

Desigualdade de Proficiência no Ensino Fundamental Público Brasileiro: Uma Análise de Decomposição

Luiz Guilherme Scorzafave

*Faculdade de Economia, Administração e Contabilidade Universidade de São Paulo,
Campus de Ribeirão Preto (FEA-RP/USP), Brasil*

Rodrigo Araujo Ferreira

SEBRAE, São Paulo, Brasil

Resumo

Este artigo analisa quais variáveis estão mais associadas à desigualdade de desempenho escolar entre alunos da 4ª série do ensino fundamental da rede pública de São Paulo. Foi realizada uma decomposição estática do índice Theil-L. Os resultados indicam que características de *background* sócio-econômico dos alunos (especialmente reprovação, se o aluno faz lição de casa, participação no programa Bolsa Família e número de pessoas na casa) são as mais relacionadas com a desigualdade de notas, correspondendo a 25% da desigualdade total. As variáveis de insumos escolares apresentaram menor influência se olhadas individualmente. Entretanto, em conjunto contribuem com cerca de 14% da desigualdade.

Palavras-chave: Desempenho Escolar, Escolas Públicas, Desigualdade, Proficiência, Insumos Escolares

Classificação JEL: I20, I21

Abstract

This article analyzes which variables most contribute to score inequality among 4th grade students from the elementary public schools from the São Paulo state. We implement a static decomposition of Theil-L index and the results show that the students' background variables are the most related to score inequality (reprobation, if the student does her homework, participation on the program *Bolsa Família*, number of people living at home), answering for 25% of total grades inequality. The school outputs variables had a minor impact if looked individually. However, together they contribute with 14% of score inequality.

1. Introdução

A partir do grande avanço obtido em termos de universalização do acesso ao ensino fundamental nos anos 90, a questão da qualidade educacional e de seus determinantes passou a ocupar papel de destaque na literatura brasileira. Essa temática é controversa desde o Relatório Coleman (Coleman 1966), que apontou uma intrigante insignificância dos insumos escolares na explicação da proficiência dos alunos nos Estados Unidos.

Grande parte da literatura busca explicar como diferentes variáveis afetam o desempenho médio dos alunos (embora Sprietsma e Waltenberg (2010), por exemplo, analise os efeitos em diferentes partes da distribuição de notas). Entretanto, há um aspecto que vem sendo negligenciado na discussão: a questão da desigualdade de desempenho e as variáveis a ela associadas. A preocupação com a desigualdade de desempenho é de fundamental importância, já que se, por exemplo, em uma mesma cidade os desempenhos dos alunos são muito diferentes, é importante sabermos em que medida são as diferenças no *background* dos alunos ou se são as questões ligadas ao ambiente escolar propriamente dito que estariam por trás desse resultado.

Alguns trabalhos como Albernaz et alii (2002), Soares (2006), Soares e Candian (2007) e Felício (2004) têm tratado a questão do impacto de insumos escolares na equidade de desempenho entre os alunos. Em geral, esses artigos procuram analisar em que medida os insumos escolares atenuam ou reforçam o impacto de características socioeconômicas médias dos alunos das escolas sobre a proficiência média dos alunos e não propriamente a contribuição desses fatores juntamente com as características de *background* para a desigualdade entre eles. Ainda quando se faz o uso de índices de desigualdade, como em Felício (2004), este é utilizado de forma auxiliar, contribuindo para análise do desempenho médio dos alunos.

Nesse sentido, o presente trabalho pretende discutir a questão da desigualdade de proficiência dos alunos da quarta série do ensino fundamental público brasileiro, buscando identificar quais fatores estariam relacionados a essa desigualdade. Há poucos estudos acerca desse tema para o Brasil. Duas exceções são os trabalhos de Soares (2006) e Soares e Marotta (2009). A vantagem do presente artigo com relação àqueles que já abordaram o tema para o Brasil é a utilização dos dados da Prova Brasil de 2005. Essa base de dados permite análises a um nível de desagregação que não era possível com as bases de dados educacionais até então disponíveis (Sistema de Avaliação do Ensino Básico, SAEB), enriquecendo as possibilidades de análise. O SAEB é uma avaliação amostral, cuja representatividade amostral se dá apenas em nível de unidade da federação. Assim, não é possível se obter resultados em nível municipal, por exemplo. Já a Prova Brasil, possui essa representatividade amostral mais desagregada, permitindo que compreendamos, por exemplo, o quanto da desigualdade da nota dos alunos está relacionado a diferenças entre os municípios.

* Recebido em janeiro de 2009, aprovado em maio de 2011.
E-mail addresses: scorza@usp.br, araujo.ferreira@gmail.com

A metodologia empregada consiste na decomposição estática da desigualdade do índice de desigualdade de Theil-L, que permite avaliar a contribuição de cada variável separadamente para a desigualdade, bem como o impacto de grupos de variáveis. A amostra desse trabalho consiste de alunos da 4^a série do ensino fundamental da rede pública do estado de São Paulo que realizaram a Prova Brasil em 2005. Por se tratar de uma base de dados com muito mais observações do que o SAEB, a Prova Brasil também é conveniente para ser aplicada a essa metodologia, que requer a criação de células definidas de acordo com uma série de características que possuam um razoável número de observações. Assim, quanto mais observações, mais detalhada pode ser a construção dessas células, permitindo uma investigação mais detalhada das possíveis variáveis por detrás da desigualdade de desempenho escolar.

O artigo está dividido em cinco seções além dessa introdução. A seguir, realizamos uma breve revisão da literatura acerca dos determinantes do desempenho escolar. A seção três apresenta a metodologia e os dados utilizados. Em seguida, é feita uma análise descritiva dos dados. A seção cinco trás os resultados e por fim, são tecidas as conclusões.

2. Revisão da Literatura

Nessa seção são apresentadas algumas evidências da literatura no que tange aos determinantes do desempenho escolar, tanto aqueles relacionados ao *background* socioeconômico dos alunos (renda familiar, escolaridade dos pais, raça, idade, sexo, etc.) como os associados aos insumos escolares (infraestrutura da escola, características dos professores e dos colegas de turma, etc.). Com isso, é possível identificar variáveis que sejam candidatas naturais para serem empregadas na decomposição da desigualdade de desempenho dos alunos. Ou seja, será possível saber qual a relação dessas variáveis com a desigualdade de desempenho escolar. Por fim, são abordados os trabalhos que já trataram da questão da desigualdade de proficiência para o caso brasileiro.

2.1. *Insumos escolares*

No que diz respeito a insumos escolares, 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. Em Hanushek e Luque (2003), o autor ainda discorda de um resultado encontrado por Krueger e Whitmore (1999), afirmando que turmas menores são benéficas apenas nas séries iniciais e não em todas as séries como Krueger e Whitmore (1999) apontara.

Dewey et alii (2000) contestam diversos resultados de Hanushek afirmando que a inclusão de variáveis de renda e qualidade de insumos escolares em uma mesma

função de produção pode gerar um problema de especificação, pois se estaria utilizando conjuntamente variáveis do lado da oferta e da demanda por educação. Assim, re-analisaram a base de dados utilizada em Hanushek (1986) buscando separar esses dois tipos de variáveis e encontram que 81% dos coeficientes referentes a insumos escolares são positivos (sendo 37% positivos e significantes, contra 20,3% em Hanushek (1986).

Observando especificamente o impacto dos professores, Hanushek e Luque (2003) diz que apenas 9% dos estudos avaliados por ele encontraram uma relação significativa e positiva entre a escolaridade dos professores e o desempenho dos alunos e 49% encontram uma relação positiva e significativa entre experiência dos professores e desempenho.

No Brasil, entretanto, os trabalhos têm encontrado impacto de diferentes insumos escolares, mesmo que pequeno. Albernaz et alii (2002), por exemplo, encontraram um impacto significativo da infra-estrutura escolar no desempenho dos alunos. Segundo os autores, 28% da diferença de desempenho entre os alunos pode ser explicada pelas diferenças entre as escolas e aspectos como a presença de ruído e o arejamento da sala impactam significativamente no seu desempenho.

Albernaz et alii (2002) encontraram um impacto positivo da escolaridade dos professores sobre as notas, mesmo controlando para diversos fatores. O salário dos professores também é significativo quando colocado no modelo como substituto da escolaridade. Além disso, o impacto dessas variáveis é maior para os alunos de melhores condições sócio-econômicas. Rios-Neto et alii (2002) também encontram evidência do efeito da escolaridade do professor sobre o desempenho dos alunos. Por sua vez, Felício (2004) obteve resultados inconclusivos tanto com relação à experiência do professor como com relação à escolaridade dos mesmos.

2.2. *Background socioeconômico*

De modo geral a literatura aponta para um impacto positivo do *background* sócio-econômico sobre o desempenho dos alunos. Albernaz et alii (2002), por exemplo, afirmam que 80% da diferença de desempenho entre os alunos é devida a diferenças no *background* socioeconômico. Além disso, também foi constatado um considerável impacto do *background* socioeconômico médio das escolas nas diferenças entre elas, ou seja, quanto mais “ricas” as escolas, melhores suas notas médias.

Outro resultado bastante interessante é o encontrado em Barros et alii (2001) referente ao impacto da escolaridade dos pais sobre os alunos. O estudo encontrou um impacto maior da educação da mãe sobre as filhas e do pai sobre os filhos (0,4 anos por ano de escolaridade).

Soares e Candian (2007) procuram avaliar a capacidade das escolas de melhorar a proficiência dos alunos. Usando dados do SAEB, os autores encontram um efeito escola de 11,4% para escolas públicas e de 17,2% para escolas particulares. É relevante destacar que os autores chegam a tais resultados após controlar pelo

nível sócio-econômico dos alunos, onde esse indicador é construído através de um índice que captura acesso a bens duráveis dos domicílios, construído por meio da Teoria de Resposta ao Item (TRI). Nenhum fator escolar isoladamente explica significativamente o efeito escola e os autores apontam para a importância da liderança da escola na gestão dos diferentes fatores escolares para o bom desempenho dos alunos.

Por sua vez, Soares e Collares (2006) apresentam uma abordagem alternativa para capturar o efeito do *background* sócio-econômico familiar sobre o desempenho dos alunos. Admitem que a condição familiar consiste em um “conceito multidimensional” (Soares e Collares 2006, p. 617) e consideram quatro dimensões que descreveriam a condição familiar: recursos econômicos, presença de bens culturais nas famílias, relação dos pais com a educação dos filhos e ausência (ou não) dos pais na família. Segundo os autores, a condição familiar teria efeito direto, mas também indireto sobre a proficiência ao possibilitar, por exemplo, a aquisição de bens culturais. Cabe destacar que os autores constroem um índice de recursos econômicos da família também usando TRI, aplicada a itens de conforto domiciliar (TV, geladeira, computador, automóvel) e com o uso de modelos estruturais, os autores conseguem estruturar uma sequência de relacionamento entre as variáveis estudadas:

“a família primeiro conquista os recursos econômicos; a seguir, (...) algumas (...) adquirem bens culturais; tendo essas duas condições, dedicam tempo para acompanhar a vida escolar de seus filhos. (...) Refletindo o ambiente favorável da casa, os estudantes desenvolvem melhor atitude em relação à escola. Tudo isto resulta em maior proficiência (...)”.

2.3. *Desigualdade de proficiência*

A literatura nacional que se debruça sobre a questão da desigualdade de proficiência é pequena se comparada àquela associada aos determinantes da proficiência. Assim, há poucos trabalhos que se preocupem em relacionar a desigualdade de notas com a desigualdade na distribuição dos diferentes insumos escolares e no *background* dos alunos.

Por exemplo, Soares e Andrade (2006) encontraram uma grande desigualdade de qualidade entre as escolas particulares de Belo Horizonte, indicando que a proficiência superior na média seria devida mais ao *background* socioeconômico dos alunos do que a influência da própria escola. Se fosse observado apenas o efeito médio das escolas, o impacto seria equivocadamente tido como nulo. Observaram também que as escolas têm pouco impacto na diminuição do efeito das desigualdades de *background* dos alunos sobre sua proficiência. Segundo os autores, “o sistema de educação básica de Belo Horizonte só consegue produzir qualidade na presença de alta iniquidade” (Soares e Andrade 2006).

Já Soares e Candian (2007) constataram que apenas alguns fatores escolares

apresentaram efeito sobre a desigualdade entre os alunos, mas atuando no sentido de reforçá-la. Segundo os autores, “os alunos com melhores condições se apropriam mais rapidamente das melhorias da escola e, assim, no final, as diferenças aumentam (...)” (Soares e Candian 2007, p. 14).

Mais na linha do presente artigo, Soares e Marotta (2009) definem equidade educacional deveria ocorrer quando a distribuição de notas fosse semelhante em todos os grupos definidos de acordo com características sociodemográficas (sexo, cor da pele, religião etc.). Eles encontram que a distribuição de notas dos alunos de cor de pele preta está mais concentrada em valores baixos de proficiência do que para os demais, demonstrando, assim, a existência de desigualdade por cor. O mesmo é válido para nível sócio econômico: aqueles com maior nível socioeconômico possuem distribuição de notas concentrada em valores mais altos do que os demais.

Por sua vez, Soares (2006) desenvolve uma medida de desigualdade de proficiência análoga ao coeficiente de Gini. Vale destacar que a distribuição de referência para o autor, não é dada por aquela em que todos os alunos possuem a mesmo nível de proficiência, mas pela distribuição de notas observada para os alunos de escolas no décimo superior da distribuição da variável que captura o nível sócio-econômico dos alunos. O autor define nível sócio-econômico usando atributos familiares como educação dos pais, presença de bens duráveis no domicílio, bem como se a família possui uma empregada doméstica trabalhando no domicílio. Calculando o equivalente ao coeficiente de Gini dentro de grupos definidos por características sociodemográficas, o autor aponta que há grandes diferenças no índice entre alunos de diferentes regiões, níveis sócio-econômicos e também em termos de cor da pele.

O presente artigo se aproxima mais de Soares (2006) ao aplicar uma metodologia consagrada no estudo da desigualdade de renda para o campo da desigualdade educacional. Embora não seja realizada uma análise por nível sócio-econômico do aluno, foi utilizada a Prova Brasil de 2005, que engloba praticamente toda a rede de ensino fundamental pública paulista e a decomposição de um índice de entropia generalizada que é exatamente decomponível, ao contrário do índice de Gini.

3. Metodologia e Dados

3.1. *Dados*

A análise da desigualdade de desempenho entre aluno da rede pública do estado de São Paulo no presente artigo é feita por meio da decomposição do índice Theil-L, com a utilização de uma base de dados formada pela Prova Brasil e o Censo Escolar, ambos de 2005. 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 4^a e 8^a 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. Além disso, a Prova Brasil avalia alunos apenas da rede pública (federal, estadual e municipal).

Cabe a ressalva que a Prova Brasil de 2005 foi amostral no estado de São Paulo, ou seja, apesar de todas as escolas da rede estadual terem participado do exame, naquelas em que havia mais do que uma turma de quarta série, nem todas as turmas participaram do exame.

Uma das grandes vantagens da Prova Brasil é que seus dados permitem inferências em nível municipal, o que não era possível no SAEB. Assim, pode-se analisar, por exemplo, o quanto da desigualdade de desempenho total está associada à desigualdade entre os municípios.

A amostra utilizada no trabalho conta com 352.248 alunos de quarta série da rede pública de São Paulo. As informações acerca da infra-estrutura e outras características das 4.552 escolas foram extraídas do Censo Escolar. Foram retiradas da 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. Entretanto, optou-se por mantê-las já que o método de decomposição do índice Theil-L permite trabalhar também com grupos contendo *missing values*, que serão indicados na categoria “não declarado”.

3.2. Metodologia: Decomposição do Índice Theil-L

Os índices de desigualdade já vêm sendo utilizados em estudos da área de educação, como em Felício (2004), Soares (2006) e Thomas et alii (2002). Para que se compreenda melhor quais fatores estão mais associados à desigualdade da distribuição de notas, a primeira escolha a ser feita diz respeito ao índice de desigualdade a ser utilizado. Optou-se pelo índice Theil-L pertencente à família das medidas de entropia generalizadas. Tais medidas possuem algumas propriedades interessantes que não são compartilhadas, por exemplo, com o coeficiente de Gini. Assim, enquanto as medidas de entropia generalizada podem ser *exatamente* decompostas de modo que se saiba o quanto cada grupo sociodemográfico contribui para a desigualdade total, o coeficiente de Gini não possui essa propriedade. Particularmente, se os alunos forem separados de acordo com atributos como sexo, cor, escola que frequenta etc, é possível saber qual desses fatores mais contribui para a desigualdade de proficiência sem haver qualquer resíduo da decomposição.

O índice de Theil-L pode ser definido como:

$$L = \frac{1}{N} \sum_{i=1}^N \ln \left(\frac{\bar{y}}{y_i} \right) \quad (1)$$

onde y_i refere-se à nota do aluno i ; N corresponde ao número total de alunos na amostra; \bar{y} corresponde à média das notas de todos os alunos e y_i é a nota do aluno i .

Se todos os alunos tivessem a mesma proficiência, não haveria desigualdade de notas e o índice seria zero. Na medida em que os alunos possuem notas diferentes com relação à média, a desigualdade medida pelo Theil-L tende a aumentar.

Uma vantagem do índice Theil-L em relação a outros índices é a sua decomponibilidade exata em dois componentes: desigualdade entre grupos (L_e) e desigualdade dentro dos grupos (L_h). Imagine que os alunos possam ser divididos em k grupos disjuntos. Seja π_h a proporção de alunos do h -ésimo grupo no total de alunos e y_h a proporção que o grupo h possui do somatório das notas de todos os alunos. Para exemplificar os conceitos de π_h e y_h , suponha que a amostra seja composta por sete alunos com as seguintes notas.

Quadro 1 – Exemplo fictício de distribuição de notas

Nota	3	1	2	6	10	10	8
Sexo	Homem	Homem	Homem	Homem	Mulher	Mulher	Mulher

Fonte: Elaboração própria.

Nesse caso, se os grupos forem divididos por sexo, $h = 1$ para homens e $h = 2$ para mulheres, $\pi_1 = 4/7$; $\pi_2 = 3/7$; $y_1 = 12/40$ e $y_2 = 28/40$.

Desse modo, pode-se definir a desigualdade entre grupos como:

$$L_e = \sum_{h=1}^k \pi_h \ln \frac{\pi_h}{y_h} \tag{2}$$

o que mede a desigualdade entre os h grupos formados a partir da amostra. Ela será tão maior quanto maior for a diferença entre o tamanho do grupo e a proporção das notas desse grupo. No exemplo do Quadro 1, as mulheres estão em menor proporção, mas possuem notas mais altas, respondendo por quase $\frac{3}{4}$ do total das notas.

O componente L_e corresponde a uma parte de L . A parte restante (L_h) se refere à parcela da desigualdade total que se deve a desigualdade de notas dentro dos grupos (no nosso exemplo, quanto da desigualdade está associada ao fato de que nem todos os homens (mulheres) tiraram as mesmas notas).

Portanto, pode-se escrever:

$$L_h = \sum_{h=1}^k \pi_h \frac{1}{n_h} \sum_{i=1}^{n_h} \ln \frac{\bar{y}_h}{y_{hi}} \tag{3}$$

onde n_h é o número de alunos do grupo h ; y_{hi} é a nota do aluno i pertencente ao grupo h e \bar{y}_h é a média de nota dos alunos do grupo h . Ou seja, a desigualdade dentro dos grupos corresponde à média ponderada do Theil-L calculada dentro de cada grupo, sendo o peso dado pelo tamanho relativo do grupo. Quanto maior L_h , menos as características que definiram a formação dos k grupos contribui para a desigualdade.

Combinando L_e e L_h , pode-se mostrar que o índice Theil-L é exatamente decomponível, tal que:

$$L = L_e + L_h \quad (4)$$

$$L = \sum_{h=1}^k \pi_h \ln \frac{\pi_h}{Y_h} + \sum_{h=1}^k \pi_h \frac{1}{n_h} \sum_{i=1}^{n_h} \ln \frac{\bar{y}_h}{y_{hi}} \quad (5)$$

Para que se possa analisar o quanto cada característica contribui para o total da desigualdade, são discutidos os conceitos de contribuição bruta e contribuição marginal, já utilizados por Ramos (2007) na decomposição da desigualdade dos rendimentos do trabalho. Se os alunos forem divididos de acordo com apenas uma característica, por exemplo, sexo, a razão L_h/L será denominada *contribuição bruta* da variável sexo para a desigualdade de notas, ou seja, corresponde à parcela da desigualdade de notas que ocorre porque homens e mulheres apresentam distribuições de notas distintas. Se a distribuição de notas fosse idêntica nos dois grupos, L_h seria zero e concluiria-se que o sexo não contribui para a desigualdade de notas. Assim, definimos a contribuição bruta de uma variável para a desigualdade de notas como:

$$CB = L_e/L \quad (6)$$

A divisão dos k grupos não precisa necessariamente ser realizada de acordo com apenas um atributo; pode ser feita a partir do uso de várias variáveis. Por exemplo, ao invés de se dividir a amostra somente por sexo, procede-se à divisão da amostra por sexo e raça. Se os indivíduos forem divididos em negros e não negros, são formados quatro grupos disjuntos ($k = 4$). É possível realizar a decomposição de L em L_e e L_h , o que resultaria respectivamente na contribuição de cada célula definida por raça e sexo para a desigualdade total (L_e/L).

Entretanto, a partir do exemplo acima, como observar quanto da desigualdade total se deve à desigualdade entre os diferentes sexos, somente? Para responder a esta pergunta, deve-se fazer a decomposição de L como anteriormente, obtendo L_e para o caso em que $k = 4$ (homem negro; homem não negro; mulher negra; mulher não negra). Em seguida, divide-se a amostra, apenas de acordo com a raça, e obtém-se L'_e para o caso em que $k = 2$ (negro; não negro).

A *contribuição marginal* da variável sexo para a desigualdade total será dada por:

$$CM = L_e - L'_e \quad (7)$$

A contribuição marginal permite identificar, portanto, qual a importância do sexo para a desigualdade total, já consideradas as diferenças de raça entre os alunos. Uma vantagem dessa abordagem é, por exemplo, que se pode identificar se após considerarmos as diferenças de *background* sócio-econômico dos alunos, ainda existe desigualdade entre os alunos negros e não negros. Se os alunos negros forem

preponderantes entre os de menor *background*, a contribuição bruta da variável raça iria mascarar as diferenças sócio-econômicas dos alunos negros e não-negros.

Por fim, cabem algumas observações relevantes quanto aos métodos aqui empregados. O uso do índice de desigualdade Theil-L permite a obtenção informações importantes relativas à contribuição de diferentes variáveis para a desigualdade observada em uma determinada variável de bem-estar dos indivíduos. Entretanto, é importante destacar algumas limitações dessa metodologia. Primeiramente, esse método não permite que se faça uma análise de causalidade que busque compreender qual o efeito causal de diferentes variáveis sobre a desigualdade de proficiência. O que se procura averiguar é quanto cada diferente variável considerada representa no total da desigualdade de notas. A segunda limitação da metodologia é que ela não estabelece uma distribuição de referência ideal, como o faz Soares (2006) e Soares e Marotta (2009), mas para o cálculo da desigualdade entre grupos, se admite que as notas dentro de cada grupo são iguais para todos do grupo.

Por fim, é importante ressaltar que fatores que estão presentes em praticamente todas as escolas não devem apresentar contribuição considerável para a desigualdade. Por exemplo, a existência (ou não) de paredes nas escolas pode representar parcela ínfima da desigualdade de proficiência, mesmo se for um determinante importante do nível de proficiência dos alunos.

3.3. Formação dos grupos para as decomposições

Uma questão importante no cálculo das contribuições bruta e marginal diz respeito à escolha das variáveis a partir das quais os grupos serão formados. No caso da contribuição marginal, a escolha de muitas variáveis para a definição dos grupos pode acabar gerando células vazias ou com poucas observações. Assim, uma primeira preocupação é evitar esse tipo de problema. Por exemplo, se os grupos forem definidos por sexo, raça, escolaridade do pai, escolaridade da mãe, se o professor corrige lição de casa, se a família do aluno recebe Bolsa-Família etc., é bem provável que não haja muitos alunos nas células formadas pela combinação das categorias dessas variáveis, principalmente se a base de dados for de tamanho reduzido.

Nesse sentido, a utilização de dados da Prova Brasil é extremamente adequada por se tratar de um banco de dados com muitas observações, permitindo gerar grupos definidos de acordo com mais variáveis. Por exemplo, enquanto o SAEB possui cerca de 50.000 observações para alunos da 4ª série para o Brasil inteiro, a Prova Brasil, possui mais de 350.000 observações de alunos da 4ª série apenas no Estado de São Paulo.

Mas, como escolher quais variáveis utilizar na montagem das células? Neste trabalho, adota-se o seguinte critério: primeiramente, realizou-se a contribuição bruta de todas as variáveis disponíveis e, em seguida, elegeu-se aquelas que apresentaram as maiores contribuições brutas (*CB*) para formarem os grupos

(células) a serem utilizadas na contribuição marginal.

A vantagem dessa escolha frente ao uso de índices sócio econômicos é que ela permite caracterizar como a adição de cada uma das variáveis separadamente contribui para a desigualdade de renda. Assim, se o interesse fosse apenas em “controlar” pelo nível sócio-econômico dos alunos, uma abordagem que utilizasse um índice sócio-econômico agregado poderia ser utilizada. No entanto, não teríamos o detalhamento acerca dos fatores mais associados à desigualdade de proficiência. Por outro lado, a desvantagem da escolha das variáveis como feita no presente artigo, ou seja, aquelas que apresentam as maiores contribuições marginais, reside exatamente no fato de ser um critério *ad-hoc*.

4. Análise Descritiva

Nesta seção, são apresentadas as estatísticas descritivas das variáveis da Prova Brasil e do Censo Escolar para alunos da 4^a série do Estado de São Paulo. Como já salientado, são reportados os resultados apenas daquelas variáveis que apresentaram as maiores contribuições brutas para a desigualdade de notas. Além das médias, é apresentada também a distribuição de cada característica entre os 10% melhores, 10% piores e os 10% em torno da mediana.

Comparando-se os resultados nas duas disciplinas, percebe-se que as notas de português têm média ligeiramente inferior nas diversas categorias analisadas. Já com relação à diferença de médias por categorias das variáveis, aplicando o teste *t* com intervalo de confiança de 95%, encontra-se que as médias são significativamente diferentes para todas as categorias.

As maiores médias encontradas foram para os alunos cuja mãe cursou o ensino superior. Por outro lado, os piores desempenhos são dos alunos com reprovação, que já abandonaram e com maior distorção idade-série (captada pela variável “idade dos alunos”). Se o histórico de reprovação dos alunos for observado, vê-se que alunos que já foram reprovados alguma vez são cerca de 40% entre os alunos do primeiro decil e apenas algo em torno de 5% entre os 10% melhores. Por outro lado, chama a atenção o fato de que entre os alunos com pior desempenho, cerca de 40% fazem sempre a lição, proporção que dobra entre os melhores.

A variável “se sente deixado de lado” levanta um ponto ainda pouco explorado pelas análises quantitativas: entre os piores alunos quase 60% dos alunos se sentem sempre excluídos no ambiente escolar, enquanto essa proporção cai para 25% entre os melhores alunos. Há também um aumento no número de alunos brancos e cujos pais cursaram ensino superior entre os melhores alunos, comparativamente aos de pior desempenho.

Se dividíssemos os alunos de melhor e pior desempenho em dois grupos, poderíamos observar o contraste em algumas características que chamam mais a atenção: o histórico de reprovação e abandono escolar, fazer a lição de casa, presença dos pais nas reuniões escolares, raça e a escolaridade dos pais.

No que tange as variáveis referentes às escolas, com base nos dados da Tabela

Tabela 1
 Estatísticas descritivas das variáveis da Prova Brasil

		Matemática				Português			
		Média	10	45-55	90	Média	10	45-55	90
Reprovação	Não	194,2	60,4	82,2	94,7	189,3	58,7	83,2	95,9
	Sim	166,4	39,6	17,8	5,3	156,8	41,3	16,8	4,1
Faz lição	Sempre	197	40,8	63,4	81,6	191,8	39,9	63,7	81,4
	Nem sempre	175,2	59,2	36,6	18,4	168,3	60,1	36,3	18,6
Participa do Bolsa Família	Não	192,3	68,2	80	89,4	187	67,4	79,6	90,6
	Sim	175,7	31,8	20	10,7	167,8	32,6	20,4	9,4
Número de pessoas em casa	De 3 a 5	193	61,6	75,1	85,9	187,6	61,1	75,3	86,6
	Outro	176,8	38,4	24,9	14,1	169,5	38,9	24,7	13,4
Se sente deixado de lado na escola	Sempre	195,5	42,7	58	73,5	189,9	42,4	58,8	72,4
	Nem Sempre	179,7	57,3	42,1	26,5	173,5	57,6	41,2	27,6
Idade	Até 10 anos	193,2	52	65,9	73,6	188,2	50,1	66,3	74,2
	11 anos	186,3	28,6	26,1	23,8	179,7	29,4	26,1	23,7
	12 ou mais	165,9	19,5	8,1	2,7	156,1	20,5	7,5	2,1
Abandono escolar	Sim	191,4	77,8	93,6	97,4	185,8	77,3	94,1	97,5
	Não	162,2	22,2	6,4	2,6	153,5	22,7	5,9	2,5
Raça	Não branco	183,1	66,6	58,9	40,5	176,9	66,9	59	41,5
	Branco	196,5	33,4	41,1	59,5	191	33,1	41	58,5
Escolaridade da mãe	Sem Faculdade	185,2	84,1	77,5	62,4	178,9	84,2	78,3	61
	Com Faculdade	200,6	15,9	22,5	37,6	196,1	15,8	21,7	39
Trabalho Infantil	Não	191,5	75,5	89,4	94,4	186,2	74,3	89,7	95,6
	Sim	170,3	24,5	10,6	5,6	160	25,7	10,3	4,4
Professor corrige lição de Matemática	Sempre	192,8	62,4	76,8	87	187	62,4	77,5	85,8
	Nem Sempre	176,4	37,7	23,2	13	170,2	37,6	22,5	14,2
Professor corrige lição de Português	Sempre	192,8	61	75,5	85	187,1	60,7	76	84,1
	Nem Sempre	177,4	39	24,5	15,1	171,1	39,3	24	15,9
Escolaridade do pai	Sem Faculdade	185,8	83,9	78,6	65	179,6	84,1	78,7	63,8
	Com Faculdade	199,3	16,1	21,4	35	194,7	15,9	21,3	36,2
Frequência dos pais em reuniões	Sempre	193,6	52,6	68,3	78,6	188,6	51,8	68	80,1
	Nem Sempre	179,1	47,4	31,7	21,4	171,6	48,2	32	19,9
Pretende estudar ou trabalhar depois do EF	Só Estudar	188,3	25,5	21	25	182,4	24,7	20,7	25,6
	Só Trabalhar	172,4	14	7,9	3,4	164,1	14,1	8,2	2,9
	Trab e Estud	195,8	27,8	38,6	48	190,8	28,3	38,6	48,9
	Não Sabe	185	32,7	32,6	23,6	178,7	32,9	32,6	22,7
Trabalho doméstico	Mais de 3h/dia	172,4	14,5	8,3	3,7	166,1	14,5	8,1	3,9
	Eventualmente	190,5	85,5	91,7	96,3	184,6	85,5	91,9	96,1
Número de quartos na casa	Até 1 Quarto	178,7	34,1	22,2	14,1	173,1	32,8	22	15,3
	Mais de 1 Quarto	191,9	66	77,8	85,9	186	67,2	78	84,8
TV	Tem	189,7	91,3	96,3	97,9	183,9	91,4	95,9	98,2
	Não Tem	171,6	8,7	3,7	2,1	163,2	8,6	4,1	1,8
Tem dicionário em casa	Tem	191,1	76,3	87,3	92,4	185,6	76,8	87,2	93,4
	Não Tem	174,6	23,7	12,7	7,6	166,9	23,2	12,8	6,6
Pais incentivam a fazer a lição	Sim	190,7	81,4	92,5	95,7	185,1	81,9	92,7	96,6
	Não	170	18,6	7,5	4,3	160,9	18,1	7,3	3,5
Sexo	Masculino	187,9	60,1	49,4	54,6	177,4	63,9	50,7	43,1
	Feminino	189,9	39,9	50,6	45,4	189,1	36,1	49,3	56,9

Fonte: Prova Brasil 2005.

2 pode-se notar que a variável rede é a única cuja diferença de médias não é significativa. Como se está trabalhando apenas com os dados de escolas públicas, isso indica que o desempenho médio de estudantes da rede municipal é semelhante aos alunos da rede estadual.

O grupo que apresentou as médias mais altas entre as variáveis escolares propriamente ditas foi o das escolas com mais de 5 vídeos por mil alunos. Assim como para as variáveis da primeira tabela, não há uma diferença significativa entre o desempenho em matemática e em português, exceto que as médias em português são sempre inferiores.

As variáveis de ciclos apontam algo interessante: embora as escolas que não adotaram o sistema de ciclos tenham médias superiores, ao compararmos estas escolas com as que adotaram um regime com dois ciclos iniciais com duração de dois anos, as médias são muito próximas. Por sua vez, famílias menores estão associadas a um melhor desempenho dos alunos.

Portanto, a análise descritiva permitiu associar que há uma correlação entre um melhor desempenho escolar dos alunos e o fato deles morarem em famílias menores, terem baixa distorção idade-série, reprovações menos e não terem abandonado a escola.

5. Resultados da Decomposição

Nesta seção é feita a apresentação dos resultados das decomposições feitas por meio do índice Theil-L. Primeiramente, serão apresentados os resultados das contribuições brutas, seguidos dos resultados das contribuições marginais.

5.1. Contribuições brutas

O valor do índice Theil-L para a desigualdade total é de 0,0263 para as notas de matemática e 0,0310 para o desempenho em português, o que está próximo do que foi encontrado por Felício (2004), 0,0331 e 0,0419, respectivamente. A diferença entre os números encontrados provavelmente se deve ao fato de a autora utilizar os dados do SAEB, que também contempla escolas particulares.

As variáveis analisadas neste trabalho estão em dois níveis, o nível dos alunos e das escolas. As variáveis do nível dos alunos são as que estão na Prova Brasil e referem-se aos alunos diretamente, enquanto que as outras são oriundas do Censo Escolar e se referem às escolas. Abaixo o resultado para as contribuições brutas das variáveis da Prova Brasil.

Para a implantação da contribuição bruta, a amostra foi dividida de acordo com a categoria de cada variável separadamente e, após isso, foi calculada a desigualdade entre os grupos. Esse exercício foi feito separadamente para cada uma das variáveis, sejam elas das escolas ou dos alunos.

Os valores das contribuições para português e matemática estão apresentados em termos percentuais da desigualdade total dos alunos nessas disciplinas. A variável

Tabela 2
Estatísticas descritivas das variáveis das Escolas

		Matemática				Português			
		Média	% dos centis			Média	% nos centis		
			10	45-55	90		10	45-55	90
Rede	Municipal	188,8	77,7	76,1	76,8	182,1	78,2	77,1	73,7
	Estadual	189,1	22,3	23,9	23,2	185,9	21,8	22,9	26,4
Ciclos: duração dos dois Primeiros Ciclos	2 e 2	195,8	11,9	16,4	21,8	188,8	12,3	17,1	19,8
	4 e 4	185,2	73,5	62,5	54	179,9	73,2	61,3	57,9
	Outro	194,5	14,6	21,1	24,3	187,6	14,6	21,6	22,3
Alunos por Sala	até 30	194,7	23,6	30,7	38,9	187,6	23,9	31,6	35,3
	de 30 a 40	186,6	73	66,7	59,6	181,2	72,7	65,9	63
	mais de 40	179,6	3,4	2,7	1,5	174,8	3,4	2,6	1,7
Ventiladores por Sala	Nenhum	181,6	27	21	13,9	176,5	26,8	20,4	15,8
	até 2	190,5	67,9	72,5	77,5	184,6	67,7	73,3	76,4
	mais que 2	195,3	5,1	6,4	8,6	187,4	5,5	6,4	7,8
Vídeos por 1000 alunos	Nenhuma	179,7	7,5	5,5	3,4	174	7,4	5,5	3,6
	até 5	188,7	87	87	85,6	183	86,7	86,9	86,6
	mais de 5	197,5	5,5	7,5	11	190	5,8	7,6	9,8
Tamanho em Alunos	até 500	194,6	12	15,93	20,06	187,1	12,21	16,47	17,49
	de 500 a 1000	192,88	36,77	43,46	51,73	187,06	36,96	43,94	50,84
	de 1000 a 1500	186,07	26,13	24,41	20,24	180,91	26,23	23,95	21,94
	mais de 1500	176,61	25,1	16,2	7,97	171,13	24,6	15,64	9,73
Fundamental Completo	Sim	181,4	47,73	36,08	24,94	175,67	48,04	34,88	27,46
	Não	193,06	52,27	63,92	75,06	187,11	51,96	65,12	72,54
Supletivo	Sim	184,82	55,76	46,93	39,34	178,56	55,89	47,32	39,84
	Não	192,54	44,24	53,07	60,66	187,01	44,11	52,68	60,16
Mimeógrafo por 100 professores	até 10	185,98	70,86	63,51	55,44	180,11	70,89	63,03	56,78
	de 10 a 25	193,36	24,04	29,98	35,44	187,35	24,17	30,21	34,19
	mais de 25	196,09	5,1	6,51	9,13	190,81	4,94	6,76	9,03
Parabólica para TV	Sim	191,75	56,47	63,52	71,97	185,84	56,47	64,05	70,3
	Não	183,85	43,53	36,48	28,03	178,04	43,53	35,95	29,7
% Alunos repetentes na série	Nenhum	187,59	39,36	38,17	35,97	181,7	39,17	38,03	36,15
	até 10%	192,14	36,13	40,61	46,45	186,53	35,99	41,12	46,38
	de 10 a 15%	188,25	11,32	11,37	10,89	182,07	11,59	11,19	10,8
TVs por 1000 alunos	Nenhuma	181,57	2,49	1,81	1,44	174,63	2,54	1,92	1,35
	Até 5	187,87	88,25	85,67	80,8	182,23	87,89	85,29	82,68
	Mais de 5	196,69	9,26	12,52	17,76	189,46	9,58	12,79	15,98

		Matemática				Português			
		% dos centis				% nos centis			
		Média	10	45-55	90	Média	10	45-55	90
% Alunos que abandonaram o curso na série	até 1%	189,74	83,35	85,69	90,01	183,84	83,28	85,71	89,42
	de 1 a 5%	183	14,6	12,69	8,98	177,32	14,72	12,61	9,46
	mais de 5%	180,95	2,05	1,62	1,01	175,74	1,99	1,68	1,12
Tamanho médio das famílias	até 4	208,3	1,6	2,9	6,7	203,2	1,6	2,8	6,4
	5	190	79,7	84,4	87	184,2	79,4	84,8	86,8
	mais de 5	176,6	18,7	12,7	6,3	169,9	19,1	12,5	6,9
% Alunos Negros ou Pardos na série	até 20%	201,1	7,7	11,2	20	195,5	7,7	11,4	19,5
	de 20 a 35%	192,8	31,2	36	42,5	186,8	31,9	36,4	41,7
	de 35 a 50%	184,1	42,6	37,4	28,3	178,3	42,2	36,8	29,1
	acima de 50%	179,3	17	13	7,1	173,6	16,7	13	7,8
	Não declarado	191,7	1,5	2,4	2,1	184,3	1,5	2,5	2
% Alunos com Idade Acima na série	até 5%	196,8	18,3	23,4	33,6	191,8	17,7	23,3	33,7
	de 5 a 10%	190,4	27,5	29,6	31,4	184,9	27,2	29,9	31,4
	de 10 a 15%	186,3	22,8	21,8	18,3	179,9	23,5	21,7	18,2
	mais de 15%	181,6	31,5	25,3	16,8	174,9	31,6	25,2	16,7
% Alunos Reprovados na série	até 5%	193,8	22	27	32,9	188,4	21,8	27,4	32,7
	de 5 a 10%	191,2	20,1	23,4	24,5	185,6	20,1	23,3	24,7
	de 10 a 15%	186,4	16,3	15,6	13,5	180,1	16,6	15,4	13,3
	mais de 15%	179,8	24,3	17,4	10,8	173,1	24,4	17	11,4
	Não declarado	189,5	17,4	16,6	18,2	183,5	17,2	16,9	17,9
TV Escola	Sim	193,8	33,1	40,2	50,1	187,9	33,3	40,4	48,7
	Não	185,5	66,9	59,8	49,9	179,7	66,7	59,6	51,3
Renda mínima	Sim	191,3	65,5	73,8	81,1	185,1	65,7	74,9	78,4
	Não	182,1	34,5	26,2	18,9	177	34,3	25,1	21,6

Fonte: Prova Brasil 2005 e Censo Escolar 2005.

que apresentou maior contribuição bruta foi reprovação, seguida das variáveis referentes a fazer a lição, participação no programa Bolsa Família, número de pessoas em casa e se o aluno se sente deixado de lado. Assim, se as notas de todos os alunos reprovados fossem iguais e o mesmo acontecesse entre os não reprovados, a desigualdade de notas seria 6,5% menor, no caso de matemática e 8,3% para português.

É interessante notar que as variáveis de comportamento escolar estão no topo da lista, seguidas de variáveis altamente correlacionadas com renda, o que indica que *background* socioeconômico parece ser um fator importante para determinação da desigualdade educacional.

A idade dos alunos, a raça e a escolaridade da mãe também se destacam entre aquelas de maior contribuição. Já uma variável aparentemente importante, mas

Tabela 3

Contribuições brutas das variáveis de *background* dos alunos

	Matemática Português	
Reprovação	6,60%	8,30%
Faz lição	6,20%	6,60%
Participa do Bolsa Família	4,50%	5,20%
Número de pessoas em casa	4,40%	5,30%
Se sente deixado de lado na escola	4,30%	4,30%
Idade	4,20%	5,20%
Abandono escolar	3,90%	4,40%
Raça	3,60%	3,60%
Escolaridade da mãe	3,50%	4,00%
Trabalho Infantil	3,20%	4,50%
Professor corrige lição de matemática	3,10%	3,10%
Professor corrige lição de português	3,10%	3,20%
Escolaridade do pai	3,00%	3,40%
Frequência dos pais em reuniões escolares	2,90%	3,60%
Pretende estudar ou trabalhar depois do EF	2,90%	3,40%
Trabalho doméstico	2,40%	2,40%
Número de quartos na casa	2,30%	2,10%
TV	2,10%	1,90%
Tem dicionário em casa	1,90%	2,30%
Pais incentivam a fazer a lição	1,90%	2,50%
Sexo	0,10%	1,60%

Fonte: Prova Brasil 2005.

pouco explorada na literatura, é se os professores corrigem a lição de casa.

O resultado das contribuições brutas das variáveis é, de maneira geral, mais significativo para a desigualdade em português do que em matemática. Isso é mais expressivo para a variável reprovação, em que a contribuição em português é dois pontos percentuais superior a de matemática. Na Tabela 4 são apresentados os resultados da contribuição bruta das variáveis referentes às escolas.

Comparativamente aos resultados obtidos para o *background* socioeconômico, as contribuições brutas das variáveis relativas às escolas são significativamente menores. As que se destacam são algumas variáveis referentes aos colegas: o

Tabela 4

Contribuições brutas das variáveis das Escolas

	Matemática Português	
% Alunos Negros ou Pardos na série	2,40%	2,10%
Tamanho médio das famílias	2,00%	2,00%
Alunos por Sala	1,70%	1,30%
% Alunos com Idade Acima na série	1,60%	1,80%
Tamanho em Alunos	1,40%	1,30%
Fundamental Completo	1,10%	0,80%
Ventiladores	1,10%	0,70%
% Alunos Reprovados na série	1,10%	1,20%
No.de Ciclos e Duração dos dois Primeiros Ciclos	0,70%	0,30%
TV Escola	0,70%	0,70%
Supletivo	0,70%	0,80%
Mimeógrafo por 100 professores	0,70%	0,60%
Parabólica para TV	0,60%	0,60%
Renda mínima	0,50%	0,30%
Vídeos por 1000 alunos	0,50%	0,30%
% Alunos Repetentes na série	0,50%	0,50%
TVs por 1000 alunos	0,40%	0,30%
% Alunos que Abandonaram o Curso na série	0,40%	0,30%
Ciclos	0,30%	0,20%
Rede	0,10%	0,20%

Fontes: Prova Brasil 2005 e Censo Escolar 2005.

percentual de alunos negros ou pardos, o tamanho médio das famílias dos colegas e o percentual de alunos acima da idade na série.

As variáveis relativas aos professores apresentaram contribuições brutas bastante baixas. Curiosamente, ao contrário do que aconteceu na tabela anterior, as contribuições das variáveis escolares apresentam geralmente contribuições menores em português.

A análise feita até agora se concentrou nas diversas características dos alunos e das escolas para explicar a desigualdade de proficiência. Uma análise alternativa consiste em se avaliar em que medida a desigualdade de notas seria reduzida caso a desigualdade de proficiência dentro da turma fosse eliminada. Desse modo, seria possível avaliar a contribuição bruta das turmas para a desigualdade. Esse

mesmo exercício pode ser feito eliminando-se a desigualdade dentro da escola. Por fim, pode-se repetir a análise para cálculo da contribuição bruta dos municípios para a desigualdade. Ou seja, em cada caso, seriam calculadas separadamente as contribuições brutas das turmas, escolas e municípios para a desigualdade de notas entre alunos.

Tabela 5

Contribuições brutas das variáveis das Escolas

	Matemática Português	
Turma	20,9%	19,4%
Escola	14,4%	13,0%
Município	5,1%	3,7%

Fonte: Prova Brasil 2005.

A variável turma é a que mais se relaciona à desigualdade de proficiência entre os alunos, ou seja, se fossem eliminadas as diferenças de notas dentro de todas as turmas, a desigualdade de notas diminuiria cerca de 20%. Entretanto a contribuição da variável turma deve ser olhada com cuidado. Apesar do tamanho médio das turmas ser de 27 alunos, o fato de haver turmas muito pequenas prejudica a inferência acerca da desigualdade dentro das turmas. Por sua vez, as diferenças de notas entre escolas correspondem a 13% e 14% da desigualdade total e os municípios sozinhos são responsáveis por 5 e 4%.

5.2. Contribuições marginais

A análise das contribuições brutas forneceu um primeiro retrato acerca de quais variáveis isoladamente estão mais relacionadas à desigualdade de notas. No entanto, cada uma das contribuições brutas pode “camuflar” a existência de outros fatores que estejam associados à desigualdade de notas. Por exemplo, o perfil da desigualdade dentro de cada grupo definido por posse ou não de Bolsa Família pode ser muito semelhante ao perfil definido de acordo com a escolaridade dos pais. Nesse sentido, é importante a realização da contribuição marginal no sentido de separar a importância de cada fator na desigualdade total.

Os resultados da Tabela 6 apresentam a contribuição marginal das variáveis de *background* socioeconômico dos alunos.¹ Nesse caso, estamos comparando o resultado de da decomposição em que a variável sob análise é considerada na divisão dos grupos com aquela em que tal variável não é considerada.

É interessante notar que fazer a lição é a variável mais importante, mesmo quando condicionamos para variáveis importantes como a escolaridade dos pais e outras variáveis de renda. A variável reprovação continua sendo importante, apesar de

¹ As variáveis que possuíam mais de uma categoria foram transformadas em binárias de modo a possibilitar a utilização de mais variáveis na formação das células.

sua contribuição marginal ser bem menor que a bruta. Ainda sobre as variáveis de caráter “mais comportamental” chama a atenção o fato de a variável “se sente deixado de lado” contribuir mais para a desigualdade do que a “escolaridade da mãe”, o que chama a atenção para um aspecto não-cognitivo que pode ser mais bem trabalhado pela escola e que contribuiria para a diminuição da desigualdade de proficiência entre os alunos.

Tabela 6

Contribuição marginal das variáveis de *background* dos Alunos

	Matemática Português	
Contribuição de todas as variáveis	22,94%	26,14%
Contribuição marginal de cada variável		
Faz a lição	2,50%	2,42%
Reprovação	2,45%	2,93%
Número de pessoas em casa	2,15%	2,34%
Se sente deixado de lado na escola	1,86%	1,66%
Trabalho infantil	1,77%	2,27%
Recebe Bolsa Família	1,71%	1,87%
Pretende trabalhar ou estudar depois do EF	1,56%	1,63%
Escolaridade da mãe	1,55%	1,56%
Raça	1,53%	1,45%
Abandono escolar	1,51%	1,52%
Número de quartos em casa	1,49%	1,30%
Frequência dos pais nas reuniões escolares	1,18%	1,32%
Idade	1,06%	1,05%

Fonte: Prova Brasil 2005.

Tomadas em conjunto, todas as variáveis de *background* explicam 23% (matemática) e 26% (português) da desigualdade entre os alunos. A contribuição das variáveis é maior em português. Nessa disciplina, as contribuições mais importantes são das variáveis reprovação e trabalho infantil. Ainda que as contribuições marginais individuais sejam baixas, em conjunto são importantes fatores associados à desigualdade de notas.

Diferentemente do trabalho de Albernaz et alii (2002) para quem 80% da diferença se deve ao *background* socioeconômico, neste trabalho a desigualdade devida a essas variáveis chega a cerca de 25%. Um dos motivos que pode explicar essa diferença de resultados pode ser que no trabalho acima citado, a

amostra também contempla escolas particulares o que acaba majorando o efeito do *background* sócio-econômico.

A Tabela 7 fornece os resultados da contribuição marginal das variáveis associadas às escolas. As variáveis “de escola” da Tabela 7 referem-se a insumos escolares, características dos professores e dos colegas de turma.

Tabela 7

Contribuição marginal das variáveis de Escola

	Matemática Português	
Contribuição de todas as variáveis	13,55%	12,13%
Contribuição marginal de cada variável		
Município	1,48%	1,27%
% Alunos Negros ou Pardos na série	0,30%	0,31%
TV Escola	0,22%	0,19%
Parabólica para TV	0,20%	0,19%
% Alunos Reprovados na série	0,19%	0,18%
Mimeógrafo por 100 professores	0,19%	0,15%
Supletivo	0,18%	0,19%
% Alunos com Idade Acima na série	0,17%	0,20%
% Alunos que Abandonaram o Curso na série	0,15%	0,16%
Ventiladores	0,15%	0,13%
Alunos por Sala	0,14%	0,13%
Tamanho médio das famílias	0,13%	0,14%
Renda mínima	0,13%	0,13%
Fundamental Completo	0,12%	0,13%
% Alunos Repetentes na série	0,11%	0,11%
Transporte de Alunos	0,09%	0,11%
O professor(a) Corrige a Lição	0,06%	0,08%
Tamanho em Alunos	0,03%	0,02%
TVs por 1000 alunos	0,02%	0,03%
Vídeos por 1000 alunos	0,01%	0,01%
Ciclos	0,01%	0,00%

Fontes: Prova Brasil 2005 e Censo Escolar 2005.

Conjuntamente, todas essas variáveis contribuem com 14% (matemática) e 12%

(português) da desigualdade de notas. Tomadas em conjunto, as variáveis têm contribuição importante, cerca de metade do impacto das variáveis de *background*. Mas se tomadas isoladamente, vemos que as contribuições marginais são muito baixas, destacando-se apenas a variável município. Ou seja, a interação de variáveis escolares é mais importante para compreensão da desigualdade de desempenho entre os alunos do que cada variável isoladamente.

Um exemplo disso é a variável referente ao tamanho da sala de aula. Nesse caso, a contribuição bruta mostrou-se bem maior que a contribuição marginal. Isto indica que existe uma correlação desta variável com outras variáveis escolares. Tendo em vista que em conjunto as variáveis escolares explicam bem mais do que qualquer uma delas isoladamente, isso pode significar que, embora o tamanho das salas possa ser pouco importante isoladamente, faz parte de um conjunto de fatores que tem uma relação relevante com a desigualdade de desempenho dos alunos.

Também foi calculada na Tabela 8 a contribuição marginal considerando-se conjuntamente grupos de variáveis de características das escolas na tentativa de investigar se algum tipo de insumo escolar seria mais importante na desigualdade de notas.

Tabela 8

Contribuição marginal das diferentes características das Escolas

	Matemática Português	
Contribuição de todas as variáveis	13,55%	12,13%
Contribuição marginal de cada variável		
Infraestrutura Escolar	2,24%	2,09%
Colegas	1,75%	1,79%
Município	1,48%	1,27%
Programas Governamentais	0,57%	0,52%
Professores	0,06%	0,08%

Fontes: Prova Brasil 2005 e Censo Escolar 2005.

Ao agruparmos as variáveis de escola em algumas categorias, aquelas ligadas a infraestrutura escolar, características dos colegas e ao município da escola representam de 1,3% a 2,3% da desigualdade de notas dos alunos. Assim, novamente não há evidências de que apenas um grupo de características ligadas ao ambiente escolar responda por parte considerável da desigualdade de desempenho.

6. Considerações Finais

O presente trabalho procurou identificar quais fatores estão mais associados à desigualdade de desempenho escolar dos alunos da 4ª série do ensino fundamental do estado de São Paulo. Foi utilizada a metodologia de decomposição do índice de

Theil-L por subgrupos populacionais e os resultados apontam que a desigualdade de *background* socioeconômico está relacionada a 25% da desigualdade total de proficiência. Embora não seja possível inferir uma relação de causalidade, podemos dizer que um quarto da desigualdade de notas está relacionada à desigualdade das condições de *background* familiar. Entretanto, nenhum fator individualmente se destaca: por exemplo, as diferenças entre alunos que fazem e os que não fazem lição de casa contribuem com cerca de 2% da desigualdade de notas.

As variáveis relacionadas ao ambiente escolar responderam por cerca de 13% da desigualdade total e, novamente, nenhum atributo escolar, dos colegas ou dos professores mostrou uma contribuição acima de 2% para a desigualdade de notas. No entanto, ficou claramente demonstrado que a desigualdade na distribuição de todo o conjunto desses atributos explica parte não desprezível da desigualdade de notas. Ou seja, a interação de variáveis escolares é mais importante para compreensão da desigualdade de desempenho entre os alunos do que cada variável isoladamente.

Apesar de não possuir escolas particulares na amostra como em outros trabalhos (Albernaz et alii 2002; Felício 2004; Soares 2006), este artigo encontrou uma contribuição importante da desigualdade entre as escolas para a desigualdade total, o que aponta que há espaço para políticas educacionais que diminuam a desigualdade de notas na rede pública de ensino. Diferenças entre municípios explicam cerca de 4% da desigualdade de notas. Esta é uma das variáveis que apresentou maior contribuição para a desigualdade e merece maior investigação futura. Como aqui nos limitamos aos dados de São Paulo, um estudo que incorpore outras regiões do país deve acentuar ainda mais essas diferenças. Um possível ponto de investigação futura consiste em melhor compreender os fatores por detrás dessa diferença dos municípios na desigualdade. Qual seria o papel de diferenças regionais, do nível de qualidade de vida e principalmente das políticas públicas municipais na área educacional como fatores indutores dessa desigualdade?

Embora nenhuma variável de *background* tenha contribuído individualmente para grande parcela da desigualdade, chama a atenção o fato de a variável indicativa de o aluno se “sentir deixado de lado” contribuir com 1,5% da desigualdade de notas. Assim, apesar dessa variável ser negligenciada em estudos sobre os determinantes do nível de desempenho médio dos alunos, ela se destacou como fator contributivo para a desigualdade, sendo mais importante, por exemplo, que escolaridade dos pais.

O fato de nenhuma variável ter se destacado sobre as demais na decomposição da desigualdade entre os alunos chama a atenção para o estudo da relação entre as variáveis, sobre sua interação para gerar a desigualdade de desempenho entre os alunos. Assim, é interessante a realização de uma investigação mais detalhada acerca da interação de diferentes insumos escolares e destes com variáveis de *background* para compreender a contribuição das escolas para a desigualdade de desempenho entre os alunos.

Referências bibliográficas

- Albernaz, A., Ferreira, F., & Franco, C. (2002). Qualidade e equidade na educação fundamental brasileira. Departamento de Economia – PUC Rio (Texto para Discussão 455).
- Barros, R. P., Mendonça, R., Santos, D. D., & Quintaes, G. (2001). Determinantes do desempenho educacional no Brasil. IPEA (Texto para Discussão 834).
- Coleman, J. S. (1966). *Equality of Educational Opportunity*. Office of Education, U. S., Washington D. C.
- Dewey, J., Husted, T. A., & Kenny, L. W. (2000). The ineffectiveness of school inputs: A product of misspecification? *Economics of Education Review*, 19(1):27–45.
- Felício, F. (2004). O efeito da qualidade da escola sobre o desempenho escolar: Uma avaliação do ensino fundamental do estado de São Paulo. Master's thesis, Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo.
- Hanushek, E. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3):1141–1177.
- Hanushek, E. & Luque, J. (2003). Efficiency and equity in schools around the world. *Economics of Education Review*, 22(5):481–582.
- Krueger, A. & Whitmore, D. (1999). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project star. *The Economic Journal*, 111(468):1–28.
- Ramos, L. (2007). A desigualdade de rendimentos do trabalho no período pós-real: O papel da escolaridade e do desemprego. *Economia Aplicada*, 11(2):281–301.
- Rios-Neto, E. L. G., César, C. C., & Riani, J. L. R. (2002). Estratificação escolar e progressão educacional por série no Brasil. *Pesquisa e Planejamento Econômico*, 32(3):395–415.
- Soares, J. (2006). Measuring cognitive achievement gaps and inequalities: The case of Brazil. *International Journal of Educational Research*, 45(3):176–187.
- Soares, J. & Andrade, R. (2006). Nível sócio-econômico, qualidade e equidade das escolas de Belo Horizonte. *Ensaio: Avaliação de Políticas Públicas Educacionais*, 14(50):107–126.
- Soares, J. & Candian, J. (2007). O efeito da escola básica brasileira: As evidências do PISA e do SAEB. *Revista Contemporânea de Educação*, 2(4).
- Soares, J. & Collares, A. (2006). Recursos familiares e desempenho cognitivo dos alunos do ensino básico brasileiro. *DADOS – Revista de Ciências Sociais*, 49(3):615–650.
- Soares, J. & Marotta, L. (2009). Desigualdade no sistema de ensino fundamental brasileiro. In Giambiagi, F., Henriques, R., Pessoa, S., & Veloso, F., editors, *Educação Básica no Brasil*, pages 73–91. Elsevier, Rio de Janeiro.
- Sprietsma, M. & Waltenberg, F. (2010). The effect of teachers wages on student achievement: Evidence from Brazil. In Backhaus, J., Eamets, R., & Sepp, J., editors, *Advances in the Economics of Education: On Markets and Post-Transformation Issues*. LIT-Verlag, Berlin.
- Thomas, V., Wang, Y., & Fan, X. (2002). *A New Dataset on Inequality in Education: Gini and Theil Indices of Schooling for 140 Countries 1960-2000*. Banco Mundial.