and 20% of the inequality in math. The variables about the peers background are responsible for 8,79 and 9,40% respectively. It is concluded that the impact of different variables in students' learning are more due to their interaction, than by their direct effects. It is reinforced the importance of a better understanding of school variables to understand children apprenticeship as shown by part of the literature. This research points to the importance of municipalities for schools and gives a basis for further discussion on this matter. Key words: Apprenticeship. Public schools. Economics of Education. Theil-L index.

SUMÁRIO

1. Introdução	1
2. Revisão Bibliográfica.....	3
2.1. Escola	4
2.1.1 Rede	10
2.2. Professores.....	11
2.3. Colegas.....	14
2.4. Background Sócio-Econômico.....	16
2.5. Comunidade	18
2.6. Outros Fatores	19
3. Metodologia.....	25
3.1. Base de Dados	25
3.2. Índices de Entropia.....	26
3.3. Índices de Theil	28
3.4. Decomposição da Desigualdade.....	31
3.5. Formação dos Grupos para Decomposição.....	33
4. Análise Descritiva	34
5. Resultados	49
5.1. Contribuições Brutas	49
5.2. Contribuições Marginais.....	56
6. Considerações Finais	70
Referências Bibliográficas	73
APÊNDICE A	76
APÊNDICE B.....	77

1. Introdução

O maior determinante da desigualdade de renda no Brasil é a desigualdade de escolaridade entre os indivíduos, Barros e Mendonça (1995). O acesso à escola têm sido uma das principais preocupações do país e assistimos a partir de meados dos anos 90 uma universalização quase completa do acesso das crianças ao ensino fundamental. O próximo passo consiste em compreender os determinantes de uma educação de qualidade, o que passa pela compreensão do processo de aprendizagem das crianças e dos seus determinantes.

A compreensão dos determinantes de uma educação de qualidade, porém, não é uma tarefa fácil. Existem controvérsias na literatura referente ao impacto das diferentes variáveis relacionadas ao desempenho escolar. O debate em torno da significância de diferentes variáveis foi inaugurado pelo relatório Coleman (Coleman,1966) que causou surpresa a não encontrar um impacto significativo de diversos insumos escolares sobre o desempenho dos alunos.

Basicamente, esse debate têm sido em torno da significância de diferentes variáveis referentes às escolas e aos alunos no aprendizado, como salas mais equipadas, laboratórios, menos alunos por sala, o impacto dos professores, o impacto do background sócio-econômico dos colegas sobre o aluno e o impacto do *background* sócio-econômico do aluno. Essa última variável é a que tem encontrado mais respaldo da literatura, sendo invariavelmente significativa desde Coleman (1960).

A respeito do impacto de diferentes insumos escolares é interessante ver o trabalho de Hanushek (1986) e a crítica de Dewey et alli (2000). Os resultados de Hanushek vão ao encontro do relatório de Coleman (1966) e Dewey et alli (2000) aponta para as possíveis questões metodológicas que poderiam estar levando a resultados não significantes dos insumos escolares.

No Brasil, os pesquisadores têm cada vez mais afirmado o impacto significativo que as escolas têm sobre o processo de aprendizagem, como se pode ver em Rios-Neto, César e Riani (2003), Albernaz, Ferreira e Franco (2002) e Felício (2004). Entretanto, perguntas como quais os determinantes do impacto das escolas, qual o impacto do município e da comunidade ou ainda qual a real efetividade de diversos programas governamentais, ainda precisam de resposta.

A natureza das diferentes variáveis relacionadas às escolas ainda não é claramente conhecida, pois mesmo quando os autores concordam sobre o impacto das escolas, eles ainda divergem sobre quais os insumos responsáveis por esse impacto. Para enfrentar este problema

os autores têm feito uso de diversos diferentes métodos para tentar captar o efeito das diferentes variáveis de interesse.

O exemplo colocado acima das questões relativas ao impacto de diferentes insumos escolares é apenas um exemplo do debate que existe em torno da importância de diversos fatores no desempenho dos alunos, como será tratado mais a frente na revisão bibliográfica.

A complexidade da questão educacional e a diversidade dos resultados encontrados permitem uma série de diferentes abordagens para que possa ser compreendida. Assim, este trabalho se insere na linha das perguntas que foram levantadas desde o relatório Coleman (Coleman, 1960) sobre quais os determinantes do desempenho dos alunos. Procura realizar uma análise exploratória do impacto de diferentes variáveis que têm sido analisadas na literatura.

Em 2007, foram disponibilizados os dados da Prova Brasil 2005. Esses microdados permitem um nível de desagregação que não era possível com as bases de dados anteriores, enriquecendo as possibilidades de análise. Além disso, por se tratar de uma base de dados com muito mais observações do que o SAEB, comumente utilizado nesses casos, permite trabalhar com um conjunto maior de variáveis na busca da causa de determinantes de desempenho escolar.

Neste trabalho, será utilizado o método de decomposição do índice de Theil sobre os dados da Prova Brasil 2005 combinados com os dados do Censo Escolar 2005, o que permite que seja analisado o impacto de diversas variáveis no nível do aluno e das escolas. O método permite que avaliemos as variáveis uma a uma e ainda que analisemos seu impacto conjuntamente com outras variáveis.

Para realizar essa análise, primeiro será apresentada uma revisão da literatura em que este trabalho se insere, depois será apresentado com detalhes o método que será utilizado. Por conta da novidade dos dados da Prova Brasil, será dedicado um espaço importante à apresentação desses dados e por fim serão apresentados os resultados obtidos por meio do índice Theil-L.

A amostra desse trabalho consistirá nos alunos da 4ª série do ensino fundamental da rede pública do estado de São Paulo que realizaram a Prova Brasil em 2005.

Espera-se que este trabalho possa contribuir na análise de diferentes fatores associados ao desempenho escolar e oferecer base para que trabalhos posteriores possam compreender ainda melhor o impacto de diferentes variáveis sobre o aprendizado. O trabalho se propõe um estudo exploratório sobre possíveis fatores que afetam o desempenho dos alunos.

2. Revisão Bibliográfica

A literatura da chamada Economia da Educação no Brasil passou a ganhar maior atenção dos economistas com a constatação de que o problema da desigualdade de renda está intrinsecamente relacionado com o problema da desigualdade na educação. O trabalho de Barros e Mendonça (1995) é uma das grandes referências a esse respeito. Neste, encontram-se evidências que mostram que é a desigualdade educacional a principal causa da desigualdade de renda no Brasil.

Seguindo a mesma linha de pesquisa, Menezes-Filho (2001) analisa a desigualdade de educação no Brasil e encontra gargalos na passagem dos alunos do ensino médio para o ensino superior no início dos anos 90, o que pode ser a causa da grande diferença de salários entre os extratos mais escolarizados da população e o extrato imediatamente inferior.

Uma segunda vertente através da qual a discussão sobre educação ganhou importância foi por meio das teorias de crescimento econômico que consideram a importância do capital humano. Segundo essas teorias, a habilidade dos indivíduos em manusear os insumos da economia determinaria a diferença de crescimento entre os países. A educação aparece então como um dos componentes, ou uma *proxy*, do capital humano, fundamental para o crescimento econômico¹.

Já na década de 60, podem-se observar os primeiros trabalhos que associam educação e crescimento econômico como citado em Aghion e Howitt (1998). Dentre esses trabalhos, o de Becker (1964) ganhou bastante evidência. Este autor foi o primeiro a apontar a relação entre capital humano e crescimento econômico, dando origem a uma linha de pesquisa que se desenvolveu até o trabalho de Lucas (1988), em que se apresenta o acúmulo de capital humano como um caminho de crescimento.

Nos anos 90, os estudos de Economia da Educação, já apresentavam autonomia com relação aos estudos de teorias do crescimento. Nessa época, já se pode encontrar nos países desenvolvidos pesquisas que se concentravam estritamente no desempenho escolar dos alunos, buscando entender quais incentivos poderiam gerar mais eficiência entre e dentro das escolas e quais seriam as práticas educacionais com melhor custo-benefício (Hoenack, 1996).

Os trabalhos encontrados na literatura recente são majoritariamente testes sobre a relevância de diferentes variáveis na determinação do desempenho escolar. Na maioria dos casos, estes testes seguem o modelo apresentado por Coleman (1966), utilizando uma função de produção escolar em que diversas variáveis referentes a características dos alunos e da

¹ Referente à discussão sobre Teorias do Crescimento, consultar Raj (1998).

escola são colocados como insumos. Algumas variáveis se destacam pelo número de trabalhos dedicados à sua análise como as referentes ao impacto da estrutura escolar, professores, desempenho dos colegas e *background* socioeconômico.

Nesta seção serão apresentados os resultados da literatura em relação à significância dos insumos em uma função de produção escolar. Esta abordagem nos parece melhor por haver quase um consenso na literatura sobre quais insumos devem ser analisados e por permitir a comparação de resultados entre estudos mais gerais, que analisam a qualidade de escolas e programas governamentais e estudos mais específicos, que atentam para o impacto de apenas um insumo. Assim, apresentaremos abaixo os resultados referentes à estrutura escolar, *background* socioeconômico, professores, colegas e outros resultados interessantes da literatura.

2.1 Escola

O relatório Coleman (Coleman, 1966), que avaliava desempenho estudantil nas escolas americanas desenvolveu o modelo que é o mais utilizado para se avaliar desempenho educacional. O modelo tem o seguinte formato: $y = f(c, m, g, p, s)$, em que y é o desempenho médio do aluno, c são as características pessoais tais como etnia e idade, m são as características de sua família, incluindo seu *background* socioeconômico, g são as características dos colegas, p são as características referentes aos professores e s são as características referentes à escola. Este modelo de “função produção” é até hoje o mais utilizado entre os pesquisadores.

O trabalho de Coleman chegou a vários resultados importantes, dentre os quais se destaca a importância observada do *background* socioeconômico dos alunos como fator mais importante para o seu desempenho. Entretanto, o estudo chegou também a uma conclusão intrigante: a ausência de um impacto significativo da infra-estrutura escolar. Segundo o relatório, apenas 2% da variância de desempenho entre os alunos pode ser atribuída a esta variável. Diversos estudos foram realizados posteriormente e corroboraram os resultados de Coleman. Hanushek (1986) concluiu, depois da análise de 187 estudos sobre o tema, que a única variável relevante da escola para o desempenho dos alunos é a experiência dos professores, descartando qualquer impacto relevante das salas, prédio, equipamentos e outros itens relacionados à infra-estrutura escolar.

Com relação à melhoria da qualidade do ensino por meio da infra-estrutura escolar, Hanushek (2003) aponta uma série de evidências na literatura de que não há resultados que indiquem que um aumento nos gastos por aluno leva a uma melhora da sua performance. Ele aponta ainda que, de 1960 a 2000, o número de alunos por professor caiu 40% nos EUA, o que não resultou num melhor de desempenho dos alunos. O autor observa que isso poderia ser explicado por um aumento das atribuições das escolas ou uma alteração do perfil dos alunos devido ao aumento dos gastos com educação especial, aumento da porcentagem de crianças morando em áreas pobres e aumento da porcentagem de crianças que não falam inglês em casa. Entretanto, afirma que o efeito desses fatos é contrabalanceado por uma queda no número de crianças por família, aumento da porcentagem de adultos com educação colegial e superior e aumento do percentual de crianças que recebem educação infantil, não ficando claro que há uma tendência que justifique o aumento de gastos por aluno.

Hanushek (2003) aponta os resultados encontrados em Hanushek e Kimko (2000), que indicam uma relação negativa e significativa entre o aumento de gastos escolares por aluno e o desempenho dos alunos de diversos países. Os autores utilizaram dados do TIMSS² e controlaram por diversas características. Ele cita ainda Woessman (2000, 2001), que utilizando dados do TIMSS de 1995 para os países desenvolvidos, conclui que as medidas tradicionais de infra-estrutura explicam consistentemente muito pouco do desempenho dos alunos. Há ainda menção feita ao trabalho de Gundlach et alli (2001), que analisando os gastos com educação e o desempenho dos alunos em diferentes países entre 1970 e 1995, encontra uma queda de produtividade das escolas, relacionando negativamente aumento de gastos e desempenho escolar.

Segundo Hanushek (2003), os estudos que corroboram o impacto da infra-estrutura sobre o desempenho dos alunos americanos cometem erros de agregação, utilizando dados nacionais e desconsiderando as grandes diferenças das políticas educacionais dos estados. Esses estudos ainda cometem segundo esse autor erros de interpretação dos efeitos dos insumos, que são cumulativos ao longo do tempo, o que faz com que a correlação com o desempenho não possa ser mensurada apenas pelo fluxo dos insumos no período de análise. Ainda assim, apenas 14% dos estudos analisados pelo autor encontraram uma relação significativa e positiva entre aumento dos gastos por aluno e desempenho escolar.

² Third International Math and Science Study (TIMSS) – exame internacional de proficiência escolar em matemática e ciências conduzidos em 38 países pela [International Association for the Evaluation of Educational Achievement](#) (IEA) desde 1995 a cada 4 anos.

Esse argumento é importante para Hanushek, pois explica, segundo o autor, a invalidade de diversos resultados encontrados na literatura americana referente ao programa STAR³, que verificam um melhor desempenho dos alunos de salas menores e uma não significância da presença de assistentes de professores nas salas de aula, além de problemas na seleção dos alunos do programa, que não teria sido aleatória. Contestando o resultado de Krueger (1999) que aponta um melhor desempenho dos alunos de salas menores, Hanushek ressalta que o valor adicionado da redução de salas é nulo exceto para o primeiro ano de estudo, sendo esperado que os alunos de salas menores continuassem a ter acréscimo de desempenho ao longo dos anos. Assim, parece que a existência de salas menores traz resultados positivos no início do processo de aprendizagem, mas seus ganhos se restringem a este período da vida escolar.

Examinando uma amostra de 17 países participantes do TIMSS, Hanushek e Luque (2003) (citado em Hanushek, 2003) não encontraram uma relação significativa entre a infraestrutura escolar e o desempenho dos alunos. A princípio, notaram uma relação significativa e negativa entre o tamanho da sala e as notas dos alunos mais jovens. A amostra contava com países em desenvolvimento e não foi encontrada uma relação diferente para estes. Isto parece contrariar a hipótese de que a não significância da infra-estrutura escolar em estudos de países desenvolvidos se deve ao fato de suas escolas já estarem atuando além do seu nível de produtividade marginal positiva, supondo produtividade marginal decrescente.

Em contraste com a literatura internacional, Albernaz, Ferreira e Franco (2002) encontraram um impacto significativo da infra-estrutura escolar no desempenho dos alunos em sua análise da qualidade do ensino fundamental brasileiro a partir de dados do SAEB de 1999. O trabalho de Albernaz, Ferreira e Franco (2002) encontrou que 28% da diferença de desempenho entre os alunos pode ser explicada pelas diferenças entre as escolas. Encontrou que aspectos como a presença de ruído e o arejamento da sala impactam significativamente o desempenho dos alunos. Além disso, encontrou uma relação significativa entre a capacidade de financiamento e a capacidade de oferecer professores para todas as matérias e o desempenho dos alunos.

Essa diferença nos resultados pode ser atribuída, segundo os autores, ao viés de agregação presente nos trabalhos anteriores ou a peculiaridades da educação no Brasil que ainda não foram identificadas. O viés de agregação consiste em supor que as características da

³ [Tennessee's Student Teacher Achievement Ratio \(STAR\)](#) – experimento de 4 anos iniciado em 1985 com alunos dos Tennesse, EUA, sobre o impacto da redução do número de alunos na sala de aula sobre os alunos. O experimento durou quatro anos e acompanhou um coorte de alunos do ensino infantil até o terceiro ano do ensino fundamental.

amostra se distribuem aleatoriamente quando na verdade sua distribuição depende do subgrupo a que pertencem. Nos estudos de Economia de Educação, significa supor que os alunos estão aleatoriamente distribuídos pelas escolas, quando, na verdade, alunos com determinadas características se alocam em certas escolas. Dessa forma, a escola tem um impacto sobre a distribuição dos alunos e explica parte da variável dependente, nesse caso o desempenho dos alunos no SAEB.

Para fazer frente a esse problema, o trabalho de Albernaz, Ferreira e Franco (2002) fez uso de um modelo hierárquico linear. O modelo hierárquico de dois níveis utilizado pelos autores consiste na estimação de duas regressões. Na primeira, o desempenho do aluno é função do desempenho médio dos alunos da escola mais um termo aleatório que depende do aluno e da escola. No segundo nível, a média de notas da escola é função da média das notas de todas as escolas e um termo aleatório que depende da escola. Estimando os dois modelos, pode-se saber quanto da variância do termo aleatório do primeiro modelo é devido às escolas.

A significância da infra-estrutura para o desempenho estudantil no Brasil também encontra respaldo em Rios-Neto, César e Riani (2003). Utilizando dados do Censo Escolar e do Censo 2000, estes autores também utilizaram um modelo hierárquico linear e encontraram um impacto positivo da presença de laboratórios de informática e ciências sobre a taxa de progressão⁴ dos alunos. Foi percebido que o impacto é mais forte entre os alunos da 5ª série, quando comparados aos da 1ª.

Diaz (2007) encontrou um resultado parecido com o de Albernaz, Ferreira e Franco (2002) referente ao percentual da variância do desempenho que se deve à variação entre instituições de ensino. Utilizando também um modelo hierárquico linear, a autora encontrou que 29% da variância das notas se deve à variação entre instituições⁵. O trabalho de Diaz utiliza uma amostra de instituições de ensino superior no Brasil obtidos no Provão⁶. Inserindo no modelo várias variáveis de controle nos dois níveis utilizados, instituição de ensino e aluno, essa correlação cai para 14,25%.

A autora encontrou ainda resultados positivos para a participação dos alunos em atividades de pesquisa e extensão e para o engajamento e dedicação dos professores, diminuindo o impacto negativo da baixa renda familiar. Uma *dummy* utilizada pela autora para indicar técnicas participativas de ensino apresentou um impacto negativo, indicando,

⁴ Número de alunos da coorte que chegaram à série escolar atual sobre o número dos que haviam chegado à série anterior.

⁶ Provão – Exame Nacional de Cursos, exame de avaliação do Ensino Superior realizado pelo INEP.

talvez, um mal uso deste recurso pedagógico. Foi encontrado ainda um impacto positivo para maiores percentuais de mestres e doutores no corpo docente e do trabalho de até 8h semanais em sala de aula.

Também analisando microdados brasileiros, Felício (2004) encontra um impacto de até 34,44% da escola no desempenho dos alunos, utilizando a desigualdade entre escolas a partir da decomposição de um índice Theil-L e dados do SAEB de 2001. Num segundo momento, fez também estimações de mínimos quadrados, tratando as escolas como efeito fixo através de *dummies*, simulando qual seria o efeito sobre o desempenho dos alunos se fossem transferidos para uma escola representativa gerada a partir da média dos insumos das cinco melhores escolas. O resultado encontrado foi que os alunos da 4ª série teriam um desempenho equivalente aos da 7ª do Ensino Fundamental do Estado de São Paulo.

Com relação à análise do impacto de diferentes insumos escolares, Barros et alli (2001) analisaram o seu impacto juntamente com variáveis de custo de oportunidade de tempo de estudo, renda familiar, recursos disponíveis na comunidade, além de controles de background socioeconômico. Os autores usaram dados da PNAD⁷ de 1996 e da PPV⁸ de 1996/97. A construção de um índice de qualidade da educação a partir de dados das escolas privadas apresentou como item mais significativo a presença de computadores na escola⁹. O uso de médias de diversos insumos escolares, bem como do índice gerado, apresentou resultado significativo em boa parte das regressões, sendo que uma escola com todos os insumos poderia aumentar em até 1,1 ano a escolaridade média dos alunos. Foi verificado ainda que a infra-estrutura escolar tem um impacto maior no desempenho dos alunos do ensino fundamental em comparação ao ensino médio na região Sudeste e que esta relação se inverte entre as escolas do Nordeste.

A respeito da não significância de insumos escolares em muitos trabalhos e em particular no resumo de trabalhos realizado em Hanushek (1986), Dewey et alli (2000) apontam que a função de produção escolar comumente utilizada gera problemas de especificação ao confundir oferta e demanda, uma vez que insumos escolares estão do lado da oferta e renda familiar do lado da demanda. A inclusão de variáveis de renda juntamente com variáveis de insumos escolares geraria, portanto, um problema de especificação.

⁷ PNAD – Pesquisa Nacional por Amostras de Domicílios realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) com abrangência nacional, que obtém informações dos domicílios referentes a características gerais dos indivíduos, migração, educação, trabalho, famílias, domicílios e rendimento.

⁸ PPV – Pesquisa de Padrão de Vida, pesquisa amostral realizada nas áreas urbanas da região sudeste e nordeste equivalentes à PNAD, com detalhamento maior referente às condições de vida e educação.

⁹ As escolas de mensalidade mais caras foram assumidas como sendo melhores e a partir disso se construiu uma regressão tendo como variável dependente a mensalidade e como explicativas, diversos insumos escolares.

Os autores analisaram novamente a base de artigos em Hanushek (1986), incluindo alguns novos estudos e retirando os trabalhos que continham os “erros” de incluir variáveis de renda e de insumos escolares na mesma equação, apresentar resultados de variáveis cuja introdução no modelo não foi suficientemente demonstrada e ainda trabalhos que não apresentavam as estatísticas de teste dos coeficientes. Nessa nova base, encontraram que 81% dos coeficientes referentes a insumos escolares são positivos e 37% são positivos e significantes, contra 20,3% em Hanushek (1986). Quase metade dos coeficientes para experiência do professor, outras características do professor e dispêndio por aluno são significantes e positivos. Ainda um sexto dos coeficientes referentes educação dos professores, professores por aluno e tamanho da escola são significantes e positivos. Foi ainda analisada a significância dos coeficientes dos insumos escolares dentro universo das regressões, utilizando o teste qui quadrado inverso para a meta-análise e os coeficientes ainda se mostraram significativos.

Os autores realizam ainda duas estimações próprias para analisar o impacto de insumos escolares sobre o aprendizado e o impacto da introdução de medidas de renda familiar sobre os coeficientes das regressões. A partir de dados do TALENT¹⁰ com dados socioeconômicos, de proficiência verbal e em matemática, encontraram coeficientes significativos e positivos para jornadas mais longas de aula e duração do ano letivo, também encontraram resultados “esperados” para tempo de estudo em casa e mudança recente de escola. Os resultados para o impacto dos professores e dos colegas também foram interessantes e estão apresentados mais abaixo. A introdução de medidas de renda, log da renda familiar e log da renda média na escola, promoveram a queda no valor dos coeficientes dos insumos escolares e em alguns casos os tornou não significativos.

Resultados parecidos foram encontrados com uma amostra menor com dados da SAT¹¹ e dados sobre gastos com educação de 33 estados americanos. Nas regressões realizadas com esta amostra, as variáveis referentes aos professores, ao gasto por aluno e ao tamanho da escola se mostraram significantes e positivas, a dummy para rede pública apresentou-se significativa e negativa. A introdução das variáveis de renda alterou os coeficientes das variáveis dos insumos escolares, tornando-as menos significativas ou até mesmo invertendo o sinal dos coeficientes. Exceção para a experiência dos professores que se

¹⁰ Talent – Programa do governo federal americano que procura aumentar o número de alunos em desvantagem sócio-econômica que terminam o ensino médio e entram no ensino superior. O programa oferece aconselhamento acadêmico, financeiro e profissional para esses alunos.

¹¹ SAT – Exame padronizado adotado pela grande maioria das universidades americanas como um dos requisitos de admissão.

tornou mais significativa com a introdução da variável renda. Por fim, os autores realizaram o teste de Hausman que indicou que os insumos escolares são endógenos ao *background* socioeconômico.

Com relação ao impacto das escolas sobre os alunos, Figlio e Kenny (2006) observaram um impacto maior e significativo das escolas que servem alunos de baixa renda. Outro resultado importante encontrado pelos autores foi a não-significância de incentivos não-financeiros como avaliações anuais e maiores taxas de demissão para explicar desempenho estudantil. Entretanto, os autores apontam que não podem dizer se o aumento de desempenho dos alunos se deve ao maior esforço dos professores em consequência dos incentivos financeiros ou à presença de práticas escolares de difícil observação correlacionadas com os incentivos.

Outro aspecto que pode ser analisado referente ao impacto das escolas sobre os alunos é o surgimento de novas práticas de gestão. A década de 80 acompanhou a introdução de inovações no sistema de educação brasileiro seguidas ao processo de democratização iniciada em 1982 com as eleições diretas para governadores. Barros e Mendonça (1998) analisam as três principais mudanças que foram introduzidas em diversos estados brasileiros: a introdução da eleição para diretor, a criação de colegiados de professores, funcionários e alunos de caráter consultivo e deliberativo e a transferência de recursos para administração direta das escolas.

Os autores usaram dados do Censo Escolar de 1981 a 1993, do SAEB de 1990 e 1993 e da PNAD de 1981 a 1993 para estimar o impacto dessas três inovações sobre diferentes indicadores de desempenho escolar através de uma *pooling cross-section*, em que as observações são os estados da federação. Do seu modelo mais completo, com a presença de efeitos fixos e redução do período da amostra, concluíram que a transferência de recursos e formação de colegiados impactaram negativamente, ainda que modestamente, a proporção de crianças fora da escola, a defasagem série-idade e a proporção de crianças com algum atraso escolar, a transferência de recursos diminuiu ainda, também modestamente, a taxa de reprovação. Já a eleição de diretores apresenta um efeito perverso, aumentando a proporção de crianças fora da escola.

2.1.1 Rede

A rede a que as escolas pertencem, (privada ou pública) é um aspecto relevante da identificação das características escolares. Na sua análise sobre a qualidade do ensino fundamental, Albernaz, Ferreira e Franco (2002) encontraram um valor significativo para a

variável referente à rede à que a escola pertence. Curiosamente, se retirada esta variável, a variável referente à impossibilidade das escolas de oferecer professores para todas as matérias passava a ser significativa. Isto parece indicar que, para o caso brasileiro, o fato da escola ser privada está correlacionado com algum tipo de ganho de eficiência.

Também no trabalho de Rios-Neto, Riani e César (2003) observa-se a importância da rede a que a escola pertence. A porcentagem de escolas municipais entre as escolas públicas, afeta negativamente a taxa de progressão escolar dos alunos. O efeito é ainda maior para alunos da 1ª série comparados aos da 5ª, sendo que o número de escolas municipais de 1ª série é maior que o de 5ª.

A importância da rede de ensino também aparece em Figlio e Kenny (2006) no seu trabalho sobre incentivos aos professores. Os coeficientes para as variáveis de incentivos financeiros para os professores são quase um terço menor entre as escolas públicas em relação às privadas.

2.2 Professores

Outra variável que tem sido examinada e também têm apresentado resultados intrigantes é o impacto dos professores no desempenho dos alunos. Particularmente não tem sido encontrada uma correlação robusta entre os salários dos professores e o desempenho dos alunos. Coleman (1966) não encontrou esta correlação. Entretanto, verificou um impacto da experiência dos professores sobre os alunos.

Já o trabalho de Hanushek (2003), fazendo a compilação de dados de diversos estudos publicados até a data, aponta que apenas 29% dos estudos que analisavam o impacto da experiência do professor sobre o desempenho dos alunos encontraram uma relação significativa. Assim, 71% dos estudos não encontraram nenhuma relação ou uma relação negativa entre as variáveis. E ainda, apenas 9% dos estudos encontraram uma relação significativa e positiva entre o desempenho dos alunos e a escolaridade dos professores. Também há a possibilidade de que a escolha das salas pelos professores seja endógena, isto é, que melhores professores escolham as melhores salas. Se levarmos em conta apenas trabalhos que utilizam valor adicionado, isto é, o ganho de aprendizagem dos alunos ao longo das séries, também não há um impacto significativo da escolaridade dos professores. Entretanto, 59% desses trabalhos encontram uma relação positiva e significativa entre desempenho dos alunos e experiência dos professores.

Há resultados na literatura, entretanto, que indicam um impacto expressivo dos professores sobre os alunos. Hanushek (1992) (in Hanushek, 2003) afirma que professores

que estão no 95º percentil tem um impacto relativo de 1,5 anos de estudo por ano letivo sobre alunos de baixa renda de centros urbanos, enquanto que professores do 5º percentil tem o impacto de 0,5 ano por ano letivo sobre os alunos. Hanushek apresenta ainda os resultados de Rivkin et alli (2001) em que a troca de um professor médio para um do 85º percentil gera um aumento no desempenho dos alunos em 4 percentis por ano. Aponta ainda que se os alunos de baixa renda tiverem 6 anos seguidos de instrução com professores do 85º percentil, a diferença de desempenho médio com os demais alunos seria superada.

O trabalho de Albernaz, Ferreira e Franco (2002) encontrou um impacto positivo dos professores sobre as notas. Mesmo depois de controlar por aspectos socioeconômicos, rede de ensino, gênero e raça e observando as diferenças entre escolas, encontraram valores significativos para a escolaridade dos professores. Embora a variável salários dos professores seja não-significante no modelo com todas as variáveis, se colocada como substituta da variável escolaridade dos professores, torna-se significativa.

Outro resultado interessante a que chegaram Albernaz, Ferreira e Franco (2002), foi o diferente impacto dos professores para grupos de diferentes *backgrounds* socioeconômicos. O aumento da escolaridade dos professores aumenta mais o desempenho dos alunos de maiores níveis de renda, relativamente aos mais pobres, aumentando a média das notas e a sua desigualdade.

Rios-Neto, Riani e César (2002), utilizando dados de PNADs de 1981 a 1997 e a taxa de progressão de alunos das 1^{as} e 5^{as} séries, examinaram a possibilidade de uma substituição entre a escolaridade da mãe e a do professor no desempenho dos alunos. Mais uma vez, os autores utilizaram um modelo hierárquico linear, o que lhes permitiu analisar o impacto do coeficiente da educação dos professores sobre a educação da mãe. Segundo os coeficientes, observa-se uma substituição entre a educação da mãe e a dos professores no desempenho dos alunos da 1ª série. Já entre os alunos das 5^{as} séries, nota-se um coeficiente maior para o nível de escolaridade do professor sem uma substituição com o coeficiente de escolaridade da mãe, que passa a ser menor. Nota-se ainda que a substituição entre a escolaridade da mãe e do professor é maior entre as mães menos escolarizadas.

Complementando seu trabalho de 2002, Rios-Neto, Riani e César (2003) repetiram a análise utilizando outra base de dados e alterando a variável escolaridade do professor para percentual de professores com nível superior. Neste novo trabalho, encontrou-se novamente um impacto positivo da variável relacionada com os professores no desempenho dos alunos, entretanto, não se observou a substituição do trabalho de 2002. Percebeu-se um impacto positivo no número de horas-aula e uma substituição semelhante ao que se notou no trabalho

de 2002, inclusive em relação ao maior impacto nas famílias mais pobres, porém, entre escolaridade materna e horas-aula. Segundo simulação dos autores, um aumento de 2,5 horas teria o impacto equivalente ao aumento da escolaridade da mãe em 6 anos.

Felício (2004), que conforme citado acima encontrou uma relação significativa entre o desempenho dos alunos e a escola, procurou também analisar de forma mais específica quais os atributos escolares mais significativos neste impacto. Após regredir o desempenho dos alunos contra suas variáveis de *background*, tratando as escolas como efeitos fixos aos quais atribuiu *dummies*, fez uma nova regressão dos coeficientes destas *dummies* contra diversos atributos escolares. O resultado foi inconclusivo para variáveis como experiência e escolaridade do professor.

Analisando dados de uma universidade canadense de 1996 a 2005, Hoffmann e Oreopoulos (2006) encontraram uma correlação positiva entre as notas de alunos universitários e as avaliações subjetivas que estes fazem de seus professores. Porém, não encontram uma relação quando observam análises objetivas, como salários e ranqueamento. Observando o efeito dos professores ao longo dos anos, notaram um impacto maior no primeiro ano, quando os alunos ainda possuem poucas informações sobre os cursos e são, portanto, mais dependentes dos professores.

Alguns autores têm apontado ou ao menos sugerido a possibilidade de explicações institucionais para a diferença de aprendizado entre alunos e entre escolas. Ao apresentar seus resultados sobre o impacto de incentivos financeiros aos professores sobre o desempenho de alunos, Figlio e Kenny (2006) mostram que esta relação só é significativa para incentivos de alta seletividade.

Segundo esse trabalho, os alunos só se beneficiam de incentivos em que os professores de uma escola apresentem uma alta variância dos salários ou em que um grupo pequeno de professores recebe aumentos e bônus. Incentivos que são amplamente distribuídos entre os professores não são significantes para o desempenho dos alunos.

O trabalho de Figlio e Kenny (2006) cita o resultado de um programa de incentivo às escolas no Kenya¹² em que os bônus eram igualmente distribuídos entre os professores. Os pesquisadores verificaram ganhos de curto prazo através das notas dos alunos, mas nenhum ganho de longo prazo, o que sugere que os professores manipularam os resultados. Com relação ao impacto dos professores sobre o desempenho dos alunos, pode-se ainda encontrar estudos sobre a relação entre incentivos aos professores e desempenho estudantil. Figlio e

¹² Glewwe et alli (2003)

Kenny (2006) mesclaram diferentes bases de dados referentes a escolas americanas na década de 90 e testaram a significância do coeficiente associado a uma variável referente à presença de programas de incentivo aos professores em uma função de produção de educação.

Os autores fizeram uma distinção entre programas que beneficiam um número grande de professores na escola, pouco seletivos, e aqueles que beneficiam poucos professores, seletivos. Sem controlar por seletividade, o coeficiente da variável de incentivo aos professores mostrou-se não-significativo, mas ao tratar como incentivo somente os programas mais seletivos, foi encontrada uma relação positiva e significativa.

Os resultados encontrados por Figlio e Kenny (2006) são bastante robustos, uma vez que eles controlaram também por rede, sindicalização dos professores e participação em programas de reforma educacional, que em conjunto servem como *proxy* para a presença de métodos inovadores de lecionar, que poderia estar correlacionado com os incentivos financeiros para os professores.

Como citado anteriormente, Dewey et alli (2000) encontraram resultados interessantes para o impacto dos professores. Analisando dados de proficiência verbal e em matemática a partir dos dados do Projeto TALENT, encontraram que o desempenho dos alunos de matemática está associado com a escolaridade do professor, enquanto que para o desempenho verbal, a experiência do professor é a variável relevante. Outras regressões realizadas com dados do SAT, apresentaram coeficientes positivos e significativos para salários, escolaridade e experiência dos professores.

2.3 Colegas

O impacto dos colegas de sala sobre o desempenho de cada aluno é tratado na literatura de diversas maneiras. Pode-se observar o efeito das notas dos colegas sobre o aluno, o efeito do *background* socioeconômico médio dos colegas ou ainda o tamanho da sala de aula. Com relação ao tamanho da sala de aula, pode-se pensar tanto no efeito da interação de um maior número de alunos sobre seu desempenho como na relação professor por aluno. Não parece ainda claro na literatura o mecanismo através do qual as variáveis dos colegas afetam cada aluno e nem como cada aluno poderia se apropriar de uma porção maior ou menor do serviço oferecido pelo professor durante a aula devido ao tamanho da sala. Segue abaixo a discussão sobre o tamanho da sala como parte da discussão sobre o efeito dos colegas no desempenho dos alunos.

Hanushek et alii (2002) apontam que a literatura sobre o efeito dos colegas tem apresentado resultados diversos devido à presença de dois problemas aos quais muitas vezes

não se deu tratamento adequado: a simultaneidade dos efeitos causados entre alunos e colegas e a correlação das variáveis dos colegas com variáveis omitidas dos alunos. Para enfrentar esses problemas, os autores utilizaram um painel que tinha entre as variáveis explicativas a média defasada em dois anos das notas dos colegas. Com esse modelo, os autores concluíram que o efeito das notas colegas sobre os alunos é significativo. Entretanto, não acharam um efeito significativo para a variância das notas dos colegas.

Observa-se um impacto positivo do nível socioeconômico médio dos alunos da escola sobre o desempenho dos alunos em Albernaz, Ferreira e Franco (2002). E ainda que quanto maior o nível socioeconômico médio dos alunos, menor o impacto do nível socioeconômico de cada aluno sobre seu desempenho.

Em Rios-Neto, Riani e César (2003), mostra-se que o tamanho da sala influencia significativamente o desempenho dos alunos de 1^a e 5^a série analisados, sendo o maior o impacto na 5^a série. Nas simulações realizadas pelos autores, vê-se que a redução do tamanho da sala de 45 para 25 alunos, teria o mesmo impacto que o aumento da educação média das mães de 3 anos sobre a progressão escolar.

Um ponto importante para as políticas públicas em relação ao efeito dos colegas sobre os alunos é como distribuir os alunos entre as salas de aula, se é preferível ter alunos de diferentes características dentro da mesma sala ou separá-los de acordo com suas características. A esse respeito Glewwe (1997) aponta que os modelos que levam em conta apenas o impacto do desempenho médio dos colegas não são suficientes para compreender o fenômeno. O autor demonstra que se considerarmos diferentes funções de desempenho escolar, teremos sugestões muito distintas de política pública.

Se o desempenho dos alunos for uma função exponencial de uma determinada característica, teremos um desempenho médio superior ao desempenho do aluno médio, o que quer dizer que, neste caso, é positivo misturar os alunos com diferentes valores desta característica em uma mesma sala. Se, ao contrário, o desempenho é o logaritmo dessa característica, o desempenho médio da sala ficará abaixo do desempenho do aluno médio, sendo melhor separar os alunos de características distintas.

Dills (2005) encontra resultados que se opõe aos de Hanushek et alli (2002) referente à introdução de *magnet schools*¹³ no subúrbio de Washington. O autor aponta um impacto significativo do desempenho dos colegas com notas mais altas sobre os alunos, em particular

¹³ Escolas americanas que selecionam os melhores alunos.

sobre os de pior desempenho. O autor analisa qual o impacto da saída dos melhores alunos sobre o desempenho dos colegas que permanecem na escola.

O trabalho de Dills utiliza dados de 1987 a 1992 referentes ao percentual de alunos da escola em cada quartil em exames nacionais padronizados do Ensino Médio. A partir disso gera duas equações para o primeiro e o último quartil, tendo como variável dependente o percentual de alunos da escola naquele quartil e como variáveis explicativas o percentual de alunos que deixaram a escola para frequentar as *magnet schools*.

O coeficiente referente ao percentual de alunos que deixam a escola foi significativo e positivo para a regressão dos alunos no primeiro quartil de notas, indicando que a saída de 1% dos alunos da escola aumenta em cerca de 9% o percentual de alunos no primeiro quartil de notas. O coeficiente para a equação dos alunos do último quartil mostrou-se negativo, porém insignificante. Os dois coeficientes encontrados foram significativamente distintos e os resultados foram também corroborados por análises de diferença em diferenças comparando os coortes entre a 8ª e 11ª série.

Dewey et alli (2000) também abrange em seu estudo a questão da distribuição dos alunos nas salas e apontam que a alocação dos alunos com habilidades homogêneas nas mesmas salas de aula é irrelevante para a proficiência verbal, entretanto, tem um impacto positivo sobre o aprendizado de matemática.

2.4. Background Socioeconômico

Desde o relatório de Coleman (1966), as únicas características que se têm mostrado incontestes no seu impacto sobre o desempenho dos alunos, são as de ordem socioeconômica. O nível de renda e a educação dos pais têm respondido pela maior parte da explicação dos resultados de exames escolares. Alguns trabalhos têm apontado de forma especial o impacto da educação da mãe sobre os filhos.

Albernaz, Ferreira e Franco (2002) apontam que 80% da diferença de desempenho entre os alunos se deve ao seu *background* socioeconômico. Além disso, também foi constatado um considerável impacto do *background* socioeconômico médio das escolas nas diferenças entre as escolas, ou seja, quanto mais “ricas” as escolas, melhores suas notas médias.

Um outro dado interessante apontado por Albernaz, Ferreira e Franco (2002) foi o de que quanto maior o nível socioeconômico médio das escolas, menor o impacto do *background* socioeconômico dos alunos sobre o seu desempenho, como se o fato de uma escola ter alunos

mais ricos compensasse parte da desvantagem que um aluno tem de ser pobre. A respeito disso, Dewey et alli (2000) encontraram um impacto negativo da desigualdade de renda entre os alunos sobre o seu desempenho.

Ao observar as diferenças de desempenho determinadas por gênero, Albernaz, Ferreira e Franco (2002) mostra que os alunos do sexo masculino tiveram, em média, uma nota sete pontos menor no SAEB de 1999, após realizados os demais controles. Além disso, os meninos apresentaram desempenho melhor em matemática, ciências e geografia, ao passo que as meninas apresentaram melhor desempenho em português.

Em seu trabalho com dados para o ensino superior, Diaz (2007) encontrou resultados interessantes para a escolaridade dos pais. Ao comparar instituições em que a escolaridade média dos pais é diferente, a autora encontra que os alunos cujos pais não fizeram faculdade se beneficiam mais da alta escolaridade dos pais dos colegas, do que do fato do pai ter cursado faculdade se os pais dos seus colegas têm baixa escolaridade. O trabalho encontrou ainda que os alunos com melhor desempenho são aqueles cujas famílias recebem entre 20 e 50 salários mínimos, sendo seu desempenho superior ao dos alunos mais ricos. Existe uma relação não linear entre o a renda da famílias e o desempenho dos aluno.

Utilizando como variável para background socioeconômico a escolaridade dos pais, o trabalho de Rios-Neto, César e Riani (2003) encontrou que tanto a escolaridade da mãe quanto a do pai são significativas para explicar a progressão escolar. Notam que o coeficiente da escolaridade do pai aumenta com o aumento da escolaridade do filho quando comparados alunos da 5ª e 1ª série. O coeficiente da escolaridade da mãe de alunos da 1ª série varia com a introdução no modelo de variáveis das escolas e do município. Parte do efeito explicado coeficiente da escolaridade materna passa a ser explicado pelo aumento de horas-aula da escola.

Sobre a escolaridade dos pais, Barros *et alli* (2001) dizem que esta pode ser um importante fator de redução dos custos educacionais para os filhos. Um resultado interessante que os autores encontraram é que ao trabalhar com amostras separadas para homens e mulheres, a educação da mãe impacta mais a escolaridade das filhas (0,21 ano por ano de escolaridade da mãe) e a escolaridade do pai a dos filhos (0,4 anos por ano de escolaridade).

Alguns autores têm também apontado para a importância de características genéticas no aprendizado e a relação destas com características de *background* socioeconômico. Miller et alli (2001) utilizaram uma amostra de gêmeos monozigóticos e dizigóticos e concluíram que as características genéticas explicam cerca de 50% da variação em anos de estudo,

corroborando uma linha de pesquisa de diversos autores de comportamento genético¹⁴, e entre 25 e 50% o impacto do *background* socioeconômico.

Em relação ao impacto da família sobre o aluno, Francesconi et alli (2005) avaliam em uma amostra da Alemanha o impacto das rupturas familiares sobre as notas, tentando compreender os resultados obtidos por pesquisadores americanos¹⁵ que encontraram um impacto negativo da ocorrência de divórcio sobre o desempenho dos filhos. Os autores testaram diferentes situações de ruptura familiar e concluíram que a ocorrência de divórcio tem um impacto negativo sobre o desempenho do aluno, mas que a morte do pai tem impacto insignificante ou até positivo.

2.5 Comunidade

Uma outra possibilidade de explicação para o desempenho escolar é o impacto do capital humano e social¹⁶ acumulado em diferentes comunidades e municípios. Rios-Neto, César e Riani (2003) testaram a significância da educação da população adulta dos municípios sobre a taxa de progressão escolar dos alunos. Para todas as séries testadas, foi encontrada uma relação positiva. Entretanto é difícil saber se este efeito é decorrente de uma melhor qualidade dos funcionários da escola e do município, o que caracterizaria maior capital humano ou da interação dos alunos com uma população adulta mais educada, o que caracterizaria a presença de capital social.

A importância da comunidade¹⁷ também foi ressaltada por Barros *et alli* (2001) que utilizou a escolaridade média dos adultos e sua renda per capita como variáveis para explicar a escolaridade dos indivíduos. Segundo regressão feita com dados da PNAD, o aumento de um ano da escolaridade média da população adulta aumenta em 0,14 anos a escolaridade da população de 11 a 25 anos. Isto corresponde à cerca de metade do impacto da educação dos pais. Por outro lado, um aumento de R\$215 diminui a escolaridade média dos indivíduos em um ano.

Outro impacto verificado pelos autores é a questão da região de residência dos indivíduos. Mesmo ao se controlar para características familiares, escolares e comunitárias, o fato dos indivíduos estarem no Nordeste afeta negativamente sua escolaridade em relação aos

¹⁴ Taubman (1976), Behrman e Taubman (1989) e Baker et alli (1996) *apud* Miller et alli (2001) apontam um impacto entre 40 e 80% dos fatores genéticos nos anos de estudo.

¹⁵ McLanahan e Sandefur (1994) e Haveman e Wolfe (1995)

¹⁶ Segundo Coleman (2000) o capital social é um bem derivado das relações interpessoais.

¹⁷ Neste caso, uma unidade censitária da PNAD correspondente a 300 domicílios.

que estão no Sudeste. Ao dividirem a amostra por região (Sudeste e Nordeste) a maioria das variáveis referentes aos professores se tornam significativas e apontam um coeficiente maior para o Nordeste.

2.6 Outros Fatores

Questões institucionais podem ser ainda relevantes para o desempenho dos alunos, melhorando o impacto da escola, como o ponto levantado em Figlio e Kenny (2006) em relação ao monitoramento dos professores pelos pais. Os autores corroboram a idéia de que há menor envolvimento dos pais de baixa renda com as escolas de seus filhos. Os autores vêem isso como uma maior possibilidade para experimentação nas escolas de baixa renda ou para maiores ganhos com incentivos para os professores, o que poderiam ser explicações alternativas para o impacto maior dos incentivos no desempenho dos alunos dessas escolas.

Segundo Hanushek (2003) um ponto pouco explorado pela literatura é a existência de incentivos dentro das escolas, o que pode ser uma linha de política pública mais interessante do que o simples aumento de gastos escolares. Segundo o autor, o aumento de gastos é pouco relacionado com a performance dos alunos. Basicamente, existem três tipos de incentivos que têm sido tentados: o pagamento por mérito para professores e escolas, o uso da iniciativa privada para prover parte ou todo serviço das escolas e o aumento da possibilidade de escolha dos alunos entre escolas, aumentando a competição entre escolas. A literatura não tem ainda se aprofundado na discussão sobre o desenho de incentivos, o que deixa em aberto todo um campo de pesquisa.

Woessman (2000, 2001) fazendo uma análise entre diversos países com dados do TIMSS, não encontra uma relação significativa entre o desempenho dos alunos e a infraestrutura escolar. Porém, encontra uma relação significativa e positiva para a competição das escolas particulares e do nível de descentralização das decisões, o que advoga em favor de uma explicação também institucional para o desempenho escolar.

O trabalho de Ludwig (1997) examina o problema da alta evasão escolar entre as classes pobres americanas a partir de uma hipótese de informação imperfeita sobre o mercado de trabalho. Apesar do aumento extraordinário das taxas de escolaridade nos EUA na última metade do século XX, em certas regiões ainda se observa taxas de evasão de 50 a 60% entre os mais pobres. Ludwig (1997) demonstra que parte da explicação para esse fenômeno se deve a subestimação dos requisitos educacionais do mercado de trabalho por parte de adolescentes pobres em decorrência do nível inferior de informação a que tem acesso e ao

isolamento geográfico das regiões mais carentes. O autor utilizou dados do NLSY¹⁸ de 1979 para jovens de 14 a 21 anos que continham informações referentes a desempenho educacional, *background* socioeconômico, aspiração profissional e educacional e notas de exames para admissão nas forças armadas, que serviam como *proxy* para habilidade individual.

Para testar a hipótese de informação imperfeita, o autor construiu indicadores de dispersão entre as informações de aspiração educacional e profissional dos jovens e a escolaridade requerida para o exercício das referidas profissões, medidas pela CPS¹⁹. Os dados mostraram que de maneira geral (81%) os jovens subestimam a escolaridade necessária para o mercado de trabalho. Os indicadores de qualidade foram estimados controlando-se por *background* socioeconômico e residência em áreas pobres, apresentando uma relação significativa com estas variáveis, indicando que o fato de morar em áreas pobres impacta negativamente a qualidade da informação sobre o mercado de trabalho. Foram ainda incorporadas variáveis referentes à participação dos pais no mercado de trabalho, como a proporção de *white collars* na ocupação dos pais, que se mostrou significativa.

Regredindo a taxa de evasão escolar no ensino médio contra o indicador de qualidade de informação, residência em áreas pobres, *background* socioeconômico e taxa de evasão na escola, foi verificada uma relação positiva e significativa para o indicador de informação, insignificante para a residência em área de pobreza e significativa para a taxa de evasão da escola, o que o autor aponta como sendo o efeito dos colegas sobre o aluno. As menores taxas de evasão se encontram entre as minorias (negros e hispânicos) depois de controlada pelas outras variáveis.

Em Severnini (2006) encontramos mais uma abordagem alternativa da questão educacional que parece bastante pertinente para o caso brasileiro. O autor examina o efeito da violência nas escolas através de dados do SAEB de 2004. Severnini (2006) construiu um modelo utilizando uma variável construída a partir das informações sobre a ocorrência de incidentes violentos nas escolas e variáveis de interação entre várias características e os índices de violência.

Os resultados encontrados indicam que a violência impacta negativamente de 1,8 a 3,1% o desempenho dos alunos. A única variável de interação que se mostrou significativa foi

¹⁸ National Longitudinal Surveys (NLS) – pesquisa realizada com amostras de homens e mulheres em diferentes pontos do tempo relativas à informação sobre a participação dos indivíduos no mercado de trabalho.

¹⁹ Current Population Survey (CPS) – pesquisa mensal realizada em 50.000 domicílios americanos conduzida pelo Bureau of the Census for the Bureau of Labor Statistics, semelhante à PNAD brasileira.

a de pós-graduação dos professores e violência, mostrando que professores com maior índice de escolaridade podem atenuar os efeitos da violência.

A diferença na proficiência da escola mais violenta e a menos violenta para o primeiro decil da distribuição de proficiência é de 1%, enquanto que para o último decil é de 2,9%. Há um impacto positivo dos professores com menos de 10 anos de experiência em atenuar o impacto da violência como pode ser observado através da variável de interação. A proporção de alunos pobres na turma, medida pela concessão de bolsa-escola, e o número de alunos na sala de aula ampliam o efeito da violência sobre os alunos. Foi ainda analisado o efeito indireto da violência via aumento da rotatividade dos professores nas escolas mais pobres. Percebeu-se que há um aumento de 7,5% na probabilidade de rotatividade dos professores quando comparamos a escola menos e a mais violenta.

Outro aspecto que tem sido abordado é o que tem sido chamado de *accountability* das escolas, referindo-se geralmente à avaliação de escolas através de testes padronizados em que se mede a escola a partir do desempenho dos alunos como o SAEB, a Prova Brasil, ou ainda o ENEM²⁰. A esse respeito, um trabalho interessante é o de Donovan et alli (2006) que analisa o efeito de um exame padronizado sobre os hábitos de estudo e as notas dos alunos.

Os autores usaram uma amostra de 2.064 alunos de 249 escolas públicas dos coortes de 1999 a 2001 de alunos americanos. Houve a introdução de um exame padronizado em algumas escolas em 2000 e este exame foi alterado em 2001, passando de um teste que avaliava o conhecimento de conteúdos minimamente necessários para um que avalia os alunos para um conjunto de conhecimentos mais extenso e mais específico, com maior seletividade. Os autores dispunham das notas dos alunos no primeiro ano da universidade e informações referentes à antecedência com que eles fazem o *downloads* de arquivos necessários para estudar para as provas. Os alunos não sabiam que suas informações eram coletadas e os nomes do estado, das escolas e da universidade foram mantidos em sigilo.

Os resultados encontrados apontam que deixar para estudar na “última hora” impacta negativamente o desempenho *ceteris paribus* e que os coeficientes para a existência de um exame padronizado são negativos para o desempenho, salvo no caso do desempenho em matemática no caso do exame mais seletivo de 2001. A existência dos exames também impacta positivamente a probabilidade de deixar para a “última hora”, sendo que o exame de 2001 tem um impacto maior que o de 2000, o que é um resultado intrigante, do qual o trabalho não oferece mais informações. Conclui-se do trabalho que exames padronizados que

²⁰ ENEM – Exame nacional de avaliação do ensino médio que consiste em uma prova aplicada aos anos que estão concluindo ou que já concluíram o ensino médio cuja participação é voluntária.

avaliam conhecimentos mínimos são uma política contraproducente. Os autores não encontraram meios para explorar qual teria sido o desempenho dos alunos caso os exames padronizados não fossem aplicados.

Há ainda a análise de programas que têm focado no “casamento” escola-aluno, como em Lavy (2006), para explicar o desempenho escolar. Ele analisou o programa iniciado em 1994, TASC²¹, da cidade de Tel Aviv, que permitiu aos alunos freqüentar escolas de outros distritos, o que antes não era permitido²². O autor comparou os alunos do distrito que recebeu o tratamento com o distrito vizinho, fazendo o *matching* entre alunos dos dois distritos que estivessem até 3000m da divisa, o que transformou o problema em um experimento natural. Contando com dados em painel, o autor utilizou o método de diferenças-em-diferenças para analisar o efeito do tratamento.

Os resultados apontam para uma queda na evasão escolar, um aumento na média das notas dos exames vestibulares e um maior taxa de participação nos exames vestibulares. Os resultados foram ainda mais fortes nos grupos menos favorecidos, medidos pela baixa escolaridade paterna: uma queda de 7% na taxa de evasão escolar, um aumento de 6,2% na taxa de participação nos exames vestibulares e um aumento de 6% na média das notas entre outros. Os resultados mostraram-se ainda diferentes para meninos e meninas, notando-se ganhos maiores para os meninos. Os efeitos positivos do programa se devem ao aumento de competição entre as escolas e da motivação intrínseca dos alunos.

Alguns trabalhos desenvolvidos no Brasil para a análise de programas específicos implantados por diferentes esferas de governo têm ganhado bastante destaque. Menezes-Filho et alli (2005), por exemplo, encontraram um resultado positivo para o regime de ciclos aplicado nas escolas brasileiras a partir de um *matching* com *propensity score* aplicado sobre os uma base de dados montada com o SAEB de 2001 e o Censo Escolar de 2002. Seus resultados indicam um impacto negativo do regime sobre a taxa de evasão escolar e a taxa de reprovação dos alunos. A redução nestas taxas é mais proeminente entre as escolas que adotaram o regime de ciclos para mais anos escolares e que têm os ciclos mais longos.

²¹ Tel Aviv School Choice Program: Programa do governo de Tel Aviv de 1994 em um de seus distritos que permitia aos alunos escolher outras escolas de ensino médio, incluindo uma fora de seu distrito. Antes os alunos podiam escolher outras escolas, mas caso escolhessem uma diferente da mais próxima de sua casa, o governo não ofereceria os passes de ônibus, o que na prática, limitava a escolha dos alunos.

²² Sobre o efeito de programas que aumentam as possibilidades de escolha dos alunos ver Rouse(1998), Peterson et alli (1998), Mayer et alli (2002), Angrist et alli (2002), Krueger and Zhu (2004), Cullen, Jacob e Levitt (2003) e Hoxby (2005)

Os autores encontraram uma redução significativa de 2,3 pontos percentuais na taxa de evasão e um aumento de 7,9 pontos percentuais na taxa de aprovação dos alunos. Percebeu-se ainda um transbordamento dos efeitos do ensino fundamental, em que foram adotados os ciclos, para o ensino médio, diminuindo a taxa de evasão. Ao mesmo tempo, o efeito foi não significativo sobre o desempenho dos alunos.

Anuatti, Fernandes e Pazello (2002) analisaram o impacto do FUNDEF do governo federal sobre a variação dos salários dos professores da rede pública no período de 1997 a 1999 a partir de dados da PNAD. Utilizando simulações de remuneração dos professores caso estivessem empregados no setor privado, na rede privada ou no funcionalismo público, apontam que a variação da sua remuneração no período foi positiva e que para os professores da rede municipal, maior do que se estivessem empregados nos outros setores.

Se forem observadas as médias do Brasil urbano, houve uma variação de 30% nos salários dos professores de 1ª a 4ª série da rede municipal e de 16% na rede estadual, maior que a variação simulada para os outros setores. Para os professores de 5ª a 8ª série, a variação foi de 25%, bem acima dos outros setores, entretanto, para os professores da rede estadual, esta variação ficou abaixo da variação dos salários simulada para a rede privada de ensino.

Ao comparar os professores da rede estadual com a municipal, perceberam uma aproximação dos salários, sendo que as variações foram maiores para os professores da rede municipal, isto pode ser percebido principalmente no nordeste, onde a desigualdade salarial dos professores das duas redes caiu de 67 para 30%. Foi ainda observada uma diminuição maior da desigualdade salarial entre professores da rede estadual e municipal nos menores municípios, os classificados como não auto-representativos pela PNAD. Conclui-se que o FUNDEF beneficiou majoritariamente a rede municipal de ensino de 1ª a 4ª que está sobre-representada nos municípios mais pobres.

Fernandes e Nantezon (2002) analisaram a hipótese de que o rendimento dos alunos brasileiros da rede pública medido pelo SAEB caiu na segunda metade da década de 90 em razão da mudança do perfil dos alunos devido a uma maior cobertura do sistema educacional, tendo em vista a queda nas taxas de repetência e evasão escolar nesse período. Os autores desenvolveram um índice que analisa o percentual de alunos na mediana, sexto decil e sétimo decil da distribuição de notas, excluindo alunos que não estivessem na série correta. A análise do índice para os alunos da 4ª série de 1995, 1997 e 1999 mostrou que o desempenho dos alunos na série correta melhorou ligeiramente ao longo dessas gerações, ao contrário da hipótese amplamente difundida citada acima. Os resultados dos autores apontam uma evolução de 9,8% em português e 7% em matemática no período analisado.

Percebe-se pelo exposto acima que há uma grande diversidade de métodos e abordagens utilizadas na Economia da Educação. Os modelos desenvolvidos pelos pesquisadores vão desde simples funções de produção lineares até sofisticados modelos hierárquicos. O uso de variáveis também é bastante diversificado, porém alguns pontos de atenção parecem já ser consenso.

Pode-se incorrer em endogeneidade, como no caso de bons professores que escolhem lecionar nas escolas com melhores alunos ou ainda colinearidade como na correlação entre infra-estrutura escolar e *background* socioeconômico. Outra questão importante é a ocorrência de viés de agregação, em que a amostra não está aleatoriamente dividida, mas em grupos, o que impacta os coeficientes das variáveis explicativas. Isso é particularmente importante se buscamos compreender qual o impacto das escolas enquanto instituições sobre o desempenho dos alunos ou das características de um determinado município sobre estes.

Um outro ponto é a seleção de uma variável dependente apropriada. A quase universalização de alguns níveis educacionais torna difícil a análise do impacto da educação a partir de medidas de escolaridade, dada a pouca variação dessa medida. Por outro lado, variáveis de valor adicionado, que parecem ser as preferidas entre os pesquisadores são de difícil acesso, uma vez que torna necessária a obtenção de dados em painel. Medidas de aprovação ou exames são extremamente subjetivos e tornam difícil a comparação em grandes amostras. Sendo assim, o uso do resultado de testes padronizados tem se proliferado e oferece uma alternativa razoável a outras variáveis de difícil mensuração ou baixa variabilidade.

Tendo em vista estas questões levantadas, este trabalho fará uso de medidas de desigualdade para abordar o tema, como forma de evitar alguns dos obstáculos já identificados pela literatura.

3. Metodologia

Como já foi colocado, este trabalho pretende identificar quais são os fatores que mais influenciam o desempenho dos alunos das 4^{as} séries do estado de São Paulo. Para proceder à análise, utilizaremos os dados da Prova Brasil para alunos das 4^{as} séries de português e matemática avaliadas neste exame. Para compreender qual o impacto de diferentes variáveis sobre o desempenho dos alunos a partir da base de dados da Prova Brasil e do Censo Escolar, utilizaremos um índice de desigualdade, o Theil – L, procurando captar através da diferença de desempenho entre os alunos os fatores mais importantes para explicar a proficiência em português e matemática dos alunos.

O índice de desigualdade que será utilizado pertence à classe dos índices de entropia generalizada. Os índices de entropia se desenvolvem a partir da diferença entre a proporção que um indivíduo possui de uma variável de bem-estar (a somatória das notas, no nosso caso) e a sua proporção na população. O índice será aplicado sobre os microdados da Prova Brasil que foram pouco explorados pela literatura até agora dado a sua recente disponibilidade e acreditamos que o uso de uma base nova com um método que têm sido ainda pouco explorado na literatura de Economia da Educação pode oferecer uma contribuição relevante para o tema.

A Prova Brasil foi realizada pela primeira vez em 2005 e seus microdados foram disponibilizados em 2007. É uma base de dados nova e muito se espera das contribuições que pode oferecer para a pesquisa em Economia da Educação no Brasil. Segue abaixo uma breve apresentação da Prova Brasil e posteriormente uma exposição do índice que será utilizado nesse trabalho.

3.1 Base de Dados

Hoje há uma grande disponibilidade de dados que permitem uma análise profunda do estado e dos determinantes da educação no Brasil. Existem várias fontes de dados governamentais cujas tabulações são de fácil acesso para os pesquisadores, tais como os dados do Sistema de Avaliação da Educação Básica (SAEB)²³, do CensoEscolar e da Prova

²³ O SAEB consiste em uma prova aplicada a cada dois anos no Brasil em alunos da quarta e da oitava série do Ensino Fundamental e em alunos do terceiro ano do Ensino Médio. Além dos resultados nas provas, obtem-se por meio de questionário respondido pelo aluno, dados sobre comportamento com relação à escola, *background* socioeconômico do aluno e o ambiente na sala de aula. Há ainda dados referentes ao questionário aplicado às escolas, respondido pelos diretores. Nestes encontram-se dados sobre a infraestrutura escolar.

Brasil, sendo que todos esses bancos de dados são desenvolvidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP)²⁴.

A Prova Brasil consiste em uma avaliação de português e matemática e um questionário dirigido aos alunos sobre seu contexto sócio-econômico e cultural. Na sua primeira edição, foram coletados dados de 5.387 municípios de todos os estados do Brasil e 3.392.880 alunos de 4ª e 8ª séries do ensino fundamental. São consideradas apenas escolas da rede pública urbana e é necessário que tenham mais de 30 alunos matriculados na série avaliada.

Os dados oferecidos pela Prova Brasil são uma extensão do SAEB que permite a desagregação das notas por municípios, além da desagregação por alunos, escolas e unidades da federação que já era possível pelo SAEB. Devido a esta nova possibilidade de desagregação dos dados, pode-se proceder a análises de distribuição e à construção de índices para análise intra e entre escolas dentro do mesmo município, o que antes não era possível. Assim, podemos observar qual o impacto de diferentes municípios e estabelecimentos escolares no desempenho dos alunos²⁵. Como as escolas dentro do mesmo município tendem a ter um conjunto maior de características comuns do que duas escolas aleatoriamente escolhidas, isso permite que através da comparação entre elas, possamos isolar melhor em que estas duas escolas diferem, superando o viés de agregação, conforme apontado por Albernaz, Ferreira e Franco (2002) e lidando melhor com variáveis não-observáveis. Da mesma maneira, comparar alunos dentro de uma mesma escola facilita a identificação das características em que eles diferem. A possibilidade de agregar alunos também por municípios é uma das grandes vantagens desta base de dados.

3.2 Índices de Entropia

Os índices de desigualdade têm sido amplamente utilizados em diversos trabalhos para verificar os determinantes de diferentes variáveis de bem-estar, principalmente renda. Também na Economia da Educação, estes índices têm sido utilizados, como em Felício

²⁴ O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) é uma autarquia federal vinculada ao Ministério da Educação (MEC), que visa à obtenção e o desenvolvimento de dados que possibilitem a formulação de políticas públicas para a educação no Brasil.

²⁵ A respeito da construção de índices e desagregação dos microdados da Prova Brasil, ver, por exemplo, Gremaud, Felício e Biondi (2006).

(2004) e Thomas et alli (2002), porém seu uso não têm sido tão difundido como para as análises de desigualdade de renda. Acreditamos que essa análise pode contribuir para superar as dificuldades encontradas nas pesquisas de economia da educação relacionadas à multicolineariedade, autocorrelação e variáveis não-observáveis.

Da mesma forma que os problemas de renda, a análise da desigualdade de desempenho escolar pode oferecer informações importantes. Para isso, um procedimento que pode ser bastante útil é o mapeamento dos insumos, fatores e características associados a um desempenho melhor.

Como colocado acima, o índice utilizados neste trabalho será uma das medidas de entropia de Theil. Esse índice possui características que são particularmente desejáveis para identificar fatores associados a desigualdades. Além de oferecer uma medida de desigualdade, esse índice é decomponível, o que lhes dá alguma vantagem em relação a outros índices, como o índice de Gini²⁶.

A possibilidade de decompor os índices de Theil é uma característica particularmente desejável. Esta nos permite observar quanto da desigualdade total medida se dá dentro de um subgrupo particular, digamos entre os alunos negros, e quanto da desigualdade se deve à uma determinada variável, por exemplo, raça. Para proceder essa análise, utilizaremos neste trabalho os conceitos de contribuição marginal e bruta.

Podemos subdividir a amostra em diversos subgrupos e observar o impacto das diferentes variáveis. Se dividirmos os alunos por raça, por exemplo, podemos observar quanto da desigualdade se dá entre os indivíduos de diferentes raças e quanto se dá dentro do grupo da mesma raça. Essa parcela da desigualdade devida à desigualdade entre o grupo de negros e brancos, chamamos de contribuição bruta da variável raça para a desigualdade total.

Podemos ainda subdividir a amostra em subgrupos delimitados por diversas variáveis, como por exemplo, raça, sexo e idade. Se quisermos saber o impacto da variável raça condicionado a essas outras variáveis, podemos verificar qual a desigualdade entre os grupos gerados pelas variáveis escolhidas, retirar a variável raça, verificar novamente qual a desigualdade entre os subgrupos e verificar a diferença. Essa diferença é o que chamaremos de contribuição marginal da variável raça.

Segue agora uma discussão detalhada sobre o índice que será utilizado:

²⁶ Para uma discussão detalhada sobre índices de desigualdade, uma boa referência é Hoffman (1992)

3.3 Índices de Theil

O índice de Theil foi criado em 1967 por Henry Theil como parte de seus trabalhos sobre economia e teoria da informação. O índice que desenvolveu utiliza os conceitos de entropia e conteúdo informativo de uma mensagem.

Segundo Hoffman (1992), o conteúdo informativo de uma mensagem se refere a quanto a ocorrência de um evento adiciona de informação para o observador tendo em vista a sua probabilidade *a priori* de ocorrência. Assim, se há uma probabilidade de 0,2 que chova, o evento “choveu”, que tem probabilidade *a posteriori* igual a 1, adicionou informação que pode ser captada pela diferença $1 - 0,2$.

Uma maneira interessante que Hoffman (1992) utiliza para falar sobre o conteúdo informativo de uma mensagem é tratá-la como uma espécie de surpresa pela ocorrência do evento. Assim, pode-se pensar que esta surpresa deve ser uma diferença entre uma medida de ocorrência do evento menos uma medida da informação que nos permitia antecipá-la, tornando-a, por assim dizer, menos surpreendente.

Iremos apresentar brevemente alguns conceitos de teoria da informação como apresentado em Hoffman (1992) para podermos desenvolver como faz o autor o conceito central por trás do índice de Theil.

Matematicamente, o conteúdo informativo de uma mensagem de ocorrência de um evento é medido em teoria da informação como:

$$h(x) = \log \frac{1}{x}$$

em que x é a probabilidade *a priori* do evento que ocorreu. Havendo duas probabilidades, uma *posteriori* e outra *a priori*, é possível medir o conteúdo informativo de uma previsão. Sendo x e y respectivamente as previsões *a priori* e *a posteriori*, temos:

$$h(x) - h(y) = \ln \frac{1}{x} - \ln \frac{1}{y} = \ln \frac{y}{x}$$

Podemos pensar ainda no conteúdo informativo para um universo de n eventos mutuamente exclusivos A_i ($i = 1, 2, \dots, n$) em que a cada evento é associada uma probabilidade x_i , de tal maneira que $\sum_{i=1}^n x_i = 1$. A esperança matemática do conteúdo informativo da mensagem ocorreu A_i é dada por:

$$H(x) = E[h(x_i)] = \sum_{i=1}^n x_i h(x_i) = \sum_{i=1}^n x_i \log \frac{1}{x_i} = -\sum_{i=1}^n x_i \log x_i \quad 27$$

Quando a probabilidade de um evento é igual a 1 e as demais são nulas, isto é, têm-se certeza do que vai ocorrer, tem-se o valor mínimo de $H(x)$, que é zero.

O valor máximo que pode ser atingido por $H(x)$ ocorrerá quando todas as probabilidades forem iguais, o que se pode traduzir como não saber qual evento A_i ocorrerá.

Matematicamente isto representa que $x_i = \frac{1}{n}$, para todo i . Portanto,

$$H(x) = \sum_{i=1}^n x_i \log \frac{1}{x_i} = \sum_{i=1}^n \frac{1}{n} \log n = \log n$$

Assim,

$$0 \leq H(x) \leq \log n$$

A esta esperança do conteúdo informativo de uma distribuição de eventos se dá também o nome de entropia da distribuição.

Façamos agora a substituição da variável x_i pela proporção que os indivíduos têm de uma determinada variável de bem-estar, no nosso caso, esta variável é a proporção que a nota do indivíduo representa do somatório de todas as notas, y_i . Mantendo as mesmas condições que tínhamos para x_i , podemos dizer que:

$$H(y) = \sum_{i=1}^n y_i \log \frac{1}{y_i} \text{ e que } 0 \leq H(y) \leq \log n,$$

uma vez que $\log n$ seria a entropia da distribuição para o caso de todos os eventos terem a mesma probabilidade, o que neste exemplo equivale a dizer que todos indivíduos teriam a mesma nota, podemos entender $H(y)$ como uma medida da desigualdade de desta variável.

Fazendo:

$$T = \log n - \sum_{i=1}^n y_i \log \frac{1}{y_i}, \text{ temos uma medida de desigualdade da distribuição de notas}$$

que é máxima quando a desigualdade é máxima, $H(y)$ é mínima, e mínima quando a desigualdade é mínima, isto é, $H(y)$ é máximo. E assim chegamos à primeira formulação do índice de Theil, conhecido também como Theil-T.

Podemos ainda observar que:

²⁷Se $x_i = 0$, define-se $x_i \log x_i = 0$, já que $\lim_{x_i \rightarrow 0} x_i \log x_i = 0$

$$T = \sum_{i=1}^n y_i \log ny_i = \sum_{i=1}^n y_i \log \frac{y_i}{\frac{1}{n}},$$

isso nos permite observar que o índice Theil-T trabalha com a informação dada pela mensagem oferecida pela probabilidade a *posteriori* dada pela distribuição de renda a partir da probabilidade a *priori* dada pela participação no indivíduo na população, isto é, supõe-se a *priori* que todo indivíduo deveria ter uma participação igual na renda. Pode-se dizer que o índice de Theil mede a “surpresa” que temos ao observar a distribuição de notas supondo que esperávamos que as notas fossem iguais para todos os indivíduos, equivalente à sua igual representatividade na população.

A ponderação dos indivíduos pela sua participação no somatório das notas tem algumas implicações importantes. Primeiro, os indivíduos não terão uma representação igualitária no índice. Os indivíduos de maior nota contribuirão mais para o índice de desigualdade. Outro ponto importante, é que ocorrendo uma mudança na nota dos indivíduos ao longo do tempo, a medida de desigualdade sofrerá dois impactos: um primeiro devido a alteração na proporção de cada indivíduo no total das notas e outro pela alteração nos fatores de ponderação, que são dados pelas mesmas proporções. Sendo assim, a análise da desigualdade de uma distribuição das notas ao longo do tempo, exige mais do que a simples comparação dos índices.

Assim, uma formulação bastante útil do índice de Theil é a seguinte:

$$L = \sum_{i=1}^n \frac{1}{n} \log \frac{1}{y_i}, \text{ em que invertemos as probabilidades a } \textit{priori} \text{ e a } \textit{posteriori} \text{ e,}$$

portanto, os fatores de ponderação. Esta segunda formulação é chamada o índice Theil-L ou desvio médio dos logaritmos. A ponderação dada pela proporção na população dos indivíduos é mais “democrática”, como aponta Hoffman (1992). Esta formulação confere a cada indivíduo uma participação igual no índice. Além do mais, no caso de redistribuição de renda, as medidas anterior e posterior à redistribuição são imediatamente comparáveis. Um outro aspecto importante, é que, geralmente, por dar representatividade igual aos indivíduos, as parcelas mais pobres da população acabam sendo melhor representados do que seriam pelo índice Theil-T, já que tendem a possuir valores menores para variáveis de bem-estar.

Uma forma intuitiva de se pensar o índice Theil-L, é imaginá-lo como uma medida da informação que obtemos observando a desigualdade na proporção do total de notas de cada

indivíduo relativamente à hipótese de que todos indivíduos tivessem a mesma proporção, isto é, notas iguais.

A partir do índice Theil-L, podemos observar como se dá a decomposição do índice para a análise de diferentes determinantes de desempenho escolar. Como já foi dito, a possibilidade de se decompor os índices de Theil é interessante porque nos permite observar quanto da desigualdade captada pelo índice se deve à desigualdade entre diferentes grupos da amostra e quanto se deve à desigualdade dentro desses grupos.

3.4. Decomposição da Desigualdade

Até aqui, a desigualdade foi abordada sempre entre as observações, entretanto, podemos separar as observações em grupos e observar a desigualdade entre esses grupos. Nesse caso, observamos quanto da desigualdade de notas se deve aos diferentes grupos, lembrando que neste caso não se fala de nota média, mas do quanto da proporção do somatório das notas pertence a cada grupo. Sendo as observações divididas em k grupos, π_h a proporção do h -ésimo grupo na população e y_h a proporção do total das notas que possui o grupo, podemos fazer:

$$L_e = \sum_{h=1}^k \pi_h \log \frac{\pi_h}{y_h}, \text{ o que mede a desigualdade entre os } h \text{ grupos formados a partir}$$

da amostra.

Com a equação acima, obtemos qual a desigualdade entre os h grupos formados. É fácil compreender que L_e é fração de L e que ainda resta uma parcela de L a ser explicada que se refere à desigualdade de cada grupo h . Um exemplo ilustrativo do que seriam esses k grupos pode ser dado pela divisão por gênero. Neste caso, a variável gênero possui dois valores, sendo que assim $h = \{\text{masculino}, \text{feminino}\}$ e $k=2$.

Ao fazermos a subtração $L - L_e$, obtemos qual a parcela da desigualdade que se deve a variação dentro dos grupos. Somando e subtraindo L_e , podemos rearranjar L de tal maneira a ser representado como a soma de duas desigualdades: L_e a desigualdade entre grupos e L_h a desigualdade dentro de cada grupo.

$$L = \sum_{h=1}^k \pi_h \log \frac{\pi_h}{Y_h} + \sum_{h=1}^k \pi_h \frac{1}{n_h} \sum_{i=1}^{n_h} \log \frac{Y_h}{n_h y_{hi}}, \text{ sendo que, portanto,}$$

$$L_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \log \frac{Y_h}{n_h y_{hi}} \text{ e}$$

$$L_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \log \frac{Y_h}{n_h y_{hi}}$$

Podemos assim observar o impacto da desigualdade entre grupos e intra grupos sobre a desigualdade medida pelos índices Theil. Um aspecto bastante útil dessa decomposição é o fato das parcelas L_e e L_h serem proporções da desigualdade total. Assim, pode-se dizer que a partir da divisão da amostra em grupos de acordo com os critérios de uma variável qualquer, conseguimos identificar L_e / L da desigualdade. E a L_e daremos o nome de **contribuição bruta** da variável cujos valores delimitam os h grupos em que foi dividida a amostra.

A divisão em grupos pode ser feita a partir do uso dos critérios de várias variáveis, pode-se ao invés de dividir a amostra somente por gênero, por exemplo, proceder a divisão da amostra por gênero e raça. Supondo que haja duas raças. Branco e não-branco, teríamos neste exemplo 4 grupos formados. Poderíamos realizar a decomposição de L em L_e e L_h , o que nos daria respectivamente qual a contribuição da desigualdade entre os grupos de raça e gênero e da desigualdade dentro desses grupos para a desigualdade total.

Entretanto, a partir do exemplo acima, como observar quanto da desigualdade total se deve à desigualdade entre os diferentes gêneros somente? Para responder a esta pergunta, temos que fazer a decomposição de L como dissemos anteriormente, obtendo L_e e L_h . Depois, realizamos uma nova divisão da amostra, excluindo os critérios da variável cujo impacto queremos observar, neste exemplo retirariamos o gênero e faríamos a divisão dos grupos de acordo com raça, e obtemos L_e' e L_h' , notando que L permanece o mesmo.

A contribuição da variável gênero para desigualdade dentro uma amostra dividida por raça e gênero será dada por : $L_e - L_e'$. A essa contribuição dá-se o nome de **contribuição marginal** da variável em questão. O termo marginal aparece por estarmos levando em conta também outras variáveis para a formação dos grupos.

Temos assim as duas formas de análise dos índices de desigualdade que serão utilizadas neste trabalho. A partir da contribuição bruta, temos uma visão geral de qual a importância daquela variável para a desigualdade observada. Já a contribuição marginal, nos permite isolar a importância dessa variável para a desigualdade do impacto das outras variáveis, nos permitindo analisar o impacto dessa variável num contexto de relação entre várias variáveis.

3.5 Formação dos Grupos para as Decomposições

O uso do índice de desigualdade Theil-L nos permite obter de forma não paramétrica informações importantes relativas ao impacto de diferentes variáveis sobre a desigualdade observada em uma determinada variável de bem-estar dos indivíduos. Entretanto, a análise de contribuição marginal tem como limitação o caráter subjetivo da escolha das variáveis para se proceder à criação dos grupos, o que se deve à impossibilidade de se produzir grupos indefinidamente de tal maneira que se possa trabalhar com todas as variáveis. A amostra deve ser dividida de tal maneira a não gerar grupos vazios, demandando um número grande de observações.

A título de ilustração, suponhamos um banco de dados que possua 50.000 observações. O primeiro limite que temos é que o número de grupos não pode ultrapassar 50.000, Se estivermos trabalhando com x variáveis de duas categorias e se pudermos formar 2^x grupos, não poderíamos trabalhar com mais de 15 variáveis, já que $\log_2 50.000 = 15,6$.

Mas também não podemos trabalhar com um número de grupos equivalente ao número de observações, já que nesse caso a desigualdade medida trata-se na verdade não da desigualdade entre os grupos, mas da desigualdade entre as observações. É necessário garantir um certo número de observações por grupo, para que possamos ter variação dentro dos grupos e possamos observar o que é de fato variação entre os grupos.

A utilização dos dados da Prova Brasil é muito importante por se tratar de um banco de dados com muitas observações, nos permitindo gerar grupos com mais variáveis. O SAEB, por exemplo, possui cerca de 50.000 observações para alunos da 4ª série do Brasil, já a Prova Brasil, possui mais de 350.000 observações de alunos da 4ª apenas do Estado de São Paulo.

Neste trabalho, a questão da subjetividade da escolha das variáveis para formar os grupos, será tratada utilizando a contribuição bruta das variáveis como critério de inclusão no grupo formado para calcular contribuição marginal de cada uma. Primeiramente, calculamos a contribuição bruta de todas as variáveis e depois elegemos as que apresentaram um valor maior para a formação dos grupos de análise da contribuição marginal.

Depois de calcular as contribuições marginais, iremos ainda realizar uma estimação por mínimos quadrados para comparar os resultados obtidos.

4. Análise descritiva

Como já foi dito anteriormente, o banco de dados utilizado neste trabalho consiste na junção dos dados da Prova Brasil e do Censo Escolar 2005 para alunos da 4ª série da rede pública estadual e municipal do estado de São Paulo. Da Prova Brasil, utilizamos os dados do questionário socioeconômico e a proficiência em português e matemática dos alunos e do Censo Escolar extraímos os dados referentes às escolas em que eles estudavam.

O banco de dados formado pela soma da Prova Brasil e do Censo Escolar é muito rico em informações e diversos autores têm trabalhado com banco de dados semelhantes, utilizando medidas de proficiência ou desempenho educacional combinados com dados de infraestrutura escolar, como em Barros e Mendonça (1998) e em César e Riani (2003).

Os dados da Prova Brasil estão começando a ser utilizados pelos pesquisadores e, por esse motivo, ganham uma atenção particular. Os dados referentes à Prova Brasil de 2005 só foram disponibilizados este ano, o que torna interessante olharmos com mais atenção para as estatísticas descritivas do banco de dados utilizado neste trabalho. Embora não esteja não seja nossa intenção discutir as qualidades do banco em si, imaginamos que uma análise mais cuidadosa dessas estatísticas pode contribuir para a percepção do ganho que se pode obter na pesquisa em economia da Educação no Brasil à partir desses novos dados.

A amostra que utilizamos no nosso trabalho conta com 352.248 observações, sendo que o aluno é a unidade de observação. Foram utilizadas informações de 4.552 escolas do Censo Escolar correspondentes aos alunos que realizaram a Prova Brasil. A Prova Brasil de 2005 foi amostral no estado de São Paulo e não censitária como em outras partes do país. Além disso, a Prova Brasil avalia alunos apenas das redes públicas federal, estadual e municipal. No estado de São Paulo, sua amostragem cobriu apenas escolas municipais e estaduais.

A nota média em matemática foi de 187,87 pontos e 43,18 o desvio padrão, no primeiro decil da distribuição de notas estão os alunos com nota inferior a 133,38 e no último decil, os com nota superior a 245,31. Os alunos colocados entre os percentis 45 e 55 têm notas entre 179,96 e 190,72 pontos. A menor pontuação em matemática da amostra é 78,09 e a maior 330,65.

A escala de notas adotada pela Prova Brasil é a mesma que é adotada pelo SAEB, o panorama apresentado acima pelas médias e percentis acima significa que os 10% piores alunos em matemática não conseguem associar numerais com a sua representação por decimais ou realizar cálculos de adição com três algarismos. Já os alunos medianos

conseguem efetuar com reserva cálculos de subtração com três Algarismos, multiplicam com reserva com um multiplicador de um Algarismo e resolvem problemas com medidas de metros e centímetros. Os melhores alunos, por sua vez, conseguem efetuar divisões com um divisor de dois Algarismos, resolvem problemas com mais de uma medida e mais de uma operação matemática, entre outras coisas.

A média das nota português é a mais baixa, mas com dispersão parecida. A média é de 181,95 pontos e o desvio padrão é de 45,03, o primeiro decil é ocupado pelos alunos com pontuação inferior a 123,33 e o último com os que estão acima de 241,76. O espaço entre os percentis 45 e 55 é ocupado pelos alunos com notas entre 174,56 e 186, 29. A melhor nota de português na amostra é de 324,61 e a menor 74,83.

As notas de português indicam, por exemplo, que os 10% piores alunos tem dificuldade em compreender a finalidade de um texto curto e identificar quem é o personagem principal. Os alunos medianos conseguem identificar a finalidade de um texto jornalístico e reconhecem os elementos que compõe uma narrativa com temática e vocabulário complexos. Por fim, os melhores alunos conseguem, entre outras coisas, distinguir o sentido metafórico do literal em uma expressão e localizar informação em um texto instrucional de vocabulário complexo.

Foram retiradas de nossa amostra as observações que não continham informação referente ao sexo, raça e idade do aluno. Há um número razoável de informações do Censo Escolar que não foram preenchidas adequadamente pelas escolas, os *missings*, entretanto, optamos por mantê-las já que o método de decomposição por meio do índice Theil-L nos permite trabalhar também com grupos contendo *missings*.

Nas estatísticas descritivas indicamos como “não declarado” os valores que não estão preenchidos adequadamente. O fato de mantermos os grupos de *missings* não prejudica a análise, já que o método utilizado trabalha com a subdivisão da amostra em grupos de acordo com os valores das variáveis e não com estes valores em si. Sendo assim, podemos trabalhar com um grupo de *missings* e observar se, eventualmente, o não preenchimento adequado dos dados do Censo é não-aleatório, explicando parte do desempenho dos alunos.

As estatísticas descritivas que seguem abaixo foram organizadas de forma a poderem ser confrontadas com a literatura do assunto, conforme foi apresentado no capítulo 2. Os dados estão dispostos independentes do seu banco de origem e organizados de forma a apresentar os dados referentes ao *background* socioeconômico dos alunos, escolas, professores, colegas e programas educacionais.

O intuito desse estudo não é o de oferecer uma análise minuciosa de cada uma das variáveis, nosso interesse é mais exploratório, no sentido de apresentar um panorama sobre as possíveis causas da diferença de desempenho entre os alunos.

Nas tabelas abaixo, são apresentadas as médias do desempenho dos alunos em matemática e em português, segundo as categorias das variáveis que julgamos mais relevantes. Foi realizada uma análise prévia com todas as variáveis da Prova Brasil e do Censo Escolar para alunos da 4ª série e aqui são apresentadas as médias que se mostraram mais significativas. Além das médias, apresentamos também o percentual dos alunos medianos, dez por cento melhores e dez por cento piores que possui a característica analisada. Esta disposição dos dados visa formar no leitor uma imagem de como são as escolas paulistas e quais as características das que tem melhor e pior desempenho.

Vejamos então na tabela 1 abaixo o desempenho dos alunos segundo algumas características de seu *background* sócio-econômico:

Tabela 1 - Estatísticas Descritivas Referentes a *Background* Socioeconômico

		MATEMÁTICA				PORTUGUÊS			
		% dos centis				% nos centis			
		média	10	45-55	90	média	10	45-55	90
Sexo	masc	187,70	52,69	51,78	51,69	181,68	52,46	51,80	51,52
	fem	188,04	47,31	48,22	48,31	182,23	47,54	48,20	48,48
Cor	branco	190,97	38,73	42,61	49,19	185,07	38,69	43,43	48,82
	amarelo	186,19	2,63	2,30	2,20	180,33	2,58	2,34	2,16
	pardo	186,04	45,25	43,55	39,43	180,00	45,49	42,95	39,42
	indígena	184,15	4,11	3,79	3,07	179,30	4,07	3,58	3,22
	preto	182,98	9,28	7,75	6,12	177,15	9,18	7,70	6,38
Mora com os pais	sim	188,62	65,69	67,68	69,84	182,77	65,61	68,11	69,76
	não	186,27	34,31	32,32	30,16	180,23	34,39	31,89	30,24
Acima de 11 anos	sim	183,79	10,47	9,31	7,60	177,18	10,53	9,12	7,49
	não	188,27	89,53	90,69	92,40	182,43	89,47	90,88	92,51
Mãe com Ensino Superior	sim	191,17	21,18	23,33	27,69	185,75	21,19	23,28	27,84
	não	186,83	78,82	76,67	72,31	180,76	78,81	76,72	72,16
Pai com Ensino Superior	sim	191,01	20,49	22,79	26,34	185,63	20,49	22,59	26,49
	não	186,93	79,51	77,21	73,66	180,85	79,51	77,41	73,51
Pessoas por quarto	Não tem quarto	181,8	3,76	3,07	2,27	176,2	3,8	3,06	2,46
	até 2	191,48	38,01	42,48	50,84	185,51	37,83	43,50	49,78
	mais de 2	185,28	58,23	54,45	46,89	179,39	58,37	53,44	47,75
Possui carro?	Sim	189,76	51,93	54,67	59,65	184,00	51,70	54,74	60,03
	Não	185,54	48,07	45,33	40,05	179,44	48,30	45,26	39,97
Possui computador com internet	Sim	190,87	18,32	19,63	23,03	185,54	18,39	19,60	23,54
	Não*	187,11	81,68	80,37	76,97	181,05	81,61	80,40	76,46

*inclui computador sem internet

A tabela acima tem dados referentes às variáveis que têm sido mais utilizadas pela literatura referente a *background* sócio-econômico. As médias das notas em português e matemática apresentam o mesmo padrão, sendo que a única diferença são as médias mais baixas das notas dos alunos em português. As médias mais altas são as dos alunos cujos pais possuem educação superior e dos alunos que vivem em casas com até duas pessoas por quarto. Os alunos que possuem essas características estão mais presentes nos último decil das notas e menos presentes no primeiro, a participação deles entre os grupos de pior e melhor performance cresce monotonicamente para ambas as notas.

A variável pessoas por quarto tenta ser uma *proxy* de renda, já que não há uma variável específica a esse respeito na Prova Brasil, tenta ainda observar o nível de vulnerabilidade social em que se encontram as crianças, já que casas com várias pessoas dividindo cômodos como dormitório podem dar uma boa noção da precariedade da qualidade de vida de uma família.

Outras características que parecem importantes são a raça dos indivíduos, a disponibilidade de computador com internet e carro e a idade dos alunos. A importância da raça vai ao encontro do que já dizia Coleman (1966), indicando uma situação de desvantagem em que se encontram os alunos não brancos.

A variável sexo não parece ser muito importante e ao contrário de resultados da literatura, como em Albernaz, Ferreira e Franco (2002), as meninas tiveram nota superior em ambas as disciplinas, quando seria de se esperar que os meninos fossem melhor em matemática.

Olhando somente as médias, a escolaridade dos pais para a nossa amostra parece com o que a literatura tem apontado referente a um dos controles que tem sido mais utilizados em funções de desempenho escolar. Aqui, a educação tanto do pai como a da mãe parecem ser relevantes, como em Rios-Neto, César e Riani (2003) e Barros *et alli* (2001). Entretanto, o *fato* do aluno morar com os pais não parece ter um efeito muito significativo como em Francesconi *et alli* (2005).

Segue abaixo algumas outras estatísticas referentes a variáveis de *background* sócio-econômico:

Tabela 2 - Estatísticas Descritivas Referentes a Background Socioeconômico (cont.)

		MATEMÁTICA				PORTUGUÊS			
		% dos centis				% nos centis			
		média	10	45-55	90	média	10	45-55	90
Frequentou Ensino Infantil	Sim	190,84	31,69	34,49	40,20	184,93	32,48	35,44	40,18
	Não	186,23	68,31	65,51	59,80	180,31	67,52	64,56	59,82
Já foi reprovado	Sim	185,04	19,90	18,16	15,79	178,46	19,49	18,32	15,08
	Não	188,48	80,10	81,84	84,21	182,71	80,51	81,68	84,92
Já abandonou a escola	Sim	182,18	8,85	7,14	5,49	176,33	8,72	7,07	5,70
	Não	188,30	91,15	92,86	94,51	182,38	91,28	92,93	94,30
Faz tarefa	Sim	188,94	70,42	72,82	76,84	183,24	66,18	69,37	72,72
	Não	184,92	29,58	27,18	23,16	179,00	33,82	30,63	27,28
Apoio dos pais	Sim	188,16	92,61	93,59	94,82	182,23	92,73	93,88	94,67
	Não	183,47	7,39	6,41	5,18	177,66	7,27	6,12	5,33
Apoio para fazer a lição de casa	Sim	188,27	89,62	91,22	92,73	182,32	89,90	91,30	92,59
	Não	183,61	10,38	8,78	7,27	177,99	10,10	8,70	7,41
Silêncio para estudar em casa	sim	188,71	77,89	80,16	83,63	182,73	77,85	80,71	82,49
	não	184,38	22,11	19,84	16,37	178,74	22,15	19,69	17,51
Trabalhos domésticos*	sim	184,53	90,28	91,03	92,55	178,71	9,58	8,58	7,74
	não	188,18	9,72	8,97	7,45	182,26	90,42	91,42	92,26
Trabalho infantil**	sim	184,07	14,05	12,45	10,23	177,98	13,93	12,09	10,43
	não	188,94	85,95	87,55	89,77	182,50	86,07	87,91	89,57
Deseja somente estudar***	sim	184,43	9,49	8,12	7,22	178,75	9,37	8,18	7,24
	não	188,17	90,51	91,88	92,78	182,23	90,63	91,82	92,76
Recebe bolsa família	Sim	185,25	22,55	20,93	18,36	178,83	22,69	21,07	18,10
	Não	188,55	77,45	79,07	81,64	182,77	77,31	78,93	81,90

* mais de 3h diárias dedicadas a afazeres domésticos

** exerce alguma atividade relacionada a trabalho fora de casa durante dias de aula

***expressa o desejo de depois de complementar o ensino fundamental, continuar apenas estudando

Algumas variáveis apresentadas na tabela acima como se o aluno faz ou não a lição, se sente apoio dos pais e se desfruta de silêncio em casa para estudar ainda têm sido pouco exploradas pela literatura. As médias, entretanto, não indica que essas variáveis tenham um impacto da magnitude das que foram destacadas na primeira tabela.

Uma variável cuja análise parece ser bastante interessante é o momento do início da vida escolar dos alunos, se no ensino infantil ou não. Os alunos que cursaram o ensino infantil apresentaram as médias mais altas do grupo acima e também a maior variação entre o percentual de alunos entre os 10% melhores e os 10% piores.

O desempenho dos alunos que têm histórico de reprovação e abandono vai ao encontro das preocupações levantadas por Menezes-Filho, Vasconcelos e Werlang (2005) em sua análise da progressão continuada nas escolas brasileiras.

Observemos agora as estatísticas descritivas referentes às escolas:

Tabela 3 - Estatísticas Descritivas Referentes às Escolas

		MATEMÁTICA				PORTUGUÊS			
		% dos centis				% nos centis			
		média	10	45-55	90	média	10	45-55	90
Rede	Estadual	188,42	21,63	24,01	23,25	185,08	21,38	23,09	26,51
	Municipal	187,00	78,37	75,99	76,75	181,00	78,62	76,91	73,49
Ciclos	sim	186,46	89,20	82,65	79,47	180,76	89,05	81,97	81,14
	não	194,70	10,80	17,35	20,53	187,84	10,95	18,03	18,86
Número de ciclos e duração dos dois primeiros ciclos	0	194,70	10,80	17,35	20,53	187,84	10,95	18,03	18,86
	2, 2 e 2	198,80	2,83	4,53	7,18	190,88	2,99	4,83	6,17
	2, 4 e 4	184,09	73,94	62,42	54,04	178,78	73,67	61,25	57,97
	3, 2 e 2	194,89	3,34	4,14	5,91	189,14	3,25	4,33	5,66
	4, 2 e 2	192,55	5,52	7,86	8,71	185,65	5,78	7,97	8,07
	Outros	188,40	3,56	3,70	3,63	181,07	3,36	3,59	3,27
Noturno	sim	176,81	2,31	1,64	0,84	172,56	2,01	1,62	0,91
	não	188,04	97,69	98,36	99,16	182,10	97,99	98,38	99,09
9 anos	sim	191,12	1,42	2,43	2,15	183,66	1,41	2,54	1,97
	não	187,79	98,58	97,57	97,85	181,91	98,59	97,46	98,03
Fundamental Completo	sim	180,00	48,18	35,20	24,56	174,16	48,31	34,25	26,83
	não	192,15	51,62	64,80	75,44	186,20	51,69	65,75	73,17
Pré-Escola	sim	191,34	10,26	13,16	14,25	184,88	10,38	13,72	13,55
	não	187,34	89,74	86,84	85,75	181,51	89,62	86,28	86,45
Ensino Médio	sim	183,04	7,21	6,55	4,88	178,98	7,24	6,44	5,58
	não	188,20	92,79	93,45	95,12	182,15	92,76	93,56	94,42
Supletivo	sim	183,55	55,12	45,58	38,28	177,20	55,22	45,88	38,52
	não	191,56	44,88	54,42	61,72	186,01	44,78	54,12	61,48
Tamanho em alunos	até 200	191,03	1,32	1,66	1,77	181,88	1,43	1,60	1,43
	de 200 a 500	194,22	15,59	21,22	27,17	187,32	15,57	21,69	24,76
	de 500 a 1000	191,85	33,38	40,60	47,25	186,14	33,71	41,17	46,68
	de 1000 a 1500	183,79	22,01	19,32	15,19	178,51	21,85	18,90	16,65
	de 1500 a 2000	176,78	16,16	11,19	5,95	171,61	16,25	10,52	7,26
	acima de 2000	171,96	11,55	6,01	2,67	165,63	11,20	6,12	3,23
Alunos por Sala*	até 20	195,29	8,19	12,02	15,45	187,70	8,48	12,09	13,73
	de 20 a 30	194,70	23,42	31,86	42,30	188,24	23,71	32,55	39,64
	de 30 a 40	189,39	22,20	24,67	25,38	184,55	21,83	25,14	26,66
	mais de 40	176,72	46,19	31,45	16,88	171,16	45,98	30,22	19,97

*alunos por sala refere-se ao número de alunos total da escola dividido pelo número de salas e dividido por dois. Supomos que existem dois turnos, matutino e vespertino.

Os dados indicam que a composição das séries na escola pode ser muito importante para o desempenho dos alunos. Escolas que possuem turmas de alunos depois da 4ª série juntamente com essas crianças têm médias mais baixas. Chama a atenção, além das médias

da notas dos alunos, o fato de 48% dos alunos no primeiro decil estarem em escolas com o fundamental completo e esse percentual cair para 24,56% no último decil.

Da mesma forma, parece que os alunos se beneficiam da presença de ensino infantil na mesma escola. A presença do ensino supletivo pode estar relacionada com a renda média da região em que a escola está instalada, que por sua vez está correlacionada com questões de infra-estrutura da escola.

Essa tabela também chama a atenção para os programas de progressão continuada. Uma análise apressada poderia nos levar a concluir que em média as escolas que adotaram o programa têm desempenho pior que as de regime seriado, entretanto, quando observamos também o formato de ciclos adotado, vemos que diferentes formatos têm impactos bastante distintos. Os regimes com ciclos de 2 anos apresentam médias muito melhores que as dos demais ciclos e, inclusive, que até do que o regime seriado.

Os alunos das escolas que adotaram o ensino fundamental de nove anos também apresentam resultado superior. Esse dado, juntamente com o que foi observado sobre os ciclos, coloca em questão a necessidade de uma análise mais profunda dos regimes e formatos adotados pelas escolas.

Diferentemente do que foi encontrado por Rios-Neto, César e Riani (2003) o fato da escola pertencer à rede municipal ou estadual parece não ser muito significativo na nossa amostra.

A variável alunos por sala é bastante utilizada na literatura e tem sido tema de bastante controvérsia. Infelizmente, não dispomos do número de alunos por turma no Censo Escolar e como nossa base de dados da Prova Brasil é amostral, não dispomos de meios de obter o número exato de alunos nas turmas de 4ª série. Entretanto, a divisão do número de alunos do período diurno por dois sub-períodos, matutino e vespertino, e pelo número de salas de aula parece uma aproximação satisfatória.

As notas médias de português e matemática são superiores no grupo de alunos cujas escolas tem até 30 alunos por sala. O resultado discorda de Hanushek (2003) e vai na direção de Krueger (1999). Além da do número de alunos por sala, o tamanho da escola também parece ser importante. Escolas maiores têm desempenho inferior.

Tabela 4 - Estatísticas Descritivas Referente ao Uso de Tecnologia nas Escolas

		MATEMÁTICA					PORTUGUÊS			
		% dos centis					% nos centis			
TVs por 1000 alunos	0	180,68	2,58	1,70	1,43	173,71	2,54	1,82	1,35	
	5	186,17	85,42	81,79	73,96	180,48	85,17	81,08	75,94	
	10	196,53	12,00	16,51	24,61	189,72	12,29	17,10	22,71	
Vídeos por 1000 alunos	0	178,63	8,01	5,82	3,49	173,13	7,69	5,61	3,87	
	5	187,30	84,49	83,51	80,62	181,48	84,38	83,37	81,49	
	10	196,83	7,50	10,67	15,89	189,98	7,93	11,02	14,65	
Parabólica para tv	não	182,66	43,94	36,62	28,01	176,80	43,81	35,74	29,70	
	analógica	190,07	41,21	45,53	49,48	184,08	41,20	45,95	48,63	
	digital	192,74	14,84	17,85	22,51	186,87	14,99	18,31	21,67	
Parabólica para internet	sim	193,12	3,88	4,75	6,04	185,27	4,03	4,71	5,20	
	não	187,60	96,12	95,25	93,96	181,79	95,97	95,29	94,80	
Retroprojektor	não	190,79	60,40	72,03	78,96	184,95	60,26	72,81	77,30	
	retroproj.	192,02	6,85	7,64	9,36	186,39	7,16	7,80	9,41	
	data show	176,22	32,75	20,33	11,68	169,95	32,58	19,39	13,29	
Impressoras por 100 professores	até 20	187,14	94,74	93,47	91,02	181,21	94,81	93,33	91,36	
	de 20 a 30	193,38	4,88	5,90	7,86	187,66	4,82	5,93	7,47	
	mais de 30	200,07	0,39	0,63	1,12	195,70	0,37	0,74	1,17	
Xerox	sim	189,60	45,46	47,15	51,92	183,66	45,57	47,36	51,30	
	não	186,28	54,54	52,85	48,08	180,39	54,43	52,64	48,70	
Ar- condicionado	sim	189,72	9,63	10,44	11,38	183,40	10,00	10,67	11,28	
	não	187,65	90,37	89,56	88,62	181,78	90,00	89,33	88,72	
Aparelho de som	sim	191,04	15,22	17,23	19,77	185,51	15,39	17,45	19,56	
	não	187,20	84,78	82,77	80,23	181,20	84,61	82,55	80,44	
Mimeógrafo por 100 professores	até 10	184,87	71,34	63,49	55,53	178,93	71,23	62,83	56,59	
	de 10 a 25	192,47	23,66	30,00	35,34	186,47	23,87	30,19	34,36	
	mais de 25	195,33	4,99	6,52	9,13	190,10	4,90	6,97	9,05	
Máquina fotográfica	sim	188,37	80,83	81,56	84,11	182,42	81,07	81,68	83,77	
	não	185,60	19,17	18,44	15,89	179,84	18,93	18,32	16,23	
Filmadora	sim	190,28	50,74	42,48	38,65	184,30	50,71	42,32	40,01	
	não	184,69	49,26	57,52	61,35	178,87	49,29	57,68	59,99	
Computadores por 1000 alunos	0	182,49	1,56	1,30	0,99	175,82	1,50	1,44	0,98	
	até 10	188,84	46,51	50,65	51,35	183,26	46,07	50,98	51,46	
	de 10 a 20	182,74	29,59	23,02	18,78	176,45	29,54	22,67	19,39	
	de 20 a 40	189,75	19,44	20,87	22,91	183,60	20,04	20,81	22,70	
	mais de 40	197,02	2,90	4,17	5,97	190,58	2,85	4,09	5,48	
Conexão com Internet	não	191,56	3,63	5,16	5,43	185,24	3,73	5,46	5,12	
	64kps	190,65	35,86	42,08	45,99	184,29	36,01	43,14	43,84	
	64 a 128kps	187,52	24,13	23,73	23,19	182,04	24,07	23,15	23,86	
	129 a 512kps	183,00	30,08	23,22	19,70	177,57	29,67	22,79	21,32	
	512kps	186,90	6,29	5,82	5,69	180,99	6,51	5,47	5,86	
Uso pela comunidade	sim	189,76	6,65	8,01	7,93	184,61	6,81	7,74	8,41	
	não	187,71	93,35	91,99	92,07	181,73	93,19	92,26	91,59	
Uso livre dos alunos	sim	182,92	4,02	3,30	2,58	177,02	3,91	3,28	2,72	
	não	188,03	95,98	96,70	97,42	182,12	96,09	96,72	97,28	
Laboratório de Informática	não	189,73	31,53	35,59	37,80	183,53	31,33	36,32	36,58	
	sem aula	186,43	66,06	61,22	58,00	180,61	66,33	60,28	59,09	
	com aula	193,48	1,73	2,13	2,69	188,96	1,71	2,25	2,90	
	aula sem lab.	197,18	0,68	1,06	1,51	192,28	0,63	1,15	1,43	

Os dados acima se referem a equipamentos que o senso comum associa a boas práticas educacionais como computadores, televisões e acesso à internet. A primeira coisa que nos chama a atenção é a grande variação do número de equipamentos por aluno na rede pública, o que é curioso, já que era de se esperar uma maior padronização na rede pública de ensino.

Outro aspecto importante que notamos quando observamos o número de computadores, televisões e vídeos por mil alunos é que apesar destes estarem sempre correlacionados com melhor desempenho, a partir de um certo número este impacto se torna muito mais perceptível, talvez indicando que existe a necessidade de um número adequado desses equipamentos para tenham o efeito desejado.

Isso pode ser observado nas médias de matemática quando observamos, por exemplo, o número de computadores e televisões por aluno. O aumento do número de computadores por mil alunos de zero para 40 está associado a uma melhora no desempenho de aproximadamente 7 pontos na média, já o aumento de 40 para mais de 40, está associado a um aumento de 8 pontos na média.

A presença de computadores em sala de aula vai na direção do que foi encontrado por Barros et alli (2001). O maior número de computadores está relacionado com um melhor desempenho dos alunos. Entretanto a presença de conexão com internet apresenta um resultado curioso: melhor acesso à rede parece contribuir para um pior desempenho dos alunos.

A presença de laboratórios de informática também apresenta um resultado interessante. As escolas que oferecem aula de informática sem a presença de laboratório possuem o melhor desempenho, entretanto o tamanho da amostra dessas escolas é bastante pequeno. O desempenho dos alunos que possuem laboratório de informática só é superior aos que não possuem para o caso em que a presença do laboratório é acompanhada pela existência de aulas, o que parece fazer sentido em termos de eficiência no uso de recursos escolares.

De maneira geral, as médias estão na linha do que se encontra em Rios-Neto, César e Riani (2003) referente à presença de laboratórios de informática nas escolas e chama atenção para a necessidade de se compreender melhor o uso desses recursos.

A exploração dos resultados associados a recursos escolares tem que ser feita com muita cautela. È difícil chegar a conclusões simplesmente através das médias. A presença e a maior intensidade de certos recursos tecnológicos podem estar correlacionadas com outras melhoras na infra-estrutura escolar ou ainda com diferentes projetos pedagógicos de difícil mensuração.

Vejam os abaixo as médias associadas a mais algumas características das escolas:

Tabela 5 - Estatísticas Descritivas Referentes às Escolas 2

		MATEMÁTICA				PORTUGUÊS			
		% dos centis							
		média	10	45-55	90	média	10	45-55	90
Videoteca	sim	191,51	4,78	5,16	6,23	184,65	4,83	5,33	6,08
	não	187,66	95,22	94,84	93,77	181,80	95,17	94,67	93,92
Laboratório	sim	184,62	14,35	11,54	10,53	178,97	14,55	11,32	11,29
Ciências	não	188,30	85,65	88,46	89,47	182,35	85,45	88,68	88,71
Outro Laboratório	sim	190,42	4,21	3,93	4,86	185,24	4,21	4,11	5,25
	não	187,75	95,79	96,07	95,14	181,81	95,79	95,89	94,75
Alimentação	Outro	189,13	0,78	1,00	0,92	182,05	0,78	1,20	0,85
	cozinha	186,41	79,32	74,57	70,32	180,26	79,50	74,09	70,11
	cozinha e cantina	192,37	19,46	24,09	28,47	187,21	19,25	24,36	28,71
	cantina	181,82	0,45	0,35	0,30	177,11	0,47	0,36	0,33
Parque Infantil	sim	191,53	10,32	12,25	14,02	181,50	10,48	12,82	13,41
	não	187,34	89,68	87,75	85,98	185,13	89,52	87,18	86,59
Quadra Esportiva	não	188,27	15,69	17,26	16,36	182,46	15,56	17,37	16,44
	descob.	187,01	55,44	52,99	51,34	181,44	55,52	51,97	52,92
	coberta	188,73	27,56	27,87	29,80	182,21	27,55	28,79	28,38
	ginásio	195,37	1,31	1,88	2,51	187,86	1,38	1,87	2,25
Piscina	sim	189,71	1,65	1,53	1,87	184,67	1,71	1,41	2,07
	não	187,84	98,35	98,47	98,13	181,91	98,29	98,59	97,93
Dependências Especiais	não	188,44	57,24	59,95	60,44	182,46	57,11	60,48	59,36
	sim	188,23	13,20	14,26	13,59	182,26	13,27	14,00	13,93
	salas	186,39	29,55	25,79	25,97	180,64	29,62	25,52	26,70
Brinquedoteca	sim	190,86	6,66	8,11	8,68	186,25	6,61	7,88	9,03
	não	187,61	93,34	91,89	91,32	181,59	93,39	92,12	90,97
Auditorio	sim	191,20	3,97	4,65	5,21	184,63	4,17	4,62	4,95
	não	187,70	96,03	95,35	94,79	181,82	95,83	95,38	95,05
Biblioteca	sim	190,63	14,83	16,76	19,01	184,24	15,00	17,29	18,08
	não	187,30	85,17	83,24	80,99	181,48	85,00	82,71	81,92
Sala de Leitura	sim	186,70	65,03	59,97	58,52	180,88	65,15	59,49	59,55
	não	189,66	34,97	40,03	41,48	183,59	34,85	40,51	40,45
Ventiladores	não	180,02	20,91	16,09	10,46	174,88	20,96	15,63	11,85
	até 2	188,53	70,73	72,78	74,81	182,63	70,20	73,05	74,35
	mais que 2	194,61	8,36	11,13	14,73	187,49	8,85	11,32	13,80
Reciclagem	sim	193,85	8,49	11,69	14,09	188,85	8,41	11,42	14,38
	não	187,09	91,51	88,31	85,91	181,05	91,59	88,58	85,62

Essa tabela também ressalta a importância de se pensar como estão sendo usados os recursos pelas escolas e sua real eficácia e eficiência. As escolas que não possuem laboratórios de ciências apresentam médias maiores do que aquelas que possuem, o que vai

contra o que consta em Rios-Neto, César e Riani (2003), e as escolas que possuem quadras descobertas têm médias piores a aquelas que não possuem quadra alguma. O impacto positivo das quadras parece que só começa a ocorrer a partir de um certo nível de qualidade, como observamos na média para quadras cobertas, e passa a ser realmente efetiva quando observamos as escolas que possuem ginásios poliesportivos ou piscinas.

A presença de bibliotecas tem um resultado positivo, mas a presença de salas de leitura não. O resultado negativo para essas últimas talvez seja um caso de mal uso de recursos, ou ainda de uma substituição inadequada de recursos, já que algumas escolas que possuem salas de leitura não possuem bibliotecas.

É importante notar que mais de 15% das escolas públicas paulistas não possuem ventiladores. A diferença entre as escolas que não possuem ventiladores e as que possuem mais de 2 chega a impressionantes 14 pontos na média. Pode ser que isso se deva ao fato de que a presença de ventiladores em salas de aula é uma condição básica de conforto para que os alunos possam ter atenção às aulas.

A presença de auditório, parque infantil e reciclagem também indicam aspectos interessantes. As médias para as escolas que possuem esses itens são positivas. Por outro lado, a presença de dependências especiais na escola deve ser melhor examinada.

As médias que encontramos para a presença de diferentes recursos apontam na mesma direção do que foi colocado por Diaz (2007), Felício (2004) e Albernaz, Ferreira e Franco (2002) que se referem ao impacto significativo que as escolas têm sobre o desempenho dos alunos. E vai contra o ponto levantado por Coleman (1966) e seguido por Hanushek (2003) de que a diferença de atributos entre as escolas era irrelevante para explicar a diferença de desempenho entre os alunos.

Outro aspecto bastante estudado pela literatura e que aqui exploramos por meio dos dados do Censo Escolar é o impacto dos professores sobre o desempenho dos alunos. Na tabela abaixo seguem algumas estatísticas a esse respeito:

Tabela 6 - Estatísticas Descritivas Referentes aos Professores

		MATEMÁTICA				PORTUGUÊS			
		% dos centis							
		média	10	45-55	90	média	10	45-55	90
Escolaridade	até 14 anos	186,97	15,89	16,40	14,87	181,18	15,95	16,44	14,96
Média dos	de 14 a 15 anos	188,29	36,02	35,77	37,17	182,76	35,35	35,87	37,45
Professores	de 15 a 16 anos	187,57	45,82	44,11	44,56	181,41	46,41	43,93	44,42
	de 16 a 17 anos	192,14	0,85	1,29	1,25	185,10	0,88	1,22	1,20
	Não declarado	191,12	1,42	2,43	2,15	183,66	1,41	2,54	1,97
% Professores com Ensino Superior	até 50%	187,69	8,45	8,81	8,45	181,46	8,45	8,93	8,17
	de 50 a 75%	188,08	21,73	23,08	22,66	182,85	21,60	22,98	23,00
Superior	de 75% a 90%	187,62	32,55	30,47	31,49	181,91	32,18	30,52	31,80
	mais de 90%	187,78	35,85	35,20	35,26	181,41	36,36	35,04	35,06
	Não declarado	191,12	1,42	2,43	2,15	183,66	1,41	2,54	1,97
O professor(a) corrige a lição	Sim	189,30	72,16	75,90	81,07	183,35	70,54	74,75	78,47
	Não	183,27	27,84	24,10	18,93	177,83	29,46	25,25	21,53
Alunos por professor no EF	até 20	192,56	11,73	13,95	17,22	185,49	11,94	14,46	15,57
	de 20 a 30	187,49	68,95	68,28	66,47	181,72	68,91	67,54	68,03
	mais de 30	185,49	19,32	17,77	16,31	180,04	19,16	18,00	16,40

A variável escolaridade média dos professores refere-se aos anos de estudo dos professores de ensino fundamental da escola do aluno, não há dados dos professores referentes às turmas no nosso banco de dados. O número de anos de estudo dos professores foi calculado considerando-se 7 anos para o ensino fundamental incompleto, 8 anos para o ensino fundamental completo, 11 anos para o ensino médio completo, 12 anos para o magistério, 15 anos para o ensino superior, 17 anos para o ensino superior com magistério e 16 anos para o ensino superior com licenciatura.

Mais uma vez as médias divergem dos resultados de Hanushek (2003) e parecem ir na direção dos resultados encontrados em Albernaz, Ferreira e Franco (2002). Percebe-se uma associação entre níveis mais altos de escolaridade média dos professores e melhor desempenho.

Ao contrário do resultado encontrado por Rios-Neto, César e Riani, a variável do percentual de professores com ensino superior parece não influenciar o desempenho dos alunos. Já correção de exercícios propostos por parte dos professores parece ser relevante para explicar alunos com pior e melhor desempenho. Talvez esta variável esteja associada com práticas mais eficientes ou eficazes adotadas em sala, ou ainda com aspectos institucionais da relação professor-aluno. Filgio e Kenny (2006) e Oreopolus (2006) levantaram a questão da importância desses aspectos.

Tabela 7 - Estatísticas Descritivas Referentes aos Colegas

		MATEMÁTICA				PORTUGUÊS			
		% dos centis				% nos centis			
		média	10	45-55	90	média	10	45-55	90
% Alunos Negros ou pardos por sala	até 20%	200,21	7,48	10,64	19,12	194,55	7,31	10,74	18,44
	de 20 a 35%	191,92	30,95	35,71	42,66	186,01	31,53	36,27	42,15
	de 35 a 50%	183,16	43,24	38,34	29,13	177,21	43,09	37,77	29,80
	acima de 50%	178,02	16,91	12,88	6,93	172,26	16,67	12,68	7,65
	Não declarado	191,12	1,42	2,43	2,15	183,66	1,41	2,54	1,97
% Alunos que abandonaram o curso*	Até 1%	188,80	83,86	86,19	90,32	182,86	83,78	86,17	89,85
	de 2 a 5%	182,04	14,17	12,30	8,72	176,17	14,33	12,22	9,09
	mais de 5%	179,82	1,97	1,52	0,96	174,70	1,89	1,61	1,06
% Alunos repetentes*	Nenhum	189,63	13,03	13,63	15,13	183,97	12,81	13,46	15,12
	de 1 a 10%	191,15	35,35	40,05	45,71	185,58	35,39	40,47	45,78
	de 10 a 15%	187,32	11,25	11,28	10,87	181,04	11,42	11,14	10,70
	mais de 15%	183,49	40,37	35,05	28,29	177,19	40,38	34,93	28,40
% Alunos reprovados*	Nenhum	187,67	15,66	13,42	15,06	181,65	15,24	13,68	14,83
	de 1 a 5%	193,04	20,77	25,88	31,86	187,81	20,51	26,19	31,84
	de 5 a 15%	188,41	35,62	38,69	37,81	182,53	35,95	38,53	37,72
	mais de 15%	180,79	27,96	22,01	15,27	174,07	28,30	21,60	15,61
% Alunos com idade acima	Nenhum	198,30	1,69	2,18	3,56	193,32	1,50	2,13	3,39
	até 5%	194,32	21,35	26,68	36,23	189,54	21,01	26,67	36,81
	de 5 a 10%	188,10	32,95	34,72	33,82	182,28	33,37	34,65	34,11
	de 10 a 15%	181,26	42,59	33,98	24,24	174,21	42,71	34,02	23,72
	Não declarado	191,12	1,42	2,43	2,15	183,66	1,41	2,54	1,97

* número no ano anterior (2004) sobre matrículas em 2005.

Os dados acima mostram uma relação importante entre o percentual de negros ou pardos na série e o desempenho dos alunos. Escolas com alto percentual desses alunos têm um desempenho médio significativamente menor. Contudo, resta examinar se esse impacto é o mesmo do que é verificado para o *background* de cada aluno ou se trata de um aspecto distinto.

O percentual de alunos com até 20% de negros ou pardos na série entre os 10% piores é em torno de 7% e salta para aproximadamente 20% entre os 10% melhores. Já o percentual de alunos com mais de 50% de negros na série entre os 10% piores é cerca de 17% e cai para 7%. Estas variações são muito expressivas e parecem indicar que existem dois perfis muito distintos de escolas públicas paulistas, uma escola de “negros” e outra de “brancos”, com desempenhos significativamente diferentes.

Também o percentual de alunos com idade acima parece ser muito importante no desempenho médio. Esta variável é particularmente importante para as discussões em torno dos programas de progressão continuada. O percentual de repetência e reprovação também se mostraram bastante significativos, entretanto, essas variáveis são de difícil interpretação, já que existem diversos regimes adotados pelas escolas o que torna difícil comparar essas taxas.

O cenário colocado pela tabela acima vai mais uma vez ao encontro com o Albernaz, Ferreira e Franco (2002) indicando como esses autores que o *background* socioeconômico dos colegas também é um importante determinante de desempenho escolar.

Por fim, podemos vejamos abaixo a tabela com as estatísticas referentes a diferentes programas adotados nas escolas:

Tabela 8 - Estatísticas Descritivas Referentes a Programas Educacionais

		MATEMÁTICA				PORTUGUÊS			
		% dos centis							
		média	10	45-55	90	média	10	45-55	90
Renda mínima	sim	190,36	64,88	73,92	81,25	184,19	65,17	75,03	78,53
	não	180,82	35,12	26,08	18,75	175,63	34,83	24,97	21,47
TV Escola	sim	192,97	32,75	40,08	50,08	187,02	32,90	40,55	48,70
	não	184,37	67,25	59,92	49,92	178,47	67,10	59,45	51,30
Programa TVE	sim	194,87	8,66	11,13	15,28	188,69	8,94	11,34	14,62
	não	186,95	91,34	88,87	84,72	181,07	91,06	88,66	85,38
Proinfo	sim	189,19	1,30	1,39	1,50	184,76	1,32	1,35	1,54
	não	187,85	98,70	98,61	98,50	181,91	98,68	98,65	98,46
Outro Programa Estadual	sim	190,54	26,62	31,63	34,29	185,60	26,88	31,84	35,59
	não	186,64	73,38	68,37	65,71	180,28	73,12	68,16	64,41
Outro Programa Municipal	sim	187,58	63,65	61,16	61,72	181,24	63,80	61,04	60,71
	não	188,32	36,35	38,84	38,28	183,09	36,20	38,96	39,29
Projeto Saúde Reprodutiva	sim	188,27	6,99	6,61	7,23	182,71	6,97	6,28	7,47
	não	187,84	93,01	93,39	92,77	181,90	93,03	93,72	92,53
Outro Programa Federal	sim	188,87	23,95	25,58	26,17	183,32	23,61	25,71	26,64
	não	187,52	76,05	74,42	73,83	181,48	76,39	74,29	73,36
Transporte de Alunos	não	186,89	36,71	38,08	33,87	182,16	36,08	37,62	36,53
	passe	192,44	8,25	11,03	12,90	186,80	8,48	11,13	12,87
	frota	184,53	39,58	31,82	30,07	177,38	40,04	31,84	28,61
	veículo locado	192,56	15,47	19,07	23,16	186,35	15,4	19,41	21,99

De maneira geral, a participação em diferentes programas parece influenciar positivamente o desempenho dos alunos. As médias que mais chamam a atenção são as da variável Renda Mínima²⁸, TV Escola e Outros Programas Municipais. A primeira variável é interessante por se tratar de um programa destinado a crianças de baixa renda. Já a segunda, está associada a uma variação bastante expressiva no desempenho dos alunos, contudo, seria interessante verificar se este impacto se deve ao fato do programa já estar selecionando as melhores escolas. A variável Outro Programa Municipal, por sua vez, indica que há um

²⁸ Renda Mínima refere-se à participação da escola em algum programa de transferência de renda condicionada municipal estadual ou federal.

número grande de programas municipais desenvolvidos nas escolas, mas seus resultados são duvidosos.

Os outros programas da tabela acima também parecem estar associados a um aumento de desempenho dos alunos. Não podemos dizer, porém, se esses resultados são significativos. A existência de transporte escolar apresenta resultados interessantes, já que frota própria tem uma média muito diferente do uso de passe e veículos locados, sendo que a única distinção a priori que com essas formas de transporte escolar é o fato de ser gerenciada pela própria escola. Talvez esta variável esteja relacionada a formas mais ou menos eficientes de gestão de programas.

Foram apresentadas até aqui estatísticas descritivas das variáveis que foram analisadas neste trabalho. O número de variáveis utilizadas é bastante grande e não pretendemos realizar um exame minucioso de cada uma delas. Desejamos apenas contribuir com a literatura com um panorama geral de possíveis determinantes de desempenho escolar.

Na próxima seção serão apresentados os resultados obtidos através do método de decomposição de desigualdade com o índice Theil-L.

5. Resultados

Nesta seção fazemos a apresentação dos resultados das decomposições feitas por meio do índice Theil-1. Primeiramente, serão apresentados os resultados das contribuições, contribuições marginais e uma regressão por mínimos quadrados para termos um parâmetro para os resultados encontrados. As contribuições brutas foram calculadas para as variáveis que se mostraram mais interessantes na análise descritiva e as contribuições marginais apenas com aquelas cujas contribuições brutas se mostraram mais significativas.

Vejam agora a contribuição bruta das variáveis referente ao *background* socioeconômico dos alunos, às escolas e aos colegas. A disposição dos resultados procura seguir o mesmo padrão apresentado na análise descritiva:

5.1 Contribuições Brutas

As contribuições brutas consistem em analisar quanto que uma determinada variável explica da desigualdade. Neste cálculo utilizamos a propriedade de decomponibilidade do índice Theil-L, como foi colocado na seção 3. O índice nada mais é do que um somatório da participação da desigualdade para cada indivíduo na desigualdade total.

$$L = \sum_{i=1}^n \frac{1}{n} \log \frac{1/n}{y_i}, \text{ em que } y_i \text{ é a proporção que cada indivíduo tem do somatório de}$$

pontos e $1/n$ sua proporção na amostra. Se dividirmos a amostra em dois grupos, por exemplo, brancos e não brancos, podemos recalculá-lo utilizando a proporção dos grupos na amostra e a proporção destes do total das notas, fazendo:

$$L_e = \sum_{h=1}^k \pi_h \log \frac{\pi_h}{y_k}, \text{ sendo } \pi_h \text{ a proporção do grupo na amostra e } k = 2, \text{ brancos e não}$$

brancos. L_e é apenas uma agregação dos termos de L . Se ao dividirmos a amostra entre brancos e não brancos, calcularmos a desigualdade L_e e obtermos uma desigualdade igual a L , dizemos que a variável cor explica 100% da desigualdade total. É disso que se trata a contribuição bruta, o valor da desigualdade que temos ao dividir a amostra segundo os valores de uma variável. Como L_e não pode ser maior que L , se L_e é menor que L , ainda resta uma desigualdade que não é explicada pela divisão da amostra em grupos segundo os valores dessa variável. Essa desigualdade que não é explicada pela divisão em grupos, deve-se a desigualdade dentro dos grupos, chamada desigualdade intra-grupos na seção 3..

Tabela 9 - Contribuições Brutas das Variáveis de Background dos Alunos

	Desigualdade entre Grupos - Theil-L		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	100,00%	100,00%
Sexo	0,00000	0,00000	0,00%	0,00%
Cor	0,00011	0,00003	0,41%	0,09%
Mora com os Pais	0,00002	0,00002	0,07%	0,06%
Acima de 11 anos	0,00003	0,00004	0,11%	0,13%
Mãe com Ensino Superior	0,00006	0,00009	0,22%	0,28%
Pai com Ensino Superior	0,00006	0,00008	0,22%	0,25%
Silêncio para Estudar em Casa	0,00004	0,00004	0,15%	0,13%
Trabalhos Domésticos	0,00002	0,00002	0,07%	0,06%
Trabalho Infantil	0,00003	0,00003	0,11%	0,09%
Deseja Continuar Somente a Estudar	0,00002	0,00002	0,07%	0,06%
Pessoas por Quarto	0,0002	0,00021	0,74%	0,66%
Possui Carro?	0,00007	0,00009	0,26%	0,28%
Possui Computador	0,00005	0,00008	0,18%	0,25%
Apoio dos Pais	0,00002	0,00002	0,07%	0,06%
Apoio para Fazer a Lição de Casa	0,00003	0,00003	0,11%	0,09%
Frequentou Ensino Infantil	0,00013	0,00014	0,48%	0,44%
Já foi Reprovado	0,00004	0,00007	0,15%	0,22%
Já Abandonou a Escola	0,00005	0,00006	0,18%	0,19%
Faz Tarefa	0,00006	0,00009	0,22%	0,28%
Recebe Bolsa Família	0,00007	0,00009	0,26%	0,28%

O valor do índice Theil-L para a desigualdade total entre os alunos é de 0,02704 para as notas de matemática e 0,03175 para o desempenho em português. Estas medidas são a desigualdade total entre os alunos. Felício (2004) também calculou índices Theil-L para proficiência em matemática e português para alunos da 4ª série do Ensino Fundamental do Estado de São Paulo, encontrando os valores de 0,0331 e 0,0419. Os números da autora são 22% e 31% maiores para matemática e português. Assim como neste trabalho a desigualdade em português é maior. A diferença entre os número encontrados provavelmente se deve ao fato da autora utilizar também escolas particulares em sua amostra de dados do SAEB.

Chama a atenção na tabela acima o valor baixo das contribuições brutas. Nenhuma das variáveis de *background* socioeconômico consegue explicar sequer 1% da desigualdade total entre os alunos. Dentre essas variáveis, podemos destacar as seguintes contribuições que se apresentaram ligeiramente maior: o número de pessoas por quarto, a cor do aluno e se este frequentou o ensino fundamental.

Mesmo que a relação L_e/L definida acima para as contribuições brutas não chegue a representar muito percentualmente, essas variáveis devem ser ainda observadas levando em conta as demais variáveis que dispomos, pois pode ser que condicionadas a outros fatores tenham seu impacto aumentado.

Há algumas diferenças interessantes entre as desigualdades para matemática e português. Primeiro, a desigualdade entre as notas de português é maior. Além disso, as variáveis explicam um pouco mais da desigualdade em matemática do que em português. O valor da contribuição da cor do aluno para o desempenho em português é três vezes menor que o valor para matemática.

As variáveis de escolaridade do pai e da mãe também não se destacam dentre as outras. Mais uma vez, é provável que esse resultado se deva ao fato dos pais com maior escolaridade estarem na rede pública, o que diminui a variação e a significância dessa variável na amostra deste trabalho. Com esta ressalva, nossos resultados não apontam o mesmo que Albernaz, Ferreira e Franco (2002), Rios-Neto, Cezar e Riani (2003), Barros *et alli* (2005) para estas variáveis.

Observemos agora os resultados para as contribuições brutas das variáveis referentes às escolas:

Tabela 10 - Contribuições Brutas das Variáveis de Infraestrutura Escolar

	Desigualdade entre Grupos - Theil-L		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	100,00%	100,00%
Departamento Administrativo	0,00000	0,00007	0,00%	0,22%
Ciclos	0,00020	0,00010	0,72%	0,33%
Número de Ciclos e Duração dos dois Primeiros Ciclos	0,00022	0,00013	0,81%	0,41%
Noturno	0,00002	0,00002	0,07%	0,06%
9 anos	0,00000	0,00000	0,00%	0,00%
Fundamental Completo	0,00040	0,00043	1,48%	1,35%
Pré-Escola	0,00002	0,00001	0,07%	0,03%
Ensino Médio	0,00007	0,00004	0,26%	0,13%
Supletivo	0,00019	0,00028	0,70%	0,88%
Tamanho em Alunos	0,00057	0,00055	2,11%	1,73%
Alunos por Sala	0,00072	0,00069	2,66%	2,17%
Videoteca	0,00001	0,00001	0,04%	0,03%
Laboratório de Ciências	0,00003	0,00001	0,11%	0,03%
Outro Laboratório	0,00001	0,00001	0,04%	0,03%
Parque Infantil	0,00002	0,00001	0,07%	0,03%
Quadra Esportiva	0,00001	0,00001	0,04%	0,03%
Piscina	0,00000	0,00000	0,00%	0,00%
Dependências Especiais	0,00000	0,00000	0,00%	0,00%
Brinquedoteca	0,00002	0,00003	0,07%	0,09%
Auditório	0,00001	0,00003	0,04%	0,09%
Biblioteca	0,00001	0,00001	0,04%	0,03%
Sala de Leitura	0,00000	0,00000	0,00%	0,00%
Ventiladores	0,00021	0,00016	0,78%	0,50%
Reciclagem	0,00004	0,00006	0,15%	0,19%

Mais uma vez, as contribuições brutas parecem explicar relativamente pouco da desigualdade de notas entre os alunos. Como na tabela anterior, as contribuições brutas são na

grande maioria menores para as notas de português. Apesar do valor pequeno, é interessante notar que a contribuição bruta da rede é zero para matemática, mas corresponde a 0,22% da desigualdade em português.

A contribuição bruta da presença e do número de ciclos representa para a desigualdade em português a metade do que representa para matemática. Em termos de diferenças entre as duas matérias, também se nota o valor ligeiramente superior da contribuição da variável supletivo para português.

O tamanho dos alunos e das escolas, o número de duração dos ciclos, a presença de alunos de 5ª a 8ª série e o número de ventiladores por sala de aula são as variáveis que apresentam contribuições mais altas. Todas essas contribuições têm valores superiores às contribuições das variáveis de background socioeconômico. E chegam a representar até cerca de 3% da desigualdade entre os alunos. O resultado das contribuições marginais para o tamanho das salas e das escolas reforçam o que foi observado nas médias das notas, indo ao encontro de Krueger (1999) e contra Hanushek (2003).

Dentre as variáveis de infra-estrutura física, a única que se destacou foi o número de ventiladores por sala. Esta mesma variável também foi uma das poucas encontradas como significantes por Felício (2004) em sua análise das características das melhores escolas do estado de São Paulo.

Abaixo, as contribuições brutas para variáveis de uso de tecnologia nas escolas:

Tabela 11 - Contribuições Brutas das Variáveis Referentes ao Uso de Tecnologia nas Escolas

	Desigualdade entre Grupos -		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	100,00%	100,00%
TVs por 1000 alunos	0,00018	0,00016	0,67%	0,50%
Vídeos por 1000 alunos	0,00018	0,00016	0,67%	0,50%
Parabólica para TV	0,00019	0,00023	0,70%	0,72%
Parabólica para Internet	0,00001	0,00000	0,04%	0,00%
Retroprojektor	0,00004	0,00005	0,15%	0,16%
Impressoras por 100 professores	0,00004	0,00004	0,15%	0,13%
Xerox	0,00005	0,00005	0,18%	0,16%
Ar-condicionado	0,00000	0,00000	0,00%	0,00%
Aparelho de Som	0,00003	0,00003	0,11%	0,09%
Mimeógrafo por 100 professores	0,00019	0,00022	0,70%	0,69%
Máquina Fotográfica	0,00001	0,00001	0,04%	0,03%
Filmadora	0,00009	0,00009	0,33%	0,28%
Computadores por 1000 alunos	0,00009	0,00012	0,33%	0,38%
Conexão com Internet	0,00009	0,00007	0,33%	0,22%
Uso pela comunidade	0,00000	0,00000	0,00%	0,00%
Uso Livre dos alunos	0,00001	0,00001	0,04%	0,03%
Laboratório de Informática	0,00004	0,00006	0,15%	0,19%

O número de vídeos e tvs e a existência de antena parabólica na escola são as variáveis que se mostraram mais importantes. É interessante notar que a contribuição dessas variáveis está no mesmo nível das variáveis mais importantes de *background* dos alunos.

As variáveis referentes ao uso de informática na sala de aula apresentaram contribuições muito baixas. Talvez o contraste com os resultados positivos encontrados por Barros *et alli* (2001) e Rios-Neto, César e Riani (2003) estejam mais uma vez relacionados à ausência de escolas particulares na amostra.

É importante lembrar que a existência de computadores e outros tipos de tecnologia nas escolas desempenha um papel marginal no sistema de ensino brasileiro, portanto, é de se esperar que a variação intra-grupos dessas variáveis seja muito maior que a variação entre. Resta examinar melhor o uso desses equipamentos mais a frente com o cálculo das contribuições marginais.

Tabela 12 - Contribuições Brutas das Variáveis Referentes aos Professores

	Desigualdade entre Grupos - Theil-L		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	100,00%	100,00%
Escolaridade Média dos Professores	0,00002	0,00001	0,07%	0,03%
% Professores com Ensino Superior	0,00001	0,00001	0,04%	0,03%
O professor(a) Corrige a Lição	0,00010	0,00010	0,37%	0,31%
Alunos por professor no EF	0,00005	0,00004	0,18%	0,13%

Nenhuma das variáveis referentes aos professores apresentou uma contribuição bruta muito relevante. Este resultado vai na mesma direção apontada por Hanushek (2003) e a literatura que indica a irrelevância da variação dos insumos escolares no desempenho dos alunos.

Também para esses resultados podemos fazer a ressalva do impacto da ausência de escolas particulares, que estariam associadas a condições melhores para os professores.

Tabela 13 - Contribuições Brutas das Variáveis Referentes aos Colegas

	Desigualdade entre Grupos - Theil-L		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	100,00%	100,00%
Média de Pessoas por quarto	0,00087	0,00087	3,22%	2,73%
% Alunos Negros ou Pardos por sala	0,00064	0,00061	2,38%	1,93%
% Alunos que Abandonaram o Curso	0,00011	0,00011	0,41%	0,35%
% Alunos Repetentes	0,00012	0,00016	0,44%	0,50%
% Alunos Reprovados	0,00025	0,00034	0,92%	1,07%
% Alunos com Idade Acima	0,00041	0,00056	1,52%	1,76%

As variáveis de *background* dos colegas estão dentre aquelas que apresentaram contribuições brutas mais altas. Destaca-se a contribuição da presença de alunos negros ou pardos e de alunos com idade acima da ideal. Um resultado interessante é o impacto um pouco maior do percentual de alunos repetentes, reprovados e com idade acima na desigualdade de proficiência em português.

O valor das contribuições brutas encontradas para essas variáveis está de acordo com o esperado na análise descritiva e está na mesma direção dos resultados encontrados em Albernaz, Ferreira e Franco (2006) sobre a importância do *background* socioeconômico dos colegas de sala.

Tabela 14 - Contribuições Brutas das Variáveis Referentes a Programas Escolares

	Desigualdade entre Grupos - Theil-L		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	100,00%	100,00%
Renda mínima	0,00014	0,00009	0,52%	0,28%
TV Escola	0,00020	0,00024	0,74%	0,76%
Programa TVE	0,00005	0,00006	0,18%	0,19%
Proinfo	0,00000	0,00000	0,00%	0,00%
Outro Programa Estadual	0,00002	0,00008	0,07%	0,25%
Outro Programa Municipal	0,00000	0,00000	0,00%	0,00%
Projeto Saúde Reprodutiva	0,00000	0,00000	0,00%	0,00%
Outro Programa Federal	0,00000	0,00001	0,00%	0,03%
Transporte de Alunos	0,00010	0,00013	0,37%	0,41%

Programas como o TV Escola, Renda Mínima e Transporte Escolar também apresentam uma contribuição da mesma magnitude das variáveis observadas do *background* dos alunos. Os programas classificados genericamente como outros não contribuem significativamente para explicar a desigualdade de desempenho entre os alunos.

Uma última abordagem das contribuições brutas é feita ao olharmos para a contribuição que escolas, município e turmas têm na desigualdade total.

Tabela 15 - Contribuição Bruta da Escola, Turma e Município

	Desigualdade entre Grupos - Theil-L		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	100,00%	100,00%
Município	0,00141	0,00122	5,21%	3,84%
Escola	0,00394	0,00421	14,58%	13,27%
Turma	0,00570	0,00481	21,09%	15,15%

Observando o efeito dessas variáveis, nota-se que o impacto maior da escola no desempenho dos alunos. O impacto da turma deve ser olhado com muita ressalva, pois existem ao todo 13.167 turmas, o que quer dizer que se está trabalhando com grupos de 37 observações na média. Esses grupos são muito pequenos e existe pouca variância dentro deles, o que quer dizer que podem estar representando apenas a desigualdade entre os alunos e não entre as turmas.

É interessante notar a contribuição bruta das escolas em torno de 15% da desigualdade total. Comparando o impacto da variável escola com o impacto das variáveis escolares, levanta-se a hipótese de que a contribuição da escola é dificilmente captada somente pela observação dos insumos escolares, o que não é o mesmo que dizer que não tenha uma contribuição significativa. A natureza complexa da contribuição da escola para a proficiência escolar talvez não possa ser captada dessa maneira.

Um exercício interessante é observar o impacto das escolas entre os alunos de diferentes *background*:

Tabela 16 - Contribuição Bruta da Escola por Número de Pessoas por Quarto

	Desigualdade entre Grupos -		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Acima de 2 Pessoas por Quarto	0,02704	0,03205	100%	100%
Escola	0,00393	0,00430	15%	13%
Até 2 Pessoas por Quarto	0,02673	0,03100	100%	100%
Escola	0,00446	0,00469	16%	15%

Não há uma diferença significativa entre o impacto das escolas nas crianças de acordo com a *proxy* de renda utilizada acima. As escolas parecem ter o mesmo impacto para crianças de diferentes níveis de renda da rede pública de ensino do estado de São Paulo.

Os municípios contribuem com 5% da desigualdade entre os alunos no desempenho em matemática. Nota-se também que os municípios explicam menos do desempenho em português.

Nenhuma variável parece explicar significativamente o desempenho dos alunos, apenas os municípios, tamanho da escola e o número de alunos por sala se destacam, apresentando uma contribuição bruta acima de 2%. É necessário ainda observar as contribuições e observar o que acontece com as variáveis quando são examinadas em relação umas com as outras.

5.1 Contribuições Marginais

As contribuições brutas levam em consideração a divisão da amostra em grupos de acordo com apenas uma variável. Mas podemos dividir a amostra em grupos segundo os critérios de várias variáveis. Podemos, por exemplo, dividir a amostra em grupos de acordo com a cor, branco ou não branco, e o gênero, masculino ou feminino. Ao dividirmos a amostra segundo essas categorias, teríamos 4 grupos. Calculando L_e saberíamos quanto da desigualdade é explicada por essas duas variáveis em conjunto.

A maneira de saber qual a contribuição da variável cor condicionada ao gênero dos indivíduos é recalculando a desigualdade entre grupos sem a variável cor, L_e' , e fazer: $L_e - L_e'$. Esta diferença nos diz qual a contribuição da variável cor na desigualdade levando em consideração o gênero, isto é, qual a contribuição marginal da variável cor condicionada a uma outra variável para a desigualdade total entre os alunos.

Isso nos permite trabalhar com várias variáveis conjuntamente e explorar o seu impacto sobre a desigualdade total dos alunos, levando em conta outras variáveis que também explicam a desigualdade. Há um limite para a formação de grupos, que é a necessidade de garantirmos um número mínimo de observações para cada um, pois se tivéssemos tantos grupos quanto observações, não poderíamos atribuir a desigualdade explicada aos grupos formados, já que estes se confundiriam com os indivíduos.

Assim, é necessário que cada grupo tenha um limite mínimo de observações, para que havendo variação dentro dos grupos, possamos identificar a desigualdade observada com os grupos formados e atribuir essa desigualdade às variáveis que os geraram.

Neste trabalho estamos utilizando uma amostra com 352.248 observações, o que nos permite trabalhar com até cerca de 9.893 grupos razoavelmente, já que esse valor, levando em

conta a ponderação da amostra, nos dá uma média de 50 observações por grupo, garantindo assim que haja variação dentro deles.

Outro ponto com respeito à criação dos grupos é que é interessante diminuir o número de categorias das variáveis para que possamos trabalhar com mais variáveis. Assim, transformamos todas as variáveis utilizadas abaixo em variáveis binárias em que é atribuído o valor 1 para a categoria que mais se destaca das demais. Para o caso da variável tamanho da escola em alunos, por exemplo, suas 6 categorias foram reduzidas a duas: escolas com 2000 alunos ou mais e escolas com menos.

Segue abaixo a análise da contribuição marginal das variáveis ao dividirmos os grupos segundo município, escola e turma:

Tabela 17 - Contribuições Marginais da Escola, Turma e Município

	Theil - L		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos (L)	0,02704	0,03175	100,00%	100,00%
Desigualdade entre Grupos (L_e)	0,00570	0,00481	21,09%	15,15%
Contribuição Marginal ($L_e - L_e'$)				
Município	0,00141	0,00122	5,21%	3,84%
Escola	0,00253	0,00299	9,36%	9,43%
Turma	0,00176	0,00060	6,51%	1,89%

Dividiu-se a amostra em grupos segundo município, escola e turma e calculou-se a desigualdade L_e entre esses grupos. Como a escola é uma subdivisão dos municípios e a turma uma subdivisão da escola, esses grupos equivalem a dividir a amostra segundo as turmas e por isso a contribuição L_e equivale à contribuição bruta das turmas mais acima. A desigualdade entre grupos explica 21% da desigualdade entre os alunos para proficiência em matemática, já em português esse percentual é menor, 15%.

Levando em consideração escolas e turmas, os municípios explicam 3,8 e 5,2% das notas dos alunos para português e matemática. Já as escolas explicam cerca de 9,5% do desempenho dos alunos em ambas disciplinas. A variável turma foi a que apresentou maior variação, explicando quase 7% do desempenho em matemática e 2% em português.

Com as três variáveis acima foram gerados 13.167 grupos. É necessário fazer uma ressalva sobre o impacto desses grupos, como colocado acima, esses número deixam poucas observações por grupo. Entretanto, é válida a análise acima como uma panorama geral do impacto das variáveis no desempenho dos alunos.

Uma outra maneira de olhar os dados acima é perceber que L_e nesse caso é a desigualdade total entre as turmas. Podemos calcular a contribuição marginal dos municípios para a desigualdade total entre as turmas fazendo $0,00141/0,00570$, o que resulta que 25% da desigualdade entre as turmas em proficiência em matemática é causada pela desigualdade entre os municípios e 45% pela desigualdade entre as escolas. Já para proficiência em português municípios e escolas respondem respectivamente por 25 e 62% da desigualdade entre as turmas. É interessante notar o impacto maior das escolas na desigualdade em matemática das turmas.

Cabe lembrar que a análise levada nos últimos parágrafos referente às turmas só é possível porque as variáveis são subdivisões exatas umas das outras. Isto é, não é possível que uma mesma turma esteja em duas escolas ou uma escola em dois municípios, o que não é verdade para as variáveis que vem a seguir.

Lembrando que a desigualdade medida pelo Theil-L é composta pela soma da desigualdade entre grupos e intra-grupos, o que a tabela acima nos diz é que 21% e 15% das notas dos alunos são explicadas pelas escolas, municípios e turmas em que estão alocados e que, portanto, 79% e 85% da desigualdade entre os alunos se deve à desigualdade entre os alunos dentro das turmas.

Como foi colocado acima, não é possível, sem prejuízo da análise, subdividir os grupos gerados por município, escolas e turmas em mais grupos segundo outras variáveis, pois ficaríamos sem observações suficientes por grupos. Assim, o que vamos fazer é procurar encontrar possíveis determinantes dos alunos e das escolas das desigualdades de notas.

O banco de dados utilizado neste trabalho também não dispõe de informações referentes aos municípios, apenas o nome do município em que a escola está situada. Assim, nos restam dois níveis para serem explorados, as escolas e os alunos.

Tabela 18 - Contribuições Marginais das Variáveis de *Background* dos Alunos

	Theil - L		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos (L)	0,02704	0,03175	100,00%	100,00%
Desigualdades entre Grupos (L_e)	0,00043	0,00054	1,61%	1,70%
Contribuição Marginal (L_e - L_e')				
Menina	0,00005	0,00005	0,17%	0,17%
Negro ou Pardo	0,00009	0,00010	0,33%	0,32%
Idade acima da ideal	0,00005	0,00006	0,17%	0,20%
Já foi reprovado	0,00005	0,00007	0,19%	0,21%
Já abandonou a escola	0,00006	0,00007	0,22%	0,23%
Mãe com Ensino Superior	0,00005	0,00007	0,17%	0,22%
Pai com Ensino Superior	0,00006	0,00007	0,20%	0,22%
Pessoas por quarto	0,00011	0,00013	0,41%	0,40%
Participa do Bolsa Família	0,00006	0,00007	0,21%	0,23%
Frequentou o Ensino Infantil	0,00009	0,00011	0,32%	0,34%

Os dados acima consistem no resultados que foi encontrado para a desigualdade entre os grupos que foram gerados segundo as seguintes variáveis: gênero, cor, idade, reprovação, abandono, escolaridade do pai, escolaridade da mãe, número de pessoas por quarto, participação no bolsa família e início da vida escolar. A variável pessoas por quarto foi alterada para que ficasse com duas categorias, o valor zero foi substituído pelo número de pessoas na casa do aluno e o valor 1 foi atribuído para os valores acima de 2 da variável assim modificada.

Foram gerados 1.011 grupos e é interessante notar o baixo percentual que a desigualdade entre grupos explica da desigualdade entre os alunos, menos de 2%. Olhando a contribuição de cada uma das variáveis para a variação entre grupos, percebemos que nenhuma delas se destaca, estando todas abaixo de 0,5%.

Observemos agora qual a contribuição marginal das variáveis de background quando as condicionamos à escola do aluno:

Tabela 19 - Contribuições Marginais das Variáveis de *Background* dos Alunos Condicionadas às Escolas

	Theil - L		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos (L)	0,02704	0,03175	100,00%	100,00%
Desigualdades entre Grupos (L_e)				
Escola + Menina	0,00430	0,00462	15,89%	14,55%
Escola + Negro ou Pardo	0,00429	0,00463	15,85%	14,59%
Escola + Idade acima da Ideal	0,00426	0,00459	15,75%	14,45%
Escola + Já foi reprovado	0,00429	0,00464	15,86%	14,62%
Escola + Já abandonou a escola	0,00426	0,00461	15,77%	14,52%
Escola + Mãe com Ensino Superior	0,00428	0,00464	15,84%	14,61%
Escola + Pai com Ensino Superior	0,00430	0,00462	15,91%	14,56%
Escola + Pessoas por quarto	0,00429	0,00462	15,88%	14,54%
Escola + Partcipa do Bolsa Família	0,00429	0,00464	15,85%	14,62%
Escola + Frequentou Ensino Infantil	0,00428	0,00463	15,84%	14,58%
Contribuição Marginal ($L_e - L_e'$)				
Menina	0,00036	0,00041	1,32%	1,28%
Negro ou Pardo	0,00034	0,00042	1,27%	1,32%
Idade acima da ideal	0,00032	0,00038	1,18%	1,18%
Já foi reprovado	0,00035	0,00043	1,28%	1,35%
Já abandonou a escola	0,00032	0,00040	1,19%	1,26%
Mãe com Ensino Superior	0,00034	0,00043	1,26%	1,34%
Pai com Ensino Superior	0,00036	0,00041	1,34%	1,29%
Pessoas por quarto	0,00035	0,00040	1,30%	1,27%
Participa do Bolsa Família	0,00034	0,00043	1,27%	1,35%
Frequentou o Ensino Infantil	0,00034	0,00042	1,26%	1,31%

Foi feita a divisão da amostra em grupos de acordo com a variável escola e uma variável de *background* do aluno. Cada linha abaixo do campo desigualdade entre os grupos na primeira coluna corresponde a uma diferente dos composição dos grupos, feita com a variável escola e a respectiva variável na linha. Na parte da contribuição marginal temos a contribuição das variáveis de acordo com sua respectiva composição no bloco acima. Para cada uma das linhas de desigualdade entre grupos foram gerados entre 8.675 e 9.104 grupos.

Condicionando as variáveis de background dos alunos à escola, a sua contribuição marginal é maior do que quando consideramos apenas variáveis de background como foi visto na tabela anterior. Entretanto, ainda assim, a contribuição dessas variáveis para a desigualdade entre os alunos é relativamente pequena, sempre abaixo de 1,5%.

As variáveis de background não parecem explicar muito da desigualdade entre os alunos, resta então buscar as variáveis que expliquem a contribuição da escola nas desigualdades de notas.

Tabela 20 - Contribuições Marginais das Variáveis Referentes às Escolas

	Theil - L		% Desigualdade entre Escolas	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	0,00394	0,00421
Contribuição Bruta das Escolas	0,00394	0,00421	100%	100%
Desigualdades entre Grupos (L_e)	0,00368	0,00391	93,28%	92,72%
Contribuição Marginal ($L_e - L_e'$)				
Município	0,00057	0,00059	14,56%	13,98%
Ciclos de 4 anos	0,00000	0,00001	0,11%	0,14%
Fundamental Completo	0,00004	0,00004	0,92%	0,85%
Supletivo	0,00004	0,00006	0,92%	1,32%
Tamanho	0,00005	0,00005	1,18%	1,22%
Alunos por Sala	0,00005	0,00006	1,18%	1,32%
Ventiladores	0,00005	0,00007	1,35%	1,66%
Tvs por Alunos	0,00003	0,00003	0,67%	0,61%
Video por Aluno	0,00002	0,00002	0,42%	0,38%
Antena Parabólica	0,00008	0,00009	1,94%	2,04%
Mimeógrafos por Professor	0,00007	0,00008	1,68%	1,80%
Professor corrige a lição	0,00009	0,00010	2,19%	2,27%
Média de Pessoas por quarto	0,00005	0,00005	1,18%	1,11%
% Pardos ou Negros	0,00006	0,00007	1,46%	1,63%
% Alunos que Abandonaram	0,00001	0,00001	0,16%	0,22%
% Alunos Repetentes	0,00004	0,00006	1,08%	1,37%
% Alunos Reprovados	0,00003	0,00003	0,88%	0,68%
% Alunos acima da idade	0,00005	0,00006	1,32%	1,54%
Renda Mínima	0,00006	0,00007	1,43%	1,56%
TV Escola	0,00008	0,00010	1,94%	2,27%
Transporte Escolar	0,00004	0,00005	0,92%	1,09%

Tabela 21 - Contribuições Marginais das Variáveis Referentes às Escolas

	Theil - L		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	100,00%	100,00%
Contribuição Bruta das Escolas	0,00394	0,00421	14,58%	13,27%
Desigualdades entre Grupos (L_e)	0,00368	0,00391	13,60%	12,30%
Contribuição Marginal ($L_e - L_e'$)				
Município	0,00057	0,00059	2,12%	1,86%
Ciclos de 4 anos	0,00000	0,00001	0,02%	0,02%
Fundamental Completo	0,00004	0,00004	0,13%	0,11%
Supletivo	0,00004	0,00006	0,13%	0,18%
Tamanho	0,00005	0,00005	0,17%	0,16%
Alunos por Sala	0,00005	0,00006	0,17%	0,18%
Ventiladores	0,00005	0,00007	0,20%	0,22%
Tvs por Alunos	0,00003	0,00003	0,10%	0,08%
Video por Aluno	0,00002	0,00002	0,06%	0,05%
Antena Parabólica	0,00008	0,00009	0,28%	0,27%
Mimeógrafos por Professor	0,00007	0,00008	0,25%	0,24%
Professor corrige a lição	0,00009	0,00010	0,32%	0,30%
Média de Pessoas por quarto	0,00005	0,00005	0,17%	0,15%
% Pardos ou Negros	0,00006	0,00007	0,21%	0,22%
% Alunos que Abandonaram	0,00001	0,00001	0,02%	0,03%
% Alunos Repetentes	0,00004	0,00006	0,16%	0,18%
% Alunos Reprovados	0,00003	0,00003	0,13%	0,09%
% Alunos acima da idade	0,00005	0,00006	0,19%	0,20%
Renda Mínima	0,00006	0,00007	0,21%	0,21%
TV Escola	0,00008	0,00010	0,28%	0,30%
Transporte Escolar	0,00004	0,00005	0,13%	0,14%

Quando foram apresentadas as contribuições brutas, vimos que a contribuição da variável escola para a desigualdade dos alunos é de 0,00394 para matemática e 0,00391 para português. Estas desigualdades correspondem à desigualdade total entre as escolas. Foi também calculada a contribuição marginal da variável escola quando geramos grupos com as variáveis município, escola e turmas e que não é possível trabalhar com esses níveis devido à limitação da amostra disponível. Assim, a análise limitou-se a tentar captar o efeito de variáveis de *background* condicionadas à escola do aluno.

Tendo em vista que o fator mais relevante na desigualdade entre os alunos encontrado até aqui é o das escolas, as tabelas acima visam apresentar quais são os determinantes da desigualdade entre as escolas, que já havia sido medida pela contribuição bruta dessa variável. A tabela 20 refere-se ao quanto as variáveis relacionadas explicam da desigualdade entre as escolas, condicionadas às outras variáveis. Já a tabela 21 refere-se ao quanto que as variáveis das escolas condicionadas umas às outras explicam da desigualdade total entre os alunos.

Foram utilizadas as seguintes variáveis: município, ciclos de 4 anos, fundamental completo, supletivo, tamanho, alunos por sala, ventiladores, tvs por alunos, vídeos por alunos,

antena parabólica, mimeógrafos por professor, se o professor corrige a lição, média de pessoas por quarto nas famílias da escola, o percentual de alunos pardos ou negros, de alunos reprovados, repetentes, que abandonaram a escola e acima da idade ideal. Também estão incluídas variáveis referentes à participação da escola em programas de renda mínima, no programa TV Escola do governo federal e a presença de transporte escolar.

Todas as variáveis são binárias. A variável tamanho indica se uma escola tem até 2000 alunos ou mais, alunos por sala se a escola tem uma média superior a 40 alunos por sala de aula, a variável ventiladores se a escola possui mais de 2 ventiladores por sala. A variável mimeógrafo por professor indica se há mais de 10 desse equipamento para cada 100 professores. O percentual alunos pardos ou negros indica se há a proporção desses alunos é igual ou superior a 60% na 4ª série de sua escola, o percentual de alunos nascido acima da idade se este esta entre 10 e 15%, o percentual de reprovação e repetentes se este é superior a 15% na série e o de abandono se este é superior a 5%.

Duas variáveis foram geradas para as escola a partir dos dados dos alunos: a variável professor corrige a lição, que é igual a um se o número de alunos que responderam que o professor sempre corrige a lição é superior a 85% na escola e a variável média de quartos por pessoa, que indica um se a média de pessoas por quarto nas famílias da escola é superior a dois.

A variável que mais chama atenção é o município, que chega a explicar 15% da desigualdade entre as escolas na proficiência de matemática. Entre as variáveis que se pode chamar de insumos escolares, nenhuma delas teve um impacto de maior destaque. É interessante o efeito da variável se o professor corrige a lição que acabou demonstrando a maior contribuição marginal entre as variáveis escolares. É interessante notar que essa variável não mede um insumo propriamente dito, mas uma forma de trabalho dentro da sala de aula. Utilizando todas as variáveis, encontra-se a explicação para 93,28% da desigualdade dos alunos em matemática e 92,72% para a desigualdade em português.

Outro aspecto importante identificado nas tabelas acima é contribuição marginal da variável TV Escola, que analisada conjuntamente com outras variáveis escolares passa a apresentar um impacto maior do que algumas variáveis associadas aos colegas. Variáveis de qualidade da infra-estrutura como o número de mimeógrafos por professor e a presença de antena parabólica também se destacaram.

Ao olharmos quanto da contribuição marginal das variáveis escolares explica do total da desigualdade entre os alunos, vemos que os percentuais são muito pequenos, como foi observado até aqui para as outras variáveis. Mas é importante notar que embora expliquem

muito pouco separadamente, as variáveis em conjunto explicam 13,6 e 12,30% da desigualdade entre os alunos conjuntamente. Esse resultado é interessante porque parece indicar que embora o efeito das variáveis pareça pequeno, quando as observamos em interação umas com as outras, este se torna mais significativo.

As tabelas acima apontam para a necessidade de se compreender a natureza complexa da relação entre insumos escolares e desempenho dos alunos. As variáveis referentes às características dos colegas tem seu impacto diminuído quando observamos todas variáveis das escolas e as variáveis de programas tem seu efeito aumentado.

Podemos ainda olhar para os dados acima no seguinte formato:

Tabela 22 - Contribuições Marginais de Diferentes Características Referentes às Escolas

	Desigualdade entre Grupos - Theil-L		% Desigualdade entre Escolas	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	0,00394	0,00421
Escolas	0,00394	0,00421	100%	100%
Todas Variáveis das Escolas	0,00368	0,00391	93,28%	92,72%
Município	0,00057	0,00059	14,46%	14,01%
Infraestrutura Escolar	0,00074	0,00085	18,68%	20,08%
Professores	0,00009	0,00010	2,28%	2,37%
Colegas	0,00035	0,00040	8,79%	9,40%
Programas Governamentais	0,00018	0,00021	4,48%	4,89%

Tabela 23 - Contribuições Marginais de Diferentes Características Referentes às Escolas

	Desigualdade entre Grupos -		% Desigualdade Total	
	Matemática	Português	Matemática	Português
Desigualdade Total entre os Alunos	0,02704	0,03175	100,00%	100,00%
Escolas	0,00394	0,00421	14,58%	13,27%
Todas Variáveis das Escolas	0,00368	0,00391	13,60%	12,30%
Município	0,00057	0,00059	2,11%	1,86%
Infraestrutura Escolar	0,00368	0,00391	2,72%	2,66%
Professores	0,00009	0,00010	0,33%	0,31%
Colegas	0,00368	0,00391	1,28%	1,25%
Programas Governamentais	0,00368	0,00391	0,65%	0,65%

Nas tabelas 22 e 23 considerou-se um grupo de variáveis como se fosse apenas uma variável. Assim, podemos calcular a contribuição marginal de um conjunto de variáveis sobre a desigualdade entre as escolas e entre os alunos. As tabelas acima consistem na estimação do índice Theil-L para a contribuição entre grupos L_e para todas variáveis das escolas conforme foi feito nas tabelas 20 e 21. Depois, ao invés de se retirar apenas uma variável e calcular L_e e

fazer $L_e - L_e'$ e verificar a contribuição marginal dessa variável, foram retiradas várias variáveis conjuntamente, como se representassem um única variável. A decomponibilidade do índice nos permite fazer isso, o que nos dá uma maneira de analisar o impacto de um grupo de variáveis conjuntamente nas desigualdades.

Tomando todas as variáveis das tabelas 20 e 21, foram consideradas variáveis de infraestrutura: município, ciclos de 4 anos, fundamental completo, supletivo, tamanho, alunos por sala, ventiladores, tvs por aluno, vídeos por aluno, antena parabólica e mimeógrafo por professor. As seguintes variáveis foram classificadas como colegas: média de pessoas por quarto, percentual de pardos e negros, percentual de abandono, percentual de alunos repetentes, percentual de alunos reprovados e percentual de alunos acima da idade. Na categoria programas governamentais consideramos os programas TV Escola, programas de renda mínima e transporte escolar. E por fim, as variáveis município e se o professor corrige a lição referente a professores.

Olhando a tabela 22, vemos que todas as categorias à exceção do município explicam mais da desigualdade em português do que em matemática. Tomando as variáveis de infraestrutura em conjunto, notamos um impacto maior da infra-estrutura escolar do que do município, o que não poderíamos ter observado nas tabelas anteriores. Chama a atenção que 8,8 e 9,4% das desigualdades entre as escolas sejam explicadas pelos colegas. Os programas quando tomados em conjunto também apresentam uma contribuição marginal maior.

Na tabela 23, percebemos que o impacto da infra-estrutura escolar na desigualdade total é relativamente pequeno, como era esperado. Mas é interessante notar que embora as variáveis tenham explicado mais da desigualdade entre as escolas em português do que em matemática, elas explicam relativamente menos da desigualdade total em português. Isso significa que embora as variáveis analisadas até agora expliquem menos da desigualdade de português, essas mesmas variáveis explicam mais do efeito da escola nas desigualdades em português.

Um resultado importante que foi possível graças ao uso dos dados da Prova Brasil é a contribuição da variável município. Esta variável sozinha explica 5,21% da desigualdade total entre os alunos na proficiência de matemática e 3,84% da proficiência de português, como pode ser observado nas contribuições brutas. Quando condicionada às variáveis da escola essa contribuição cai para 2 e 1,86%. É interessante observar que o município explica 14,56% da desigualdade em matemática entre as escolas e 13,98% da desigualdade em português.

Mesmo quando agregamos as variáveis que foram utilizadas para calcular as contribuições marginais das escolas e formamos os grupos que passam a explicar mais da

desigualdade entre as escolas do que as variáveis sozinhas, não temos uma alteração considerável da contribuição marginal do município. Isso quer dizer que a interação das variáveis de infra-estrutura, colegas e programas governamentais não captura parte uma parcela da desigualdade que era capturada pela variável município na formulação anterior.

Podemos dizer, então, que a variável município esta associada a fatores que não aqueles associados a essas variáveis que são comumente observadas. A seguir, foi realizada uma estimação por mínimos quadrados, com as variáveis que foram utilizadas até aqui e atribuiu-se uma *dummy* para cada município:

Tabela 22 - MQO com *dummies* de Municípios

	Matemática		Português	
	Coef.	P> Z	Coef.	P> Z
nota				
média de pessoas por quarto	-15,267	0,000	-17,110	0,000
2 ciclos de 4 anos	-4,200	0,000	-5,228	0,000
departamento administrativo	1,876	0,000	4,476	0,000
renda mínima	-1,874	0,000	-2,286	0,000
transporte escola	-1,813	0,000	-1,848	0,000
Professor corrige a lição	1,559	0,000	1,091	0,000
tv escola	1,543	0,000	1,583	0,000
ventiladores por sala	1,292	0,000	1,152	0,000
já abandonou a escola	-1,095	0,006	-0,982	0,019
fundamental	-1,040	0,001	-1,236	0,000
frequentou o ensino infantil	0,682	0,001	0,486	0,026
idade acima da ideal	-0,556	0,149	0,096	0,812
2 primeiros ciclos de 2 anos	-0,391	0,705	0,423	0,695
% abandono	-0,318	0,000	-0,320	0,000
supletivo	-0,286	0,303	-0,555	0,057
negro ou pardo	-0,270	0,171	-0,555	0,007
já foi reprovado	-0,243	0,405	-0,876	0,004
% alunos pardos ou negros	-0,161	0,000	-0,170	0,000
escolaridade da mãe	0,121	0,000	0,195	0,000
vídeo por aluno	-0,120	0,126	-0,060	0,464
alunos por sala	-0,112	0,000	-0,120	0,000
% alunos repetentes	-0,107	0,000	-0,134	0,000
menina	0,103	0,593	-0,306	0,131
tv por aluno	0,085	0,217	-0,102	0,157
% alunos reprovados	-0,082	0,000	-0,092	0,000
antena parabólica	0,073	0,675	0,281	0,122
número de pessoas por quarto	0,053	0,358	0,140	0,021
% alunos com idade acima	0,047	0,001	0,052	0,001
mimeógrafos por professor	0,014	0,273	0,047	0,001
constante	250,511	0,000	257,116	0,000

Nessa primeira tabela estão os controles utilizados na regressão sem as *dummies* por município, já que como possuímos 526 municípios na amostra, seria impraticável apresentá-las todas. A regressão não foi feita de forma rigorosa, seu intuito é apenas apresentar o potencial explicativo que os municípios podem oferecer. As variáveis referentes aos programas provavelmente sofrem de endogeneidade, mas não é o intuito discutir os parâmetros da equação, mas apresentar os coeficientes das *dummies* de município como indicativo que a introdução dessa variável pode oferecer ao modelo.

Um breve comentário sobre as variáveis de controle acima que é interessante de ser feito é que as variáveis que apresentaram maiores coeficientes são referentes às escolas, fornecendo mais um indício de que os insumos escolares são importantes na explicação do desempenho dos alunos. Vejamos abaixo as tabelas com os 20 maiores e 20 menores

coeficientes das *dummies* de cidades, nos apêndices A e B temos duas tabelas com os 100 maiores e os 100 menores coeficientes:

Tabela 24 - MQO com *dummies* de Municípios

Matemática		Português	
20 maiores coeficientes		20 menores coeficientes	
cidade	coef.	cidade	coef.
BARRA DO CHAPEU	45,26	1 SAGRES	23,17
VALPARAISO	35,48	2 SETE BARRAS	20,30
PONGAI	35,37	3 DOLCINOPOLIS	17,70
CASSIA DOS COQUEIROS	34,21	4 BOM JESUS DOS PERDOES	17,20
BOM JESUS DOS PERDOES	32,91	5 SAO JOAO DAS DUAS PONTES	12,48
RUBIACEA	32,24	6 SUD MENNUCCI	12,21
FATURA	31,44	7 CABREUVA	11,10
DOLCINOPOLIS	29,81	8 TORRINHA	10,19
SAO JOAO DAS DUAS PONTES	28,23	9 RUBIACEA	10,19
TORRINHA	27,81	10 ORINDIUA	10,17
SETE BARRAS	27,73	11 INDAIATUBA	8,92
OURO VERDE	25,84	12 OURO VERDE	8,72
SEBASTIANOPOLIS DO SUL	22,92	13 ALUMINIO	8,16
CAJURU	22,25	14 TABATINGA	7,77
SUD MENNUCCI	21,82	15 VOTUPORANGA	7,21
CABREUVA	20,31	16 EMBU	6,16
LUPERCIO	17,77	17 PIRACAIA	5,88
SANTO ANTONIO DO JARDIM	17,61	18 AMERICO BRASILIENSE	4,77
AMERICO BRASILIENSE	16,93	19 OSCAR BRESSANE	4,24
OSCAR BRESSANE	16,85	20 ITAPUI	3,85
ITAPIRA	-47,14	1 ITAPIRA	-51,34
EMILIANOPOLIS	-43,55	2 RIOLANDIA	-48,43
REGINOPOLIS	-41,56	3 EMILIANOPOLIS	-47,74
NOVA GUATAPORANGA	-39,45	4 ALTAIR	-47,43
ALTAIR	-38,59	5 FLORINEA	-46,45
PONTES GESTAL	-36,81	6 SALES	-45,82
SANTA CLARA D'OESTE	-36,29	7 NOVA GUATAPORANGA	-45,74
SAO JOSE DO BARREIRO	-33,88	8 SANTA CLARA D'OESTE	-45,18
SERRA NEGRA	-33,55	9 TRES FRONTEIRAS	-43,83
RIOLANDIA	-33,42	10 PIRAJUI	-43,34
PORTO FERREIRA	-33,38	11 PAULISTANIA	-43,01
PLATINA	-33,34	12 PONTES GESTAL	-42,58
BORACEIA	-33,03	13 SANTO EXPEDITO	-41,89
SABINO	-32,21	14 REGINOPOLIS	-41,59
ITAOCA	-32,19	15 TAQUARAL	-41,40
PIRAJUI	-31,93	16 ARAPEI	-41,35
SARUTAIA	-30,77	17 APARECIDA D'OESTE	-40,76
PEDRINHAS PAULISTA	-30,54	18 ITAPIRAPUA PAULISTA	-40,66
HOLAMBRA	-30,48	19 AVAI	-40,61
AREIAS	-30,21	20 CATIGUA	-40,57

As *dummies* de município para as cidades acima tem um impacto muito maior nas notas do que todas as outras variáveis. A variável dependente do modelo são as notas na

Prova Brasil, nota-se que a residência em um município pode impactar as notas na média em até 50 pontos, o que é um impacto bastante considerável.

Podemos resumir o que foi obtido por meio das contribuições brutas e marginais dizendo que a escola é a variável mais relevante para explicar o desempenho do aluno. As variáveis de *background* sócio-econômico pouco acrescentaram em poder explicativo. Ao buscar compreender quais são os determinantes da desigualdade entre as escolas, vimos que nenhuma variável olhada isoladamente parece contribuir muito para essa desigualdade, entretanto, tomando as variáveis escolares em grupos e conjuntamente conseguimos explicar mais de 90% do seu efeito sobre a desigualdade entre os alunos.

A variável que mais explica da desigualdade entre os alunos é o município. É interessante notar que seu impacto não está claramente relacionado com as variáveis apresentadas, mas se trata de um aspecto distinto. É interessante para trabalhos posteriores, aprofundar a compreensão do município na desigualdade entre as escolas e como se dá a interação entre os grupos de variáveis escolares que explicam muito mais em conjunto do que separadamente.

6. Considerações Finais

Os resultados da análise das contribuições brutas e marginais mostrou que 14,58% da desigualdade de desempenho dos alunos e 13,27% da desigualdade de notas em português podem ser explicados pela contribuição bruta da variável escola. O percentual total da desigualdade que se conseguiu explicar com as variáveis do Censo Escolar e da Prova Brasil de 2005 foi de 15,9% da desigualdade em português quando geramos grupos a partir das variáveis escola e pai com ensino superior e 14,6%, da desigualdade em português, quando analisamos os grupos formados pelas variáveis escola e já foi reprovado e escola e recebe o bolsa família.

Olhando apenas as contribuições brutas, observamos que todas as variáveis separadamente, à exceção da escola, contribuem muito pouco para explicar a desigualdade entre os alunos. Nas contribuições brutas destacaram-se as variáveis relacionadas ao *background* dos colegas e as variáveis de *background* dos alunos apresentaram as menores contribuições. As variáveis dos alunos ficaram no mesmo nível de variáveis referentes ao uso de tecnologia nas escolas.

Formando grupos com diferentes variáveis de *background* dos alunos com a variável escola, foi pouca a alteração na desigualdade explicada. Isso parece indicar que as variáveis utilizadas condicionadas às escolas onde estão os alunos representam condições semelhantes de vulnerabilidade e que deve haver forte correlação entre essas características. Assim, o impacto na desigualdade de desempenho representado por variáveis de cor, idade, gênero, renda, início da vida escolar, histórico de reprovação e abandono participação no programa bolsa família são muito parecidos e todos igualmente baixos em termos de percentual da desigualdade entre os alunos explicada.

As variáveis de *background* dos alunos quando condicionadas às escolas apresentaram contribuição marginal sempre inferior a 1,5%. Quando observamos a contribuição marginal das variáveis de *background* condicionadas às outras variáveis de *background*, notamos que pouca informação acrescentam na explicação da desigualdade e que todas conjuntamente ainda explicam apenas 1,6% da desigualdade em matemática e 1,7% da desigualdade em português.

A desigualdade entre as turmas é menor em português do que em matemática. E as escolas respondem por 62% da desigualdade entre as turmas de português, enquanto que esse percentual é de 45% para matemática.

Tendo em vista a baixa contribuição das variáveis de *background* e a contribuição significativa das escolas, pareceu mais conveniente observar quais são os fatores que determinam a desigualdade entre as escolas. As contribuições brutas indicaram que as variáveis do nível das escolas mais importantes para a desigualdade eram o tamanho da escola e número de alunos por sala, que contribuíam com 2,11 e 2,66% da desigualdade total de matemática. Outras variáveis que apresentaram resultados importantes foram as referentes ao percentual de negros ou pardos na série e o percentual de alunos acima da idade.

As variáveis referentes ao uso de tecnologia nas escolas apresentaram contribuição pequena, entretanto, da mesma magnitude das variáveis de background dos alunos. Entre as variáveis analisadas, as que apresentaram menores contribuições brutas foram as do grupo referentes aos professores.

Segundo as contribuições brutas, todas variáveis tomadas separadamente explicam muito pouco da desigualdade entre os alunos. Agregando as variáveis para formar grupos que explicassem a desigualdade entre as escolas, conseguiu-se explicar 93,2% da desigualdade entre as escolas na proficiência de matemática e 92,7% da desigualdade em português.

Foram utilizadas as variáveis dos colegas, dos professores e de infra-estrutura da escola para explicar a desigualdade entre as escolas. As contribuições marginais das variáveis mostraram-se ainda relativamente pequenas, apesar de terem seu valor aumentado quando condicionadas às outras variáveis, ainda assim são muito pequenos. À exceção ficou por conta da variável município, que explica 14,5% da desigualdade em matemática e 13,9% da desigualdade em português. As outras duas variáveis que responderam por mais de 2% da variação entre as escolas foram a média de pessoas por quarto na escola e o programa TV Escola.

O fato das variáveis explicarem muito conjuntamente e pouco separadamente indica que o efeito que elas desempenham sobre a proficiência alunos se dá através da interação entre elas mais do que por fatores isolados. Percebe-se isso também quando se observa a contribuição marginal das variáveis de infra-estrutura escolar tomadas conjuntamente. Embora separadamente nenhuma tenha um impacto superior a 1%, quando tomadas em conjunto respondem por 20% da desigualdade em português, mais do que a contribuição marginal de 14% dos municípios.

As variáveis que explicam mais da desigualdade entre as escolas são as variáveis de infra-estrutura escolar, seguida do município e das variáveis de *background* dos colegas. A variável utilizada para representar o impacto dos professores explica menos de 2,5% da

desigualdade entre as escolas. Um resultado interessante é o impacto que os programas escolas têm sobre o desempenho dos alunos, entre 4,5 e 5%.

Pode-se graças ao banco de dados utilizado observar o impacto da variável município, que pareceu bastante interessante. Isoladamente, município é a variável que mais explica a desigualdade entre as escolas.

A possibilidade de desagregação dos dados por município oferece muitas possibilidades de trabalhos posteriores.

Comparando com a literatura brasileira de análise de proficiência escolar, nota-se que este trabalho corroborou os resultados que apontam para a importância da escola na determinação do desempenho dos alunos. Além disso, mostrou que a observação da interação das variáveis pode trazer benefícios para a compreensão do fenômeno. Outro ponto levantado na análise das contribuições é que as variáveis de *background* dos alunos comumente utilizadas nos trabalhos explicam pouco da variação entre os alunos. Isso não quer dizer que as características dos alunos não sejam importantes para seu desempenho, mas, que essas variáveis não conseguiram explicar essa desigualdade.

De acordo com o que foi exposto acima, é interessante aprofundar o estudo do tema a partir da análise de caráter mais exploratório que se realizou neste trabalho procurando compreender mais da relação entre escola e município, entre as variáveis escolares e das características de *background* dos alunos com essas.

Referências Bibliográficas

AGHION, P.; HOWIT, P. *Endogenous growth theory*. Cambridge MA: MIT Press, 1998.

ALBERNAZ, A.; FERREIRA, F.; FRANCO, C. *Qualidade e equidade na educação fundamental brasileira*. Rio de Janeiro: Departamento de Economia PUC Rio, 2002 (Texto para Discussão nº 455).

ANGRIST, J. E., BETTINGER, E., BLOOM, E., King, Kremer, M., “Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment,” *American Economic Review*, December 2002.

BARROS, R.; MENDONÇA, R. *Os determinantes da desigualdade no Brasil*. Rio de Janeiro: IPEA, 1995 (Texto para Discussão nº 377).

BECKER, G. *Human Capital*. NBER, 1975.

BEHRMAN, J.; TAUBMAN, P. *Is schooling mostly in the genes? nature-nurture decomposition using data on relatives*. *Journal of Political Economy*, v 97(6), p 1425-1446, 1989.

COLEMAN, J.S. *Equality of educational opportunity*. Washington D.C : Office of Education, 1966.

CULLEN, J. B., JACOB, B.A., LEVITT, S.D. The Effect of School Choice on Student Outcomes: evidence from randomized lotteries. *NBER*, 2003. (Working Paper 10133)

DE FRIES, J. C.; FULKER, D. W. *Multiple regression analysis of twin data*. *Behaviour Genetics*, v 22, p 467-473, 1985.

DIAZ, M. *Efetividade no ensino superior brasileiro: aplicação de modelos multinível à análise dos resultados do exame nacional de cursos*. *Revista Economia A*, v8, p 93 120, 2007.

FELÍCIO, F. *O efeito da qualidade da escola sobre o desempenho escolar: uma avaliação do ensino fundamental do estado de São Paulo*. São Paulo: USP, 2004. (Dissertação de Mestrado)

FERNANDES, R. ; ANNUATTI NETO, F. ; PAZELLO, E . *Avaliação dos Salários dos Professores da Rede Pública de Ensino Fundamental em Tempos de FUNDEF*. *Economia Aplicada*, v. 8, n. 3, 2004.

FERNANDES, R. ; NATENZON, Paulo Esteban . *A Evolução Recente do Rendimento Escolar das Crianças Brasileiras: Uma Reavaliação dos Dados do SAEB*. In: XXX Encontro Nacional de Economia, 2002, Nova Friburgo RJ. Anais do XXX Encontro Nacional de Economia, 2002.

FIGLIO, D.; KENNY, L. *Individual teacher incentives and student performance*. National Bureau of Economic Research, 2006.

FRANCESCONI, M.; JENKINS, S.; SIEDLER, T. *Childhood family: structure and schooling outcomes: evidence for Germany*. Essex: Institute for Social Economics Research, 2005 (ISER Working Paper 2005-22).

GLEWWE, P. *Estimating the impact of peer group effects on socioeconomic outcomes: does the distribution of peer group characteristics matter?* Economics of Education Review. Pittsburgh, v 16(5): p 39-43, 1997.

GLEWWE, P.; ILIAS, N.; KREMER, M. *Teacher incentives*. National Bureau of Economic Research, 2003.

GREMAUD, A.; FELICIO, F.; BIONDI, R. *Indicador de efeito escola : uma metodologia para a identificação dos sucessos escolares a partir dos dados da Prova Brasil*, Brasília: INEP, 2005.

HANUSHEK, E. *The economics of schooling: production and efficiency in public schools* . Journal of Economic Literature, v 24, p 1141-1177, 1986.(Working Paper 10133)

HANUSHEK, E. *The impact of differential expenditures on school performance*. Educational Researcher, v p 45-52, 1989.

HANUSHEK, E.; KIMKO, D. *Schooling, labor force quality, and the growth of nations*. American Economic Review, vol. 90 (5) , p 1184–208, 2000.

HANUSHEK, E. *The trade-off between child quantity and quality*, Journal of Political Economy, v 100 (1), p 84–117, 1992

HANUSHEK E.; KAIN, J.; MARKMAN, J.; RIVKIN,S. *Does peer ability affects student achievement?*. Journal of Applied Econometrics, 2002.

HANUSHEK, E.; LUQUE, J. *Efficiency and equity in schools around the world*, Economics of Education Review, vol. 22 (4) . 2003.

HOENACK, S. *The economics of education in developing countries: an assessment of the state of the art*. Economics of Education Review, v 15(4), p 327-338, 1996.

KRUEGER, A. *Economic considerations and class size*. The Economic Journal, v113, p 34–63, 2003.

KRUEGER, A.; ZHU, P., *Another Look at the New York City School Voucher Experiment*. American Behavioral Scientist, v 47, 2004.

LAVY, V. *From forced busing to free choice in public schools: quasi-experimental evidence of individual and general effects*. National Bureau of Economic Research, 2006.

LUCAS, R. . *On the mechanics of economic development*. Journal of Monetary Economics., v 22(1), p 3-42, 1988.

MAYER, D. et alli. *School Choice in New York City after Three Years: An Evaluation of the School Choice Scholarships Program*, Mathematica, 2002.

MILLER, P.; MULVEY, C. MARTIN, N. *Genetic and environmental contributions to educational attainment in Australia*. *Economics of Education Review* v 20, p 211-224, 2001.

McLANAHAN, S.; SANDEFUR G. *Growing Up with a Single Parent: What Hurts, What Helps*. Cambridge, 1994.

MENEZES-FILHO, N. Educação e Desigualdade. In: MENEZES-FILHO, N.; LISBOA, M.(Org.). *Microeconomia e sociedade no Brasil*. Rio de Janeiro: Contra Capa, 2001.

MENEZES-FILHO, N., VASCONCELLOS, L., WERLANG, S. R. C. *Avaliando o impacto da progressão continuada no Brasil* [Resumo completo]. In Sociedade Brasileira de Econometria (Org.), *Anais do XXVII Encontro Brasileiro de Econometria*. Natal: SBE.

NELSON, R. PHELPS, E.; Investing in humans, technological diffusion, and economic growth. *American Economic Review*. Nashville, v 61: p 69-75, 1966.

PETERSON, P. et alli. *An Evaluation of the New York City School Choice Scholarship Program: The First Year*. Mathematica, 1998.

PSACHAROPOULOS, G. *Economics of education: a research agenda*. *Economics of Education Review*, v 14(4), p 399-344, 1996

RAJ, D. *Development Economics*. Princeton :Princeton Press,1998.

RIVKIN, S.; HANUSHEK, E.; KAIN, J. *Teachers, schools, and academic achievement*, National Bureau of Economic Research, 2001. ,(Working Paper No. 6691)

ROUSE, C. *Private school vouchers and student achievement: an evaluation of the milwaukee parental choice program*. *Quarterly Journal of Economics*, v 118, p 553 – 602, 1998.

SERVENINI, E. – *O impacto da violência nas escolas sobre a proficiência dos alunos*. Brasília: Prêmio IPEA-Caixa, 2006

THOMAS, V.;WANG, Y.; FAN, X. *A new dataset on inequality in education: Gini and Theil indices of schooling for 140 countries, 1960-2000*, World Bank, 2002.

TAUBMAN, P. *The determinants of earnings, genetics, family and other environments: a study of white male twins*. *American Economic Review*, v 68(5), p 858-870, 1976

Theil, H. *Economics and Information Theory*. Chicago: Rand McNally and Company, 1967.

APÊNDICE A - 100 Maiores Coeficientes Estimados para as Dummies de Município

Matemática		Português	
100 maiores coeficientes cidade	coef.	cidade	coef.
1 BARRA DO CHAPEU	45,26	1 SAGRES	23,17
2 VALPARAISO	35,48	2 SETE BARRAS	20,30
3 PONGAI	35,37	3 DOLCINOPOLIS	17,70
4 CASSIA DOS COQUEIROS	34,21	4 BOM JESUS DOS PERDOES	17,20
5 BOM JESUS DOS PERDOES	32,91	5 SAO JOAO DAS DUAS PONTES	12,48
6 RUBIACEA	32,24	6 SUD MENNUCCI	12,21
7 FARTURA	31,44	7 CABREUVA	11,10
8 DOLCINOPOLIS	29,81	8 TORRINHA	10,19
9 SAO JOAO DAS DUAS PONTES	28,23	9 RUBIACEA	10,19
10 TORRINHA	27,81	10 ORINDIUVA	10,17
11 SETE BARRAS	27,73	11 INDAIATUBA	8,92
12 OURO VERDE	25,84	12 OURO VERDE	8,72
13 SEBASTIANOPOLIS DO SUL	22,92	13 ALUMINIO	8,16
14 CAJURU	22,25	14 TABATINGA	7,77
15 SUD MENNUCCI	21,82	15 VOTUPORANGA	7,21
16 CABREUVA	20,31	16 EMBU	6,16
17 LUPERCIO	17,77	17 PIRACAIA	5,88
18 SANTO ANTONIO DO JARDIM	17,61	18 AMERICO BRASILIENSE	4,77
19 AMERICO BRASILIENSE	16,93	19 OSCAR BRESSANE	4,24
20 OSCAR BRESSANE	16,85	20 ITAPUI	3,85
21 SANTO ANTONIO DA ALEGRIA	16,69	21 SANTO ANTONIO DO JARDIM	3,77
22 GUARATINGUETA	16,67	22 VALPARAISO	3,67
23 VALENTIM GENTIL	16,66	23 SANTA MARIA DA SERRA	3,30
24 SANTA ADELIA	16,64	24 SERTAOZINHO	3,13
25 SANTA RITA D'OESTE	16,10	25 SANTANA DE PARNAIBA	2,99
26 ITAPUI	15,90	26 GUARAREMA	2,07
27 BREJO ALEGRE	15,42	27 CAMPO LIMPO PAULISTA	1,97
28 TAGUAI	15,42	28 DESCALVADO	1,77
29 SARAPUI	14,43	29 JOANOPOLIS	1,67
30 BROTAS	14,30	30 FARTURA	1,58
31 INDAIATUBA	14,07	31 LIMEIRA	1,54
32 POTIRENDABA	13,88	32 VISTA ALEGRE DO ALTO	1,25
33 SANTA MARIA DA SERRA	12,88	33 PEDRA BELA	0,75
34 JUMIRIM	12,80	34 GUARATINGUETA	0,72
35 VIRADOURO	12,67	35 CORDEIROPOLIS	0,72
36 RUBINEIA	12,46	36 RAFARD	0,57
37 VOTUPORANGA	12,31	37 SAO ROQUE	0,56
38 SAO VICENTE	11,23	38 VALENTIM GENTIL	0,32
39 SERTAOZINHO	11,17	39 SAO VICENTE	0,32
40 PARISI	10,72	40 SANTO ANTONIO DA ALEGRIA	0,13
41 SANTA LUCIA	10,67	41 CESARIO LANGE	-0,24
42 GUARACI	9,78	42 JUMIRIM	-0,32
43 ALUMINIO	9,69	43 RIO CLARO	-0,41
44 PONTAL	9,36	44 FLOREAL	-0,58
45 POMPEIA	9,23	45 ARUJA	-0,72
46 RESTINGA	9,22	46 RIBEIRAO PIRES	-0,78
47 FLOREAL	9,15	47 PINHALZINHO	-0,87
48 PIRACAIA	9,14	48 SAO JOSE DOS CAMPOS	-0,88
49 ORINDIUVA	8,84	49 ITATINGA	-1,23
50 RIO CLARO	8,82	50 CAJURU	-1,35
51 CAMPO LIMPO PAULISTA	8,43	51 CEDRAL	-1,89
52 AVANHANDAVA	8,43	52 CACHOEIRA PAULISTA	-1,95
53 SAO ROQUE	8,33	53 PIRAPORA DO BOM JESUS	-1,97
54 ARARAS	8,18	54 CASSIA DOS COQUEIROS	-1,99
55 PIQUEROBI	8,17	55 BROTAS	-2,06
56 ITARIRI	7,85	56 SUZANO	-2,17
57 ELIAS FAUSTO	7,59	57 PAULINIA	-2,25
58 DESCALVADO	7,47	58 ATIBAIA	-2,27
59 CEDRAL	7,27	59 PIQUEROBI	-2,32
60 SAGRES	7,24	60 MARABA PAULISTA	-2,32
61 ITATINGA	7,24	61 SARAPUI	-2,33
62 REGISTRO	7,19	62 APARECIDA	-2,74
63 ARACATUBA	7,17	63 ITUPEVA	-2,81
64 ILHABELA	7,15	64 JOAO RAMALHO	-2,87
65 DOIS CORREGOS	7,15	65 CERQUILHO	-3,00
66 PEDRA BELA	7,09	66 BARRA DO CHAPEU	-3,06

APÊNDICE B - 100 Menores Coeficientes Estimados para as Dummies de Município

Matemática		Português	
100 maiores coeficientes			
cidade	coef.	cidade	coef.
1 ITAPIRA	-47,14	1 ITAPIRA	-51,34
2 EMILIANOPOLIS	-43,55	2 RIOLANDIA	-48,43
3 REGINOPOLIS	-41,56	3 EMILIANOPOLIS	-47,74
4 NOVA GUATAPORANGA	-39,45	4 ALTAIR	-47,43
5 ALTAIR	-38,59	5 FLORINEA	-46,45
6 PONTES GESTAL	-36,81	6 SALES	-45,82
7 SANTA CLARA D'OESTE	-36,29	7 NOVA GUATAPORANGA	-45,74
8 SAO JOSE DO BARREIRO	-33,88	8 SANTA CLARA D'OESTE	-45,18
9 SERRA NEGRA	-33,55	9 TRES FRONTEIRAS	-43,83
10 RIOLANDIA	-33,42	10 PIRAJUI	-43,34
11 PORTO FERREIRA	-33,38	11 PAULISTANIA	-43,01
12 PLATINA	-33,34	12 PONTES GESTAL	-42,58
13 BORACEIA	-33,03	13 SANTO EXPEDITO	-41,89
14 SABINO	-32,21	14 REGINOPOLIS	-41,59
15 ITAOCA	-32,19	15 TAQUARAL	-41,40
16 PIRAJUI	-31,93	16 ARAPEI	-41,35
17 SARUTAIA	-30,77	17 APARECIDA D'OESTE	-40,76
18 PEDRINHAS PAULISTA	-30,54	18 ITAPIRAPUA PAULISTA	-40,66
19 HOLAMBRA	-30,48	19 AVAI	-40,61
20 AREIAS	-30,21	20 CATIGUA	-40,57
21 SAO FRANCISCO	-30,19	21 SERRA AZUL	-40,22
22 SANTO EXPEDITO	-29,81	22 SAO PEDRO DO TURVO	-39,72
23 ORIENTE	-29,47	23 RINOPOLIS	-39,71
24 PALMARES PAULISTA	-28,47	24 PORTO FERREIRA	-38,95
25 ALFREDO MARCONDES	-28,41	25 MIRA ESTRELA	-38,20
26 RINOPOLIS	-28,10	26 ALTO ALEGRE	-37,68
27 PAULICEIA	-27,94	27 MAGDA	-37,22
28 SANTO ANTONIO DO ARACANGU.	-27,32	28 SAO JOAO DE IRACEMA	-37,06
29 CAIABU	-27,31	29 OLEO	-37,06
30 ALVARES FLORENCE	-27,22	30 JULIO MESQUITA	-36,89
31 SERRA AZUL	-27,03	31 FLORA RICA	-36,72
32 SAO PEDRO DO TURVO	-26,35	32 BORACEIA	-36,58
33 GUAICARA	-26,12	33 SANDOVALINA	-36,04
34 REDENCAO DA SERRA	-26,01	34 SUZANAPOLIS	-35,19
35 SAO JOAO DE IRACEMA	-25,95	35 SANTO ANTONIO DO ARACANGU.	-34,53
36 PAULISTANIA	-25,85	36 PLANALTO	-34,48
37 ALAMBARI	-25,50	37 SAO FRANCISCO	-34,35
38 CAIUA	-25,49	38 HOLAMBRA	-34,18
39 SALES	-25,13	39 SARUTAIA	-34,10
40 ITAPIRAPUA PAULISTA	-25,09	40 PONTAL	-33,88
41 MIRA ESTRELA	-24,80	41 MARAPOAMA	-33,28
42 ARAPEI	-24,56	42 TANABI	-33,10
43 JULIO MESQUITA	-24,22	43 MACEDONIA	-32,96
44 SAO BENTO DO SAPUCAI	-24,19	44 PLATINA	-32,36
45 PALMITAL	-24,14	45 SERRA NEGRA	-32,30
46 RIBEIRAO DOS INDIOS	-24,10	46 PARANAPANEMA	-31,87
47 TEJUPA	-24,03	47 UCHOA	-31,68
48 URU	-24,02	48 NARANDIBA	-31,64
49 BURITAMA	-24,02	49 TURIUBA	-31,09
50 SUZANAPOLIS	-23,72	50 ALFREDO MARCONDES	-30,99
51 LUCIANOPOLIS	-23,67	51 TAMBAU	-30,92
52 CATIGUA	-23,58	52 IGUAPE	-30,88
53 APARECIDA D'OESTE	-23,27	53 PEDRANOPOLIS	-30,57
54 RIBEIRAO GRANDE	-23,26	54 PAULICEIA	-30,48
55 IPIGUA	-23,24	55 PEDRINHAS PAULISTA	-30,30
56 NANTES	-23,05	56 ORIENTE	-30,03
57 TRES FRONTEIRAS	-23,04	57 TEJUPA	-29,96
58 SANTA CRUZ DA ESPERANCA	-22,97	58 NHANDEARA	-29,90
59 IEPE	-22,83	59 GUARANI D'OESTE	-29,85
60 FLORIDA PAULISTA	-22,72	60 PARANAPUA	-29,83
61 PARANAPUA	-22,40	61 PONTALINDA	-29,59
62 GUZOLANDIA	-22,16	62 INUBIA PAULISTA	-29,47
63 COLOMBIA	-22,15	63 SAO BENTO DO SAPUCAI	-29,39
64 ROSANA	-21,97	64 RIBEIRAO CORRENTE	-29,35
65 TAMBAU	-21,86	65 AMERICO DE CAMPOS	-29,24
66 NARANDIBA	-21,70	66 NIPOA	-29,16
67 AVAI	-21,54	67 IEPE	-29,15
68 GUARANI D'OESTE	-21,48	68 PAULO DE FARIA	-28,93