

Desigualdade de Oportunidades e o Papel da Escola na Formação de Habilidades

Igor Vieira Procópio – PPGEA/UFJF¹

Ricardo da Silva Freguglia – PPGEA/UFJF

Flávia Chein – PPGEA/UFJF

RESUMO

Este artigo investiga o papel da escola no processo de geração/redução das desigualdades de habilidades entre os indivíduos, com base na hipótese de que a desigualdade de habilidades em que as crianças iniciam o ensino fundamental limita o papel da escola na formação de habilidades. É utilizado um banco de dados longitudinal fornecido pelo projeto Geres, onde uma amostra de crianças é acompanhada durante os 4 primeiros anos do Ensino Fundamental. Especificamente, o presente estudo objetiva (i) medir a evolução da desigualdade e do efeito das escolas sobre as oportunidades na formação de habilidades e (ii) encontrar evidências empíricas para o Brasil do processo de dinâmica do aprendizado. O primeiro ponto é realizado comparando a evolução do aprendizado para grupos com características socioeconômicas distintas e através da decomposição da desigualdade de proficiência com a utilização do índice de Theil-T. Já no segundo ponto, utiliza-se o modelo da tecnologia de formação de habilidades (Cunha e Heckman, 2007) como base teórica e técnicas de painel dinâmico para as estimações. Os resultados encontrados indicam uma relação entre nível de habilidades e origem socioeconômica. Além disso, há evidências de um efeito dinâmico da aprendizagem, dado o impacto do nível de habilidades de períodos anteriores sobre o desempenho presente. Este resultado alerta para o papel limitado da escola regular na redução das desigualdades de habilidades e reforça a importância de investimentos na educação infantil.

Palavras chave: Desigualdade de Oportunidades; Formação de Habilidades; Efeito Escola.

Código JEL: I21, I24

ABSTRACT

This paper investigates the schools' role in the generation / reduction of inequalities of students' skills. The central assumption is that the initial level of skills inequality limits the effects of schools' intervention in skills formation. The data come from a longitudinal database provided by the project Geres, where a sample of children is monitored during the first 4 years of elementary school. Specifically, this study aims to (i) measure the evolution of inequality and the effect of schools on the opportunities in skills formation and (ii) find empirical evidence for Brazil' dynamic process of learning. The first point is performed by comparing the evolution of learning for groups with different socioeconomic characteristics and by the decomposition of inequality of proficiency using the Theil-T. While the second point uses the model of technology of skill formation (Cunha and Heckman, 2007) as a theoretical basis and techniques for dynamic panel. The results indicate a relationship between skill level and socioeconomic background and also show evidence of the dynamic effects of learning, highlighting the role of the skill level of prior periods. These findings shed some light on the limited role of the regular school in reducing inequalities of skills and reinforce the importance of investments in early childhood education.

Key words: Inequality of Opportunities; Skills Formation; School Effect.

Área ANPEC: 11 Economia Social e Demografia Econômica

¹Email para contato: igor.procopio@ufjf.edu.br

DESIGUALDADE DE OPORTUNIDADES E O PAPEL DA ESCOLA NA FORMAÇÃO DE HABILIDADES

1 – Introdução

O acesso à educação é considerado um dos aspectos fundamentais para o desenvolvimento individual e de uma nação. Indivíduos mais escolarizados têm, por exemplo, melhores oportunidades no mercado de trabalho, menor probabilidade de ficar desempregado e menor taxa de participação em crimes. Um país com uma população com alto nível de escolaridade apresenta maiores taxas de crescimento econômico, melhores taxas de inovação e facilidade de adaptação em relação a inovações introduzidas por outros países. Diversos estudos empíricos têm apontado evidências de que fatores pré-mercado de trabalho são determinantes de parcela significativa da desigualdade de rendimentos. As habilidades geradas antes da entrada do indivíduo no mercado de trabalho determinam a produtividade do indivíduo. Testes de proficiência realizados em crianças e adolescentes apresentam forte ligação com o sucesso futuro no mercado (Leibowitz, 1974; Murnane, Willett, e Levy, 1995; Neal e Johnson, 1996; Keane e Wolpin, 1997; Cameron e Heckman, 1998; Hanushek e Rivkin, 2006; Cunha e Heckman, 2007, 2009).

No Brasil destacam-se o trabalho seminal de Langoni (1973), e os trabalhos de Barros e Mendonça (1996) e Ferreira (2000), apontando a desigualdade educacional como o principal determinante da desigualdade salarial brasileira. Esses estudos trazem evidências de que a desigualdade de salários apresentada no mercado de trabalho é um reflexo da diferença de produtividade dos trabalhadores, e que esta diferença não é gerada no mercado de trabalho. Barros e Mendonça (1996) desenvolvem um modelo dividindo o ciclo de vida do trabalhador em dois períodos, um período pré-mercado de trabalho, onde as habilidades são formadas, e um segundo período correspondendo ao mercado de trabalho, onde as habilidades são transformadas em produtividade e remuneradas através dos salários. Os autores concluem que a maior parcela da desigualdade é gerada no primeiro período do ciclo de vida, ou seja, na fase de formação das habilidades. Portanto, os trabalhadores entram no mercado de trabalho em desigualdade de condições. A divisão proposta por Barros e Mendonça (1996) contribui para o entendimento da geração de desigualdades salariais, evidenciando o papel da desigualdade de habilidades.

Constata-se, pois, a relevância em se estudar políticas educacionais a partir da análise do nível de aprendizado das crianças. O objetivo deste artigo é investigar a dependência do nível de habilidades das crianças dentro da escola em relação às habilidades passadas e aos insumos escolares recebidos. Para isso, é utilizado um banco de dados pouco explorado na literatura, fornecido pelo projeto Geração escolar 2005 – GERES. Este é um banco longitudinal que acompanha crianças da 1ª a 4ª séries do ensino fundamental nos municípios de Belo Horizonte (MG), Rio de Janeiro (RJ), Campinas (SP), Campo Grande (MT) e Salvador (BA). Considerando a dinâmica do aprendizado e o caráter longitudinal dos dados, são utilizadas técnicas de painel dinâmico para a estimação dos parâmetros da função de produção educacional.

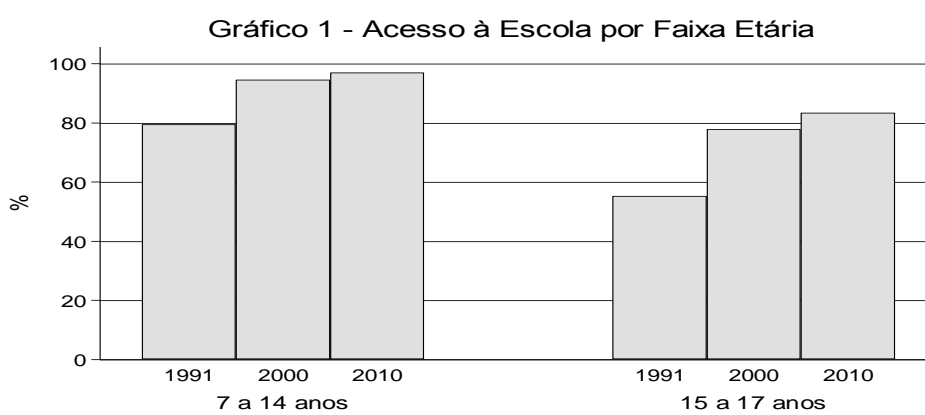
Os principais resultados encontrados apontam para uma desigualdade de oportunidades educacionais, evidenciando uma forte relação entre níveis de habilidades das crianças e sua origem social. Em relação à dinâmica do aprendizado, os resultados das estimações mostram a importância dos níveis de habilidade passados na formação das habilidades presentes.

Deve-se enfatizar que o presente artigo parte da premissa de que estudar os determinantes da formação de habilidades é essencial para a formulação de políticas públicas de redução da desigualdade de rendimentos. Os principais determinantes para a formação de habilidades são os fatores genéticos, as características familiares e os insumos escolares (Todd e Wolpin, 2003).

A eficácia dos insumos escolares é alvo de grande debate na literatura e grande divergência de opiniões e resultados encontrados nos trabalhos empíricos. O papel da escola na promoção da igualdade educacional tem como primeiro estudo o Relatório Coleman de 1966, que investigou se o desempenho

dos alunos norte americanos estava relacionado com características da escola. Os resultados deste estudo apontaram para a inexistência desta relação e que o principal determinante do aprendizado das crianças estava relacionado com o *background* familiar. No entanto, após este relatório, diversos estudos foram sendo desenvolvidos e trazendo evidências contrárias, ou seja, que a escola faz sim diferença (ver Soares e Brooke, 2008).

O acesso à escola no Brasil vem crescendo em todos os níveis de ensino. Apesar de na prática o acesso ao ensino fundamental ainda não ter atingido os 100% das crianças de 7 a 14 anos, existe a meta de universalização e garantia legal de acesso a todos que buscarem a matrícula. Nos outros níveis também vem ocorrendo um aumento no número de matrículas. O gráfico 1 exemplifica esta evolução com alguns dados. São apresentados os números da evolução do percentual de crianças de 7 a 14 anos na escola e dos adolescentes de 15 a 17 anos no período de 1991 a 2010. Para as crianças de 7 a 14 anos, nota-se que em 1991 praticamente 80% frequentavam a escola e que este percentual se aproxima dos 100% em 2010. Já para os adolescentes de 15 a 17 anos, apesar do percentual de matrícula ser inferior, a evolução no período foi mais acentuada, passando de aproximadamente 55% em 1991 para um pouco mais de 83% em 2010.



Fonte: Elaborado pelos autores a partir de Censo IBGE (1991,2000,2010)

No entanto, apesar do inegável avanço com o aumento nas taxas de matrícula em todos os níveis de ensino no Brasil, a qualidade da educação no Brasil ainda está longe dos países desenvolvidos. Resultados do Programa Internacional de Avaliação de Alunos – Pisa - mostram que o Brasil, apesar de ser um dos países que mais avançaram entre 2000 e 2009, ainda se encontra nas últimas posições no *ranking* dos países avaliados. Em 2009, o Brasil ocupava a 53ª posição em uma lista com 65 países.

O problema da qualidade da educação deve ser analisado sob dois aspectos, os valores médios da proficiência das crianças brasileiras são baixos na comparação internacional, e existe uma desigualdade no aprendizado relacionado às características familiares e regionais. À parcela da desigualdade relacionada com o ambiente em que a criança se desenvolve, dá-se o nome de desigualdade de oportunidades na formação de habilidades, que pode ser entendida como uma desigualdade de acesso aos fatores determinantes da formação de habilidades em função de características de origem familiar e local de residência. A Teoria da Igualdade de Oportunidades (Roemer, 1998) analisa o quanto da desigualdade de certos resultados individuais na vida socioeconômica está relacionado com características em que o indivíduo não tem controle, como origem familiar, e quanto às características sob o controle do indivíduo, como escolhas e esforço. A partir desta distinção, propõe-se políticas para reduzir a desigualdade até que as características fora do controle do indivíduo não determinem mais a desigualdade.

Para se estudar a formação de habilidades, este artigo é baseado no modelo desenvolvido por Cunha e Heckman (2007). Neste modelo, define-se a formação de habilidades através de um processo dinâmico, em que o nível de habilidades adquiridos em uma etapa serve como insumo para a formação de habilidades do período subsequente. Cunha e Heckman (2007 e 2009), Cunha *et al.* (2006) argumentam que desigualdades de habilidades entre as pessoas se iniciam logo no início do ciclo de vida e persistem.

Crianças que recebem baixos investimentos no início da vida apresentam dificuldades de aprendizado em etapas mais avançadas do ciclo de vida.

Logo, se as habilidades desenvolvidas no período escolar forem dependentes de habilidades passadas, ou seja, do nível de habilidades em que as crianças ingressam na escola, a escola não será capaz de diminuir o *gap* entre os alunos que iniciam com níveis de habilidades mais elevados para aqueles que iniciam com níveis mais baixos. Desta forma, as escolas, se fossem todas homogêneas em suas características que determinam o aprendizado, iriam revelar e perpetuar uma desigualdade gerada em um período pré-escolar. No entanto, as escolas não são homogêneas, existindo escolas que contribuem mais para o aprendizado das crianças do que outras. Se houver uma seleção não aleatória das escolas em relação às crianças de melhores níveis de habilidades, a escola poderá contribuir para um aumento da desigualdade de habilidades nas crianças caso as crianças mais habilidosas forem selecionadas para as escolas de melhor qualidade.

Além dessa seção introdutória, o artigo está estruturado em mais cinco seções. A segunda seção apresenta o modelo de formação de habilidades e a estratégia adotada para a estimação da função de produção educacional. A seção três traz uma descrição do banco de dados utilizado. A quarta seção apresenta as estatísticas básicas e alguns resultados preliminares, a seção cinco os resultados econométricos. A sexta seção conclui com as considerações finais.

2 - Modelo de Formação de Habilidades

Esta seção apresenta um modelo para a estimação da função de produção educacional com base na teoria do ciclo de vida de formação de habilidades, desenvolvida em Cunha e Heckman (2007). A principal característica deste modelo é incorporar que o aprendizado é um processo dinâmico, onde o nível de conhecimento em uma etapa serve como insumo para a formação de habilidades da etapa posterior. Os autores desenvolvem um modelo econômico para explicar as evidências encontradas na literatura de psicologia, educação e neurociência sobre o desenvolvimento de crianças.

O modelo desenvolvido permite explicar alguns fatos evidenciados na recente literatura empírica; i) as diferenças de habilidades entre indivíduos e entre grupos socioeconômicos se iniciam logo no início do ciclo de vida e persiste; ii) existe evidência de períodos críticos e sensíveis no desenvolvimento da criança; iii) apesar dos baixos retornos de intervenções voltadas para adolescentes em situação de desvantagem, a literatura empírica mostra altos retornos para investimentos compensatórios em crianças novas em situação de desvantagem.

A tecnologia da formação de habilidades é representada por uma função de produção que determina a relação entre os insumos e o nível de habilidades. A formação de habilidades é um processo dinâmico, onde o nível de habilidades de uma etapa serve como insumo para a formação de habilidades da próxima etapa. Os autores destacam a existência de períodos sensíveis e críticos. Um período é dito sensível para determinada habilidade se neste período o investimento tem um retorno maior do que em qualquer outro período. Um período é considerado crítico quando um investimento só exibe retorno neste período.

A equação definida em (1) representa a forma geral da tecnologia de formação de habilidades. A função f_t relaciona os insumos I_t (vetor de investimentos), u_0 (habilidade inata) e θ_t (habilidades no início do período), com o produto θ_{t+1} (habilidade ao final do período t). A função f_t pode assumir diferentes formas de acordo com o período t .

$$(1) \quad \theta_{t+1} = f_t(u_0, \theta_t, I_t)$$

Assume-se que f_t seja estritamente crescente e estritamente côncava em I_t e duas vezes continuamente diferenciável em todos os seus argumentos. A tecnologia (1) pode ser reescrita em forma recursiva, ou seja, em função dos investimentos até o período t conforme a equação (2).

$$(2) \quad \theta_{t+1} = m_t(u_0, I_1, \dots, I_T), t=1, \dots, T.$$

Para simplificar a exposição do modelo, consideram-se dois períodos na formação de habilidades da criança. Com dois períodos, a equação (2) passa a ser escrita conforme a equação (3).

$$(3) \quad \theta_3 = m_2(u_0, I_1, I_2)$$

No modelo desenvolvido por Cunha e Heckman (2007), o vetor de investimentos I_t pode conter qualquer investimento na formação de habilidades. Neste artigo, este vetor de investimentos é separado em investimentos direto da família e investimentos em insumos escolares. Ainda com base na simplificação do modelo em dois períodos, conforme apresentado na equação (3), considera-se que o primeiro período corresponde ao período antes da criança iniciar o ensino fundamental, e o segundo período corresponde ao período escolar². Portanto, no primeiro período o investimento na formação de habilidades é composto apenas por investimentos diretos da família, ou seja, $I_1 = F_1$, enquanto que no segundo período ocorrem investimentos tanto da família quanto da escola, $I_2 = (S_2, F_2)$. Com isso, a equação (3) passa a ser escrita conforme a equação (4).

$$(4) \quad \theta_3 = m_2(u_0, F_1, F_2, S_2)$$

Ou considerando os insumos correntes:

$$(4') \quad \theta_3 = f_2(u_0, \theta_2, F_2, S_2)$$

Cunha e Heckman (2007) destacam duas características da tecnologia de produção. A complementaridade dinâmica é a primeira das características, que ocorre quando $\partial^2 f_t(u_0, \theta_t, I_t) / \partial \theta_t \partial I_t > 0$, ou seja, quando o estoque de habilidades adquiridos no período anterior aumenta a produtividade do investimento no período presente. A segunda característica é a auto-produtividade, que ocorre quando $\partial f_t(u_0, \theta_t, I_t) / \partial \theta_t > 0$, ou seja, o nível de habilidades de um período contribui para a formação de habilidades do período seguinte. Atuando juntas, a complementaridade dinâmica e a auto-produtividade produzem efeitos multiplicadores. Esse é o mecanismo pelo qual habilidades produzem habilidades. Estes aspectos implicam um *trade-off* entre equidade e eficiência somente para investimentos tardiamente realizados em crianças, mas não para investimentos no início do ciclo de vida.

Se a complementaridade dinâmica e a auto-produtividade forem verificadas na prática, a formação de habilidades no período escolar (segundo período) irá depender do nível de habilidades do período anterior, ou seja, havendo uma desigualdade de habilidades entre as crianças no momento em que se inicia o período escolar, os investimentos em insumos escolares só serão capazes de reduzir a desigualdade se for abandonada a busca pela eficiência.

3 – Base de Dados

Neste artigo utiliza-se a base de dados proveniente do Projeto Geres – Geração Escolar – 2005. Este projeto acompanha alunos de uma amostra de escolas durante os quatro primeiros anos do Ensino Fundamental. Portanto, é um estudo longitudinal onde é possível acompanhar a evolução do nível de habilidades das crianças através do nível de proficiência, ou seja, é possível acompanhar o aprendizado da criança ao longo do tempo. Esta é a primeira pesquisa, no Brasil, a acompanhar uma amostra de alunos durante a primeira fase do Ensino Fundamental, logo, a análise de seus resultados é uma contribuição para a literatura sobre avaliação educacional no Brasil.

Os objetivos do Geres são: (i) identificar as características escolares que maximizam a aprendizagem dos alunos e que minimizam o impacto da origem social sobre o aprendizado; (ii)

² No Brasil a matrícula da criança na escola só é obrigatória a partir do Ensino Fundamental.

identificar os fatores escolares que diminuem a probabilidade de repetência dos alunos; e (iii) identificar aquelas características da escola que reduzem a probabilidade do absenteísmo.

O Geres acompanhou alunos de 2005 a 2008, e foram aplicados testes anuais para estimar os níveis de proficiência em matemática e português. Vale ressaltar que em 2005 foram aplicados dois testes, um no início do ano, para ser utilizado como diagnóstico, e outro no final do ano. A aplicação dos testes, portanto, corresponde a um total de cinco ondas. Além dos testes, foram aplicados questionários para as famílias, escolas, professores e diretores. Os questionários aplicados às famílias e às próprias crianças coletaram informações sobre o nível socioeconômico e características individuais das crianças, os questionários para as escolas buscavam obter informações sobre características das escolas (infra estrutura), dos professores, diretores e turmas.

Apesar do caráter longitudinal do banco de dados obtido pelo Geres, os questionários não foram aplicados em todos os anos, com exceção do questionário de professores. Em função disso, não é possível acompanhar a evolução das características das escolas e nem do nível socioeconômico das famílias. Esta deficiência do banco de dados prejudica o acompanhamento da evolução dos investimentos, tanto em insumos escolares quanto em insumos familiares, nas crianças.

Visando minimizar esta deficiência, buscaram-se informações sobre as escolas no Censo Escolar, realizado anualmente pelo Instituto Nacional de Estudos e Pesquisas Anísio Teixeira – Inep. Também foi utilizada a RAIS-Migra³ para obtenção do salário médio dos professores por escola.

O acompanhamento longitudinal foi realizado para os alunos que em 2005 estavam matriculados na primeira série do Ensino Fundamental. As escolas da amostra estão localizadas em cinco grandes cidades brasileiras – Belo Horizonte (MG), Rio de Janeiro (RJ), Campo Grande (MS), Salvador (BA) e Campinas (SP). O universo amostral estabelecido para o Projeto Geres consta dos alunos das turmas regulares do período diurno das escolas urbanas públicas com 20 ou mais alunos de 2ª série em 2003 e das escolas particulares com 10 ou mais alunos e com no máximo 3 turmas de 2ª série também em 2003, localizadas nos municípios de Belo Horizonte, Campinas, Campo Grande, Rio de Janeiro e Salvador.

4 - Estatísticas Descritivas

Esta seção apresenta as estatísticas básicas das variáveis utilizadas na função de produção educacional e alguns fatos estilizados sobre a evolução do aprendizado das crianças da amostra durante o período do projeto Geres. Também são apontadas evidências da relação entre esta evolução e algumas características dos indivíduos.

As tabelas 5 e 6 apresentam a média, desvio padrão mínimo e máximo das variáveis definidas na seção 3. As informações de proficiência representam valores em relação à escala de proficiência do Geres. O nível socioeconômico é uma variável fornecida junto com o banco de dados e é formada por informações sobre a escolaridade da mãe, a ocupação do pai e presença de bens no domicílio⁴. O salário dos professores está em valores reais de 2008. As demais variáveis são *dummies* e, portanto o valor médio representa o percentual de cada categoria.

A tabela 5 apresenta as estatísticas das variáveis invariantes no tempo. Nesta categoria encontram-se as variáveis de características das crianças e nível socioeconômico da família e informação do local de residência. As crianças da amostra são em sua maioria não brancas e em relação ao gênero existe um balanceamento. O valor médio da variável de nível socioeconômico não apresenta muito significado, mas o desvio padrão mostra que existe uma grande variação entre as famílias. Em relação ao município de residência, 30% das crianças moravam em Campinas, 27% em Belo Horizonte, 27% no Rio de Janeiro e 15% em Campo Grande.

A tabela 6 apresenta as estatísticas para as variáveis não constantes ao longo do tempo. São as informações de escolas e professores, além da variável de proficiência dos alunos. Para as informações

³ Banco de dados administrativo do Ministério do Trabalho e Emprego com características longitudinais dos trabalhadores formalmente empregados no mercado de trabalho brasileiro.

⁴ Para detalhes do procedimento de construção ver Andrade e Miranda (2004).

variantes no tempo, a variabilidade (desvio padrão) geral pode ser decomposta em uma parcela entre indivíduos e uma parcela intra indivíduos ao longo do tempo. A última coluna da tabela apresenta o percentual da variação geral atribuída a cada parcela. Esta informação sobre a variação entre e principalmente da variação intra é utilizada para a identificação dos modelos estimados no artigo.

Tabela 5 – Estatísticas Básicas das Variáveis Invariantes no Tempo

Variáveis	Observações	Média	Desvio Padrão	Mínimo	Máximo
Características das crianças					
Nse (nível socioeconômico)	43025	0.11	0.62	-1.45	1.66
Branco	43080	0.34	0.47	0	1
Não branco	43080	0.66	0.47	0	1
Raça não respondida	43080	0.01	0.11	0	1
Meninos	43080	0.49	0.50	0	1
Meninas	43080	0.49	0.50	0	1
Gênero não respondido	43080	0.02	0.12	0	1
Municípios					
Belo Horizonte	43080	0.29	0.45	0	1
Campinas	43080	0.30	0.46	0	1
Campo Grande	43080	0.15	0.35	0	1
Rio de Janeiro	43080	0.27	0.45	0	1

Fonte: Elaborado pelos autores a partir de Geres, Censo Escolar, RAIS-Migra (2005-2008)

Nota-se que o valor médio e o desvio padrão da proficiência em matemática são superiores aos da proficiência em português. O salário dos professores apresenta apenas 10% de variação ao longo dos anos, mas um alto desvio padrão entre os indivíduos, de mais de 50% do valor médio.

Os gráficos de 2 a 8 apresentam a evolução média da proficiência em matemática e em português, dos alunos do GERES ao longo dos quatro anos do primeiro ciclo do ensino fundamental, considerando também a avaliação diagnóstica. O gráfico 2 apresenta o resultado para todos os alunos da amostra utilizada de forma agregada, enquanto que os demais gráficos apresentam a evolução separada para determinados grupos. A divisão da evolução do nível de habilidades por grupos busca apresentar evidências que fatores fora do controle da criança influenciam no seu aprendizado e, portanto, que há uma desigualdade de oportunidades na formação de habilidades.

Analisando o gráfico 2 percebe-se uma diferença no comportamento das duas curvas. Enquanto o crescimento da proficiência em matemática se dá a taxas crescentes, o crescimento da proficiência em português ocorre a taxas decrescentes. As crianças entram na escola com praticamente o mesmo nível de habilidades nas duas disciplinas – aproximadamente 100 pontos na escala de proficiência do Geres. No entanto, após quatro anos, a média de proficiência em matemática passa a ser de quase 250 pontos, enquanto que a de português fica em torno de 170 pontos. Em função desta diferença entre a evolução das duas proficiências, torna-se importante a análise separada para a proficiência em matemática e para a proficiência em português.

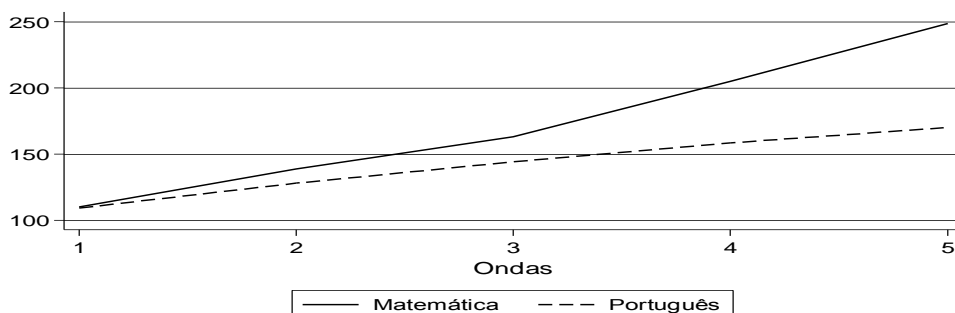
Os gráficos 3 e 4 apresentam a evolução das proficiências em matemática e português, respectivamente, dividindo-se as crianças em brancas e não brancas. As curvas para cada raça apresentam um formato semelhante. Para matemática, a distância entre os grupos aumenta ao longo dos anos, ou seja, a diferença entre o nível de conhecimento dos brancos vai aumentando em relação ao dos não brancos. Já para português a distância entre os dois grupos permanece praticamente a mesma ao longo dos anos, sendo que em todo o período o grupo dos brancos possui média de proficiência mais elevada. Este resultado indica que a escola, além de não estar sendo capaz de reduzir a desigualdade entre as raças no conhecimento em português, está contribuindo para um aumento destas desigualdades nas habilidades de matemática.

Tabela 6 – Estatísticas Descritivas para as Variáveis Variantes no Tempo

Variáveis		Média	Desvio Padrão	Mínimo	Máximo	% de variação
	geral	173.02	70.91	15.31	396.55	
Proficiência em matemática	entre		44.46	44.88	281.95	0.39
	intra		55.24	-0.96	342.86	0.61
	geral	141.92	33.37	47.80	222.96	
Proficiência em português	entre		22.84	79.35	196.09	0.47
	intra		24.33	39.15	213.17	0.53
	geral	2,157.63	1,259.40	0.00	6,731.06	
Salário do professor	entre		1,196.95	0.00	5,120.12	0.90
	intra		391.82	-264.71	5,645.58	0.10
	geral	0.02	0.13	0.00	1.00	
<i>Dummy</i> de informação de salário	entre		0.11	0.00	0.80	0.76
	intra		0.06	-0.78	0.82	0.24
	geral	0.70	0.46	0.00	1.00	
Professor com ensino superior (<i>dummy</i>)	entre		0.26	0.00	1.00	0.32
	intra		0.38	-0.05	1.45	0.68
	geral	0.44	0.50	0.00	1.00	
Experiência do professor (<i>dummy</i> – mais de 5 anos)	entre		0.30	0.00	1.00	0.38
	Intra		0.39	-0.31	1.19	0.62
	geral	0.08	0.27	0.00	1.00	
Escolas Especiais	entre		0.27	0.00	1.00	1.00
	Intra		0.01	-0.42	0.75	0.00
	geral	0.24	0.43	0.00	1.00	
Escolas Estaduais	entre		0.43	0.00	1.00	0.99
	Intra		0.04	-0.51	0.99	0.01
	geral	0.51	0.50	0.00	1.00	
Escolas Municipais	entre		0.50	0.00	1.00	0.99
	Intra		0.04	-0.24	1.26	0.01
	geral	0.17	0.37	0.00	1.00	
Escolas Privadas	entre		0.37	0.00	1.00	1.00
	Intra		0.03	-0.58	0.92	0.00
	geral	0.73	0.45	0.00	1.00	
Laboratório de informática (<i>dummy</i>)	entre		0.36	0.00	1.00	0.67
	intra		0.26	-0.02	1.48	0.33
	geral	0.29	0.45	0.00	1.00	
Laboratório de ciências (<i>dummy</i>)	entre		0.43	0.00	1.00	0.92
	intra		0.13	-0.46	1.04	0.08
	geral	0.74	0.44	0.00	1.00	
Biblioteca (<i>dummy</i>)	entre		0.33	0.00	1.00	0.57
	intra		0.29	-0.01	1.49	0.44
	geral	2.18	1.15	1.00	5.00	
Série do aluno	entre		0.17	1.20	3.20	0.02
	intra		1.14	0.38	4.43	0.98

Fonte: Elaborado pelos autores a partir de Geres, Censo Escolar, RAIS-Migra (2005-2008)

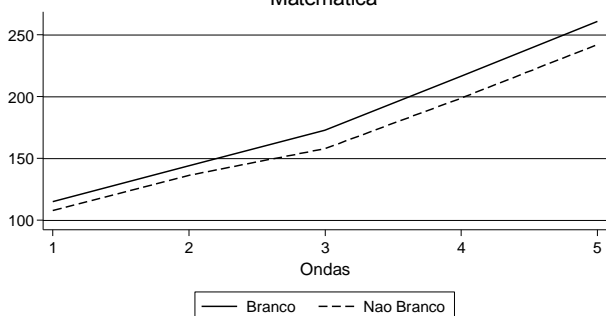
Gráfico 2 - Evolução da Proficiência Média dos Alunos do Geres



Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

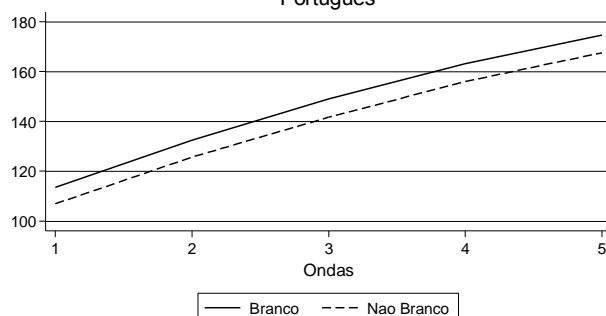
Os gráficos 5 e 6 apresentam a evolução dos níveis de proficiência para cada quintil de nível socioeconômico. Da mesma forma como acontece com a raça, o formato das curvas é semelhante para cada quintil e a distância entre as curvas na proficiência em matemática vai se ampliando, ou seja, a desigualdade entre os grupos socioeconômicos se amplia ao longo dos primeiros anos do Ensino Fundamental.

Gráfico 3 - Evolução da Proficiência Média por Raça Matemática



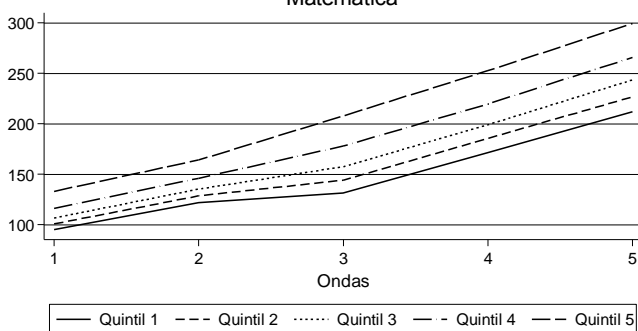
Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

Gráfico 4 - Evolução da Proficiência Média por Raça Português



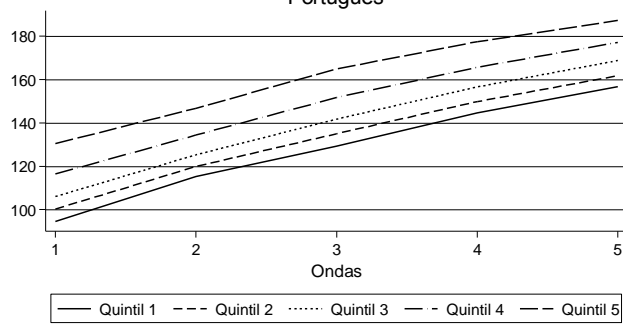
Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

Gráfico 5 - Evolução da Proficiência Média por Nível Socioeconômico Matemática



Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

Gráfico 6 - Evolução da Proficiência Média por Nível Socioeconômico Português



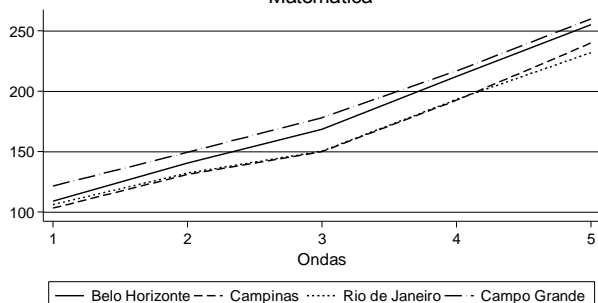
Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

Os gráficos 7 e 8 apresentam a evolução da proficiência média por município em matemática e português, respectivamente. Destaque para os municípios de Belo Horizonte e Campo Grande que se iniciam praticamente com o mesmo nível de proficiência, no entanto, na última onda, esta distância é visualmente perceptível, ou seja, as escolas de Belo Horizonte apresentam um maior nível de aprendizado do que as escolas de Campo Grande, este efeito é mais acentuado na proficiência em matemática. O

município do Rio de Janeiro inicia com maiores níveis de proficiência e se mantém assim em todas as ondas, tanto para português quanto matemática. Campinas apresenta o pior resultado.

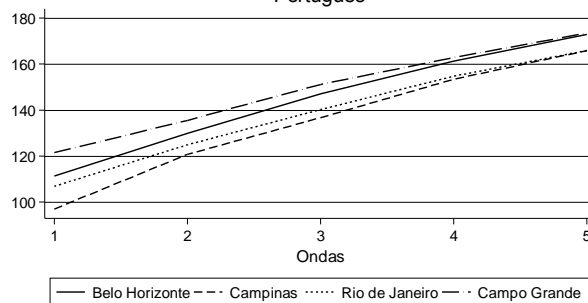
Os gráficos 3 a 8 apresentam evidências de que a raça da criança, o nível socioeconômico da família e o município de residência tem relação com o aprendizado das crianças e, portanto, são características que contribuem para a desigualdade de habilidades. Para dar robustez a esses resultados, é apresentada, na sequência, uma análise da decomposição do índice de desigualdade de Theil-T para quantificar a contribuição destas características na desigualdade de habilidades.

Gráfico 7 - Evolução da Proficiência Média por Município Matemática



Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

Gráfico 8 - Evolução da Proficiência Média por Município Português



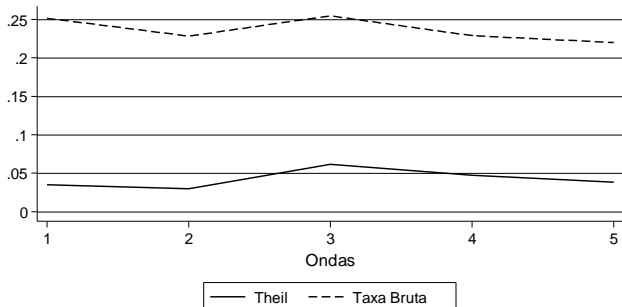
Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

O índice de Theil permite mensurar a desigualdade e decompô-la em parcelas referentes à desigualdade entre determinados grupos e dentro destes grupos. À contribuição da desigualdade entre grupos dá-se o nome de taxa bruta, que indica em termos percentuais o quanto da desigualdade total pode ser atribuída à característica que se está controlando. Os grupos podem ser formados também por combinação de duas ou mais características. Neste caso, a taxa bruta indica a contribuição conjunta destas características na desigualdade.

Os gráficos 9 e 10 apresentam a evolução do índice de Theil e da taxa bruta conjunta para a raça da criança, o nível socioeconômico e o município de residência. Nota-se que, na proficiência em matemática a desigualdade aumenta ligeiramente, mas a taxa bruta diminui, ou seja, a escola está contribuindo para o aumento da desigualdade, mas reduzindo o efeito da origem familiar. Já para a proficiência em português, tanto o índice de Theil quanto a taxa bruta se reduzem, ou seja, a desigualdade de habilidades está se reduzindo depois do início do Ensino Fundamental e também o efeito das circunstâncias, consideradas conjuntamente.

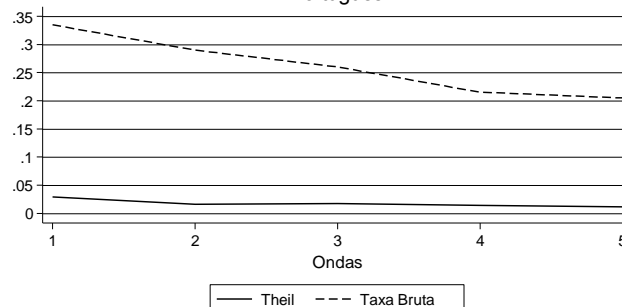
Como a taxa bruta só fornece informações sobre o efeito agregado dos três fatores que estão sendo controlados, calcula-se também a taxa marginal, que mede o efeito isolado de cada variável. Os gráficos 11 e 12 apresentam a evolução da contribuição marginal de cada variável. Percebe-se que para a proficiência em matemática apenas o efeito do município é decrescente. Ou seja, a redução na taxa bruta foi puxada pela redução do efeito do município na desigualdade. As outras duas variáveis apresentam crescimento de sua contribuição. O efeito da desigualdade entre as raças cresce de forma monotônica, enquanto a contribuição do nível socioeconômico oscila, mas termina a série com valores próximos aos valores iniciais. Estes resultados corroboram aqueles apresentados nos gráficos de 3 a 8, em que já era indicada que as desigualdades entre as raças e nível socioeconômico estavam aumentando enquanto se reduzia a desigualdade entre os municípios. Para a proficiência em português a contribuição do nível socioeconômico também se reduz, indicando que para esta habilidade o período escolar ameniza os efeitos da origem social.

Gráfico 9 - Evolução do Índice de Theil e Taxa Bruta Matemática



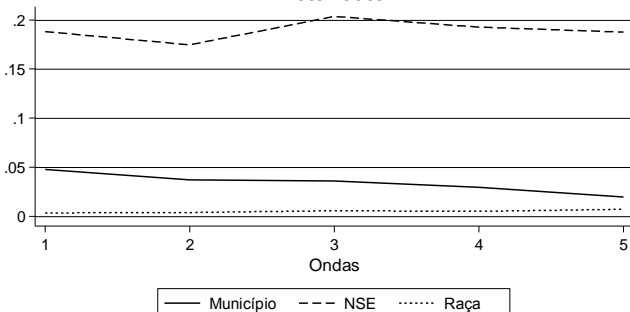
Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

Gráfico 10 - Evolução do Índice de Theil e Taxa Bruta Português



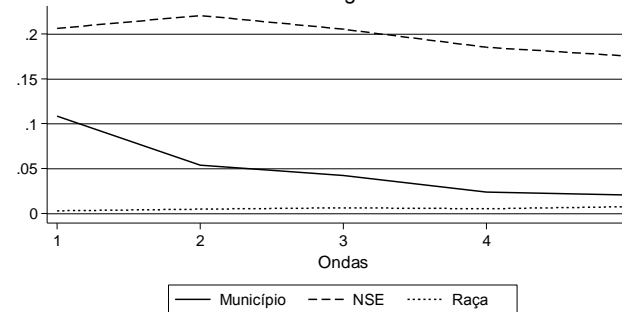
Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

Gráfico 11 - Evolução da Taxa Marginal Matemática



Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

Gráfico 12 - Evolução da Taxa Marginal Português

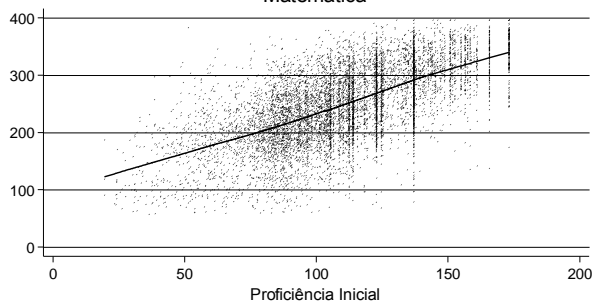


Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

De acordo com a teoria do ciclo de vida da formação de habilidades, o estoque de habilidades ao final de um período serve com insumo da função de produção de habilidades do período seguinte. Este estoque de habilidades atua em dois sentidos, diretamente na produção de habilidades do período seguinte e indiretamente através do aumento da produtividade dos investimentos. Os gráficos 13 e 14 apresentam o comportamento das habilidades em matemática e português, respectivamente, no último teste como função do nível de habilidades no início do período escolar. Nos dois casos a proficiência final é uma função crescente em relação à proficiência inicial. No entanto, a proficiência em matemática cresce a taxas maiores do que a proficiência em português, o que indica uma maior persistência do conhecimento em matemática.

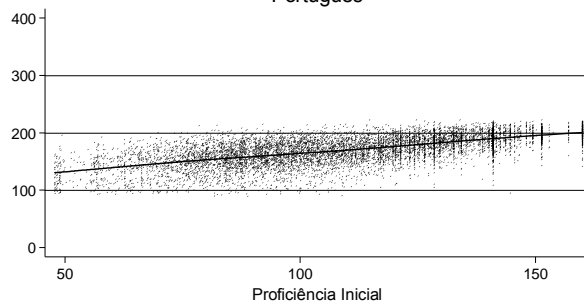
Os resultados apresentados nesta seção dão uma ideia do comportamento da formação de habilidades em português e matemática. Algumas características são parecidas para as duas disciplinas. Em ambas, a desigualdade no início do período escolar obrigatório é associada às características familiares, o que indica uma desigualdade de oportunidades na formação destas habilidades. A evolução dos índices de desigualdade mostra uma redução na desigualdade de proficiência em português para as crianças durante os quatro anos analisados e um aumento da desigualdade na proficiência em matemática. Por fim, mostra-se que a proficiência em matemática apresenta uma dependência maior em relação aos níveis de proficiência passados do que a proficiência em português.

Gráfico 13 - Lowess - Proficiência Final X Proficiência Inicial Matemática



Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

Gráfico 14 - Lowess - Proficiência Final X Proficiência Inicial Português



Fonte: Elaborado pelos autores a partir dos dados do Geres (2005-2008)

5 – Estratégia Empírica

Considerando o modelo de Cunha e Heckman (2007), Todd e Wolpin (2003) apresentam diversas formas de se estimar a função de produção educacional de acordo com a disponibilidade de informações no banco de dados. Os principais desafios para a estimação da função de produção educacional são: i) a ausência de bancos de dados contendo toda a história de investimentos familiares e escolares; ii) a habilidade genética específica de cada criança não é observada; iii) os resultados de testes padronizados (proficiência) são uma medida imprecisa do nível de conhecimento das crianças.

Assim como Andrabi *et al.* (2011), que partem da abordagem de Todd e Wolpin (2003) e incorporam técnicas de painel dinâmico para a análise do aprendizado de crianças no Paquistão, o presente artigo aplica tais técnicas para a estimação dos parâmetros da função de produção educacional para os dados do projeto Geres.

A estimação do modelo seguirá uma especificação linear, conforme Todd e Wolpin (2003, 2007), e a função f_t será a mesma para todos os períodos. Deste em ponto em diante, o nível de habilidades no final do período t será representado por y^*_t , em substituição a θ_{t+1} . Esta alteração de notação facilita a exposição da estratégia quando se utiliza variáveis defasadas. Com esta notação, y^*_t representa o nível de habilidades ao final do período t , enquanto X_t representa os insumos investidos durante o período t . Para simplificar a notação os insumos escolares e familiares são incluídos em um único vetor $X_{it} = (F_{it}, S_{it})$.

A equação (5) apresenta o modelo linear na forma de erro.

$$(5) \quad y^*_{it} = X_{it}\beta_1 + X_{it-1}\beta_2 + \dots + X_{it}\beta_t + \sum_{t=1}^T \gamma_t u_{it0} + \sum_{t=1}^T v_{it}$$

onde y^*_{it} representa o verdadeiro nível de conhecimento, sem erros de medida, γ_t representa o efeito da habilidade fixa não observada u_{it0} , e v_{it} um choque de produtividade. Esta especificação implicitamente impõe uma restrição aos parâmetros, assume que um insumo aplicado em um período tem o mesmo efeito no aprendizado do que o mesmo insumo aplicado em outro período⁵.

Os três desafios para a estimação dos parâmetros da função de produção são abordados na sequência. O primeiro é a ausência de banco de dados com toda a história de investimentos, portanto, em um primeiro momento, assume-se que a habilidade não observada não é correlacionada com os insumos e que não há erro de medida.

A estratégia mais simples seria considerar que os insumos passados não tem efeito no nível presente de habilidades. Esta estratégia é conhecida como especificação contemporânea e forma funcional ficaria de acordo com a equação (6), ou na forma de erro composto conforme (7).

$$(6) \quad y^*_i = X_i\beta + \gamma u_{i0} + v_i$$

⁵ Ver Todd e Wolpin (2003) para uma forma mais flexível.

$$(7) \quad y^*_i = X_i\beta + n_i, \quad n_i = \gamma u_{i0} + v_i$$

Para que as estimativas de (7) por OLS fossem consistentes, ter-se-ia que assumir que os insumos passados não determinam o aprendizado presente ou que os insumos passados não são correlacionados com os insumos contemporâneos, e que a habilidade inata não seja correlacionada com o nível de insumos. De acordo com o modelo da formação de habilidades toda a história de investimento tem que ser considerada. E é de se esperar que a escolha dos insumos esteja correlacionada com a habilidade fixa não observada, pois em um contexto de maximização da utilidade os investimentos dependem do retorno esperado.

A estratégia mais utilizada na literatura de economia de educação para suprir a ausência da história de investimentos, é a utilização de uma defasagem da medida de proficiência. Assume-se que esta defasagem capta a história dos investimentos, ou seja, $E[y_{it} | X_{it}, X_{it-1}, \dots, X_{i1}, y_{it-1}] = E[y_{it} | X_{it}, y_{it-1}]$. Esta hipótese é conhecida como completude dinâmica. No entanto, se o efeito da aprendizagem for persistente ao longo do tempo, tal hipótese pode não se sustentar. Para testar esta hipótese a estimação deve conter mais de uma defasagem da variável dependente.

Esta especificação é conhecida como valor adicionado. A forma funcional geral é dada por (8).

$$(8) \quad y^*_{it} = \alpha y^*_{it-1} + X_{it}\beta + \gamma u_{i0} + v_{it}$$

A estimação de (8) por MQO produzirá estimativas inconsistentes, mesmo que a habilidade não observada não seja correlacionada com os insumos, pois, necessariamente estará correlacionada com a variável defasada. Portanto, na presença de efeitos não observados a variável defasada sempre será endógena.

O segundo desafio na estimação dos parâmetros da função de produção educacional é controlar pelo efeito da habilidade não observada. Assume-se que o efeito da habilidade não observada é o mesmo em todos os períodos, ou seja, γ não depende de t . Para eliminar o efeito fixo de habilidade, realiza-se uma transformação na equação (8), tirando a primeira diferença. A equação resultante é representada por (9), sem a presença de u_{i0} .

$$(9) \quad y^*_{it} - y^*_{it-1} = \alpha(y^*_{it-1} - y^*_{it-2}) + (X_{it} - X_{it-1})\beta + v_{it} - v_{it-1}$$

A equação (9) pode ser representada através da notação de diferenças, conforme (10).

$$(10) \quad \Delta y^*_{it} = \alpha \Delta y^*_{it-1} + \Delta X_{it}\beta + \Delta v_{it}$$

Estimar (9) diretamente através de MQO produz estimativas inconsistentes, pois, o termo é correlacionado com Δy^*_{it-1} . Arellano e Bond (1991) utilizam y^*_{it-2} como instrumento para Δy^*_{it-1} , pois por construção é correlacionado com Δy^*_{it-1} , mas não é correlacionado com Δv_{it} .

Uma hipótese que implicitamente estava sendo imposta nos modelos descritos até este momento, é que o vetor de insumos X_{it} é estritamente exógeno, ou seja, $E(X_{it}v_{is}) = 0$ para qualquer $s, t = 1, 2, \dots, T$. No contexto educacional, esta é uma hipótese pouco realista, pois implica dizer que nem as famílias e nem as escolas irão reagir a choques de produtividade na escolha dos insumos do próximo período.

O estimador proposto por Arellano e Bond permite relaxar esta hipótese. É exigida uma hipótese menos restritiva conhecida como exogeneidade sequencial. A exogeneidade sequencial exige que $E(X_{it}v_{is}) = 0$ para $s \geq t$, mas permite que $E(X_{it}v_{is}) \neq 0$ para $s < t$. Neste caso, o vetor $(X_{i1}, X_{i2}, \dots, X_{is-1})$ é um instrumento válido para X_{it} . Se os insumos forem endógenos, ou seja, correlacionados com o erro idiossincrático do mesmo período, a correção é realizada utilizando mais uma defasagem do vetor de insumos.

O terceiro desafio para estimar os parâmetros na função de produção educacional é considerar erros de medida na variável dependente. Os testes de proficiência apresentam apenas uma aproximação

do verdadeiro nível de habilidades, portanto, a variável observada de habilidades possui erro de medida, ou seja, $y_{it}^* = y_{it} + \varepsilon_{it}$.

Portanto, (10) passa a ser escrita como:

$$(11) \quad (y_{it} + \varepsilon_{it}) - (y_{it-1} + \varepsilon_{it-1}) = \alpha[(y_{it-1} + \varepsilon_{it-1}) - (y_{it-2} + \varepsilon_{it-2})] + (X_{it} - X_{it-1})\beta + v_{it} - v_{it-1}$$

Rearranjando os termos e colocando em função das diferenças:

$$(12) \quad \Delta y_{it} = \alpha \Delta y_{it-1} + \Delta X_{it} \beta + (\Delta v_{it} + \Delta \varepsilon_{it} - \alpha \Delta \varepsilon_{it-1})$$

A segunda defasagem da proficiência proposta como instrumento anteriormente é correlacionada por construção com $\Delta \varepsilon_{it-1}$. Andrabi *et al.* (2011) propõem 3 alternativas para resolver esta endogeneidade. A primeira e mais direta é utilizar uma terceira defasagem da variável de proficiência como instrumento, a segunda é instrumentalizar utilizando a proficiência de uma disciplina diferente, e a terceira é corrigir o erro de medida analiticamente utilizando informações sobre erro padrão da medida de proficiência fornecida pela TRI.

No banco de dados utilizado neste trabalho a informação do erro padrão não foi disponibilizada, portanto a terceira alternativa proposta por Andrabi *et al.* (2011) não será testada. Serão testadas as duas primeiras alternativas.

O uso do estimador desenvolvido por Arellano e Bond (1991) pode não ser o mais adequado quando a variável dependente apresenta uma persistência ao longo do tempo. Neste caso, é preferível utilizar o estimador desenvolvido por Blundell e Bond (1998). Este estimador utiliza instrumentos adicionais realizando a estimação um sistema, com a equação em primeira diferença e a equação original em nível. As variáveis da equação em nível são instrumentalizadas com a defasagem de suas próprias primeiras diferenças.

A correção da endogeneidade no estimador de Arellano e Bond (1991) e Blundell e Bond (1998) é realizada através do uso de instrumentos. Para esta estratégia produzir estimativas consistentes, os instrumentos utilizados devem ser exógenos. Quando o número de instrumentos utilizados é superior ao número de parâmetros a serem estimados, a validade dos instrumentos pode ser testada através do teste de Sargan, em que sob a hipótese nula os instrumentos são válidos.

Realizar a estimação do modelo dinâmico através de Mínimos Quadrados Ordinários ou Efeitos Fixos produz estimativas inconsistentes. No entanto, os valores das estimativas servem para checar os valores encontrados com a estimação usando a metodologia de Arellano e Bond (1991) ou Blundell e Bond (1998). De acordo com Roodman (2009) e Bond (2002), as estimativas produzidas por Mínimos Quadrados Ordinários são enviesadas para cima enquanto que as estimativas obtidas pelo uso do estimador *within* envia os resultados para baixo. Portanto, bons valores para as estimativas devem estar no intervalo entre estas duas estimações.

6 – Resultados

Esta seção apresenta os resultados dos exercícios econométricos. A tabela 8 apresenta os resultados para a estimação da especificação contemporânea por Mínimos Quadrados Ordinários - MQO. A estimação foi realizada considerando as informações para a última onda. Para que as estimativas sejam consistentes, é necessário assumir que os insumos passados não são correlacionados com os insumos presentes e que a heterogeneidade individual também não é correlacionada com os insumos presentes. Esta estimação desconsidera a dinâmica do aprendizado. Apesar da teoria da formação de habilidades e de evidências empíricas não sustentarem estas hipóteses, os resultados da estimação contemporânea sem controle da heterogeneidade individual são apresentados como base de comparação.

As variáveis utilizadas são as mesmas apresentadas nas tabelas 5 e 6 da seção 4 (estatísticas descritivas). Para as variáveis sobre raça e gênero da criança, salário e experiência do professor, nem todos os indivíduos tinham a informação. Para não perder observações, foi acrescentada uma variável

dummy para cada uma das variáveis informando se havia ou não resposta para o indivíduo. Foram realizadas três estimações com especificações diferentes, incluindo controles de forma gradativa para as proficiências em matemática e português. As três especificações foram controladas pela série do aluno e município de residência.

No modelo 1, foram utilizadas apenas características da criança e da família (nível socioeconômico, raça e gênero). Os resultados mostram uma forte dependência do aprendizado em relação ao nível socioeconômico e há um diferencial positivo dos brancos em relação aos não brancos, tanto para a proficiência em matemática quanto para português. Em relação ao gênero nenhuma diferença significativa foi encontrada para a proficiência em matemática e um diferencial positivo em favor das meninas na proficiência em português.

O modelo 2 mostra as estimativas da especificação acrescentando informações sobre a rede de ensino da escola (pública/privada) e da infraestrutura (existência de laboratório de informática, laboratório de ciências e biblioteca). A magnitude das variáveis individuais se reduz, evidenciando uma associação entre características da criança com características das escolas. A informação sobre a rede de ensino mostrou uma contribuição positiva das escolas especiais, privadas e estaduais em relação às escolas municipais para as duas disciplinas analisadas, o mesmo ocorrendo para as escolas com laboratório de ciências e biblioteca. As estimativas também mostram uma contribuição negativa da existência de laboratório de informática.

No terceiro modelo foram adicionadas informações sobre os professores (salário, experiência e escolaridade). A magnitude das variáveis individuais se reduz novamente, no entanto de forma menos acentuada, e também há uma redução dos coeficientes das variáveis de escola. O salário do professor apresenta um efeito negativo, enquanto que o aumento na escolaridade e experiência possuem um efeito positivo. Essa análise é a mesma para as duas disciplinas. De uma forma geral, a principal diferença entre as duas disciplinas é a magnitude dos parâmetros, sendo sistematicamente superior para a proficiência em matemática.

Os próximos resultados são obtidos utilizando a especificação da dinâmica do aprendizado. São apresentados resultados para a estimação por MQO, por efeitos fixos e pelo estimador de Blundell e Bond (1998) para modelos com uma defasagem da variável dependente, duas e três defasagens. A inclusão gradativa de defasagens tem o objetivo de mostrar o comportamento da dinâmica do aprendizado. Conforme mencionado na seção 5 (Estratégia Empírica), em modelos dinâmicos, as estimativas por MQO e efeitos fixos são enviesadas. No entanto, servem como limites para as estimativas de Blundell e Bond, pois as estimativas por MQO são enviesadas para cima, enquanto que as estimativas por efeitos fixos são enviesadas para baixo. Portanto, os valores para os coeficientes das defasagens da variável dependente devem estar no intervalo entre as estimativas de MQO e efeitos fixos. Em todas as estimações as variáveis utilizadas são as mesmas para a estimação do modelo 3 da tabela 8.

O estimador de Blundell e Bond permite considerar três situações na relação entre as variáveis explicativas e o termo de erro (já expurgado dos efeitos fixos). As variáveis podem ser exógenas, pré-determinadas (correlacionadas com o erro do período anterior) e endógenas (correlacionadas com erro do mesmo período). Nas estimativas apresentadas na tabela 9, as variáveis de características individuais foram consideradas exógenas. O mesmo tratamento foi dado para as *dummies* de município, e para o salário do professor, laboratório de ciências e rede de ensino da escola⁶. A informação de série cursada pelo aluno foi considerada como pré-determinada, pois o resíduo do período anterior certamente está relacionado com a série do período atual. E por fim, as variáveis de escolaridade e experiência do professor, e existência de laboratório de informática e biblioteca foram consideradas como endógenas. Como exemplo de uma possível fonte de endogeneidade, suponha uma política aplicada no período atual que simultaneamente altere estas características dos professores e da escola e que também altere outras características que podem influenciar no aprendizado e que não estão sendo controladas pelo modelo,

⁶ Estas três variáveis, embora possam existir argumentos para serem consideradas pré-determinadas ou endógenas, foram tratadas como exógenas devido à baixa variação ao longo do tempo, o que impede a identificação se usadas defasagens destas como instrumentos.

como alguma nova prática didática. Em uma situação como esta, os coeficientes das variáveis observadas podem estar refletindo, na verdade, a alteração na prática didática não observada.

Tabela 8 – MQO – Especificação contemporânea

Variável dependente - proficiência	Matemática			Português		
	(1)	(2)	(3)	(1)	(2)	(3)
Nível socioeconômico	44.958*** (0.996)	25.552*** (1.243)	25.059*** (1.245)	15.780*** (0.370)	9.310*** (0.466)	9.032*** (0.466)
Branco	6.565*** (1.275)	5.124*** (1.231)	4.854*** (1.229)	2.815*** (0.475)	2.363*** (0.461)	2.219*** (0.459)
Feminino	1.548 (1.187)	1.236 (1.143)	1.015 (1.140)	5.782*** (0.442)	5.701*** (0.428)	5.611*** (0.426)
Escola Federal		54.223*** (2.974)	60.004*** (3.212)		18.625*** (1.114)	20.845*** (1.200)
Escola Estadual		9.533*** (1.515)	3.816** (1.945)		1.550*** (0.568)	-0.664 (0.727)
Escola Municipal			Omitida			
Escola Privada		43.198*** (2.201)	39.418*** (2.674)		13.528*** (0.825)	12.028*** (0.999)
Laboratório de informática		-8.123*** (1.711)	-7.831*** (1.717)		-2.115*** (0.641)	-2.056*** (0.642)
Laboratório de Ciências		4.193** (1.862)	4.231** (1.888)		1.592** (0.698)	1.609** (0.706)
Biblioteca		6.784*** (1.874)	6.272*** (1.903)		2.384*** (0.702)	2.427*** (0.711)
Salário do Professor			-0.004*** (0.001)			-0.001*** (0.000)
Salário do Professor não informado			-11.180*** (4.272)			-3.363** (1.597)
Professor com Ensino Superior			0.677 (1.731)			1.808*** (0.647)
Experiência do Professor – menos de 5 anos			-8.263*** (2.540)			-4.854*** (0.949)
Experiência do Professor – 5 a 10 anos			-4.023** (2.002)			-2.272*** (0.748)
Experiência do Professor – 11 a 15 anos			-4.856*** (1.744)			-1.890*** (0.652)
Experiência do Professor – mais de 15 anos			Omitida			
Experiência do Professor – não informada			-9.607*** (2.302)			-3.725*** (0.861)
Constante	11.988* (7.177)	2.117 (7.272)	39.063*** (9.387)	69.097*** (2.671)	65.342*** (2.724)	81.317*** (3.509)
Observações	8,605	8,605	8,605	8,605	8,605	8,605
R-quadrado	0.335	0.384	0.388	0.350	0.389	0.396

Fonte: elaborado pelos autores.

Nota 1: *** significância a 1%, ** significância a 5%, * significância a 10%

Nota 2: desvio padrão entre parênteses

Nota 3: todas as estimações as foram controladas por efeitos fixos de município e pela série do aluno.

Feitas estas considerações, os resultados das estimações dinâmicas são apresentados na sequência. A primeira estimação contém apenas uma defasagem da proficiência. As colunas de 1 a 3 da tabela 9 apresentam os resultados da estimação por Mínimos Quadrados Ordinários (MQO), Efeitos Fixos (EF) e Blundell e Bond (BB), respectivamente⁷. No painel A, são apresentados os resultados para a proficiência em matemática. O valor do coeficiente da variável defasada na estimativa por BB é de 0,740 (significativo a 1%) e está no intervalo entre as estimativas de MQO e EF. No entanto, o teste de Sargan rejeita o modelo apresentando um valor muito alto para a estatística de teste. Este resultado indica que o uso dos instrumentos não é válido. Isto pode ter ocorrido em função do processo dinâmico da aprendizagem não ser captado com o uso de apenas uma defasagem.

O próximo passo foi considerar duas defasagens para a variável dependente. Novamente foram realizadas três estimativas de acordo com o acontecido para apenas uma defasagem. Os resultados são apresentados nas colunas de 4 a 6 da tabela 9. Os valores para as variáveis defasadas foram de 0,398 para a primeira defasagem e 0,426 na segunda defasagem, ambas significativas a 1% (resultados de Blundell e Bond, coluna 6). Estes resultados vão ao encontro do argumento de um processo dinâmico da aprendizagem de forma mais persistente, ou seja, a segunda defasagem da proficiência contribui para explicar a proficiência atual. Da mesma forma como ocorreu para o uso de uma única defasagem, os valores dos coeficientes da variável defasada estão no intervalo entre as estimativas de MQO e EF⁸. No entanto, o teste de Sargan rejeita novamente o modelo. Apesar de o teste continuar rejeitando o modelo, o valor da estatística de teste teve uma forte redução, apontando indícios que o uso de mais uma defasagem contribui para aproximar o modelo do verdadeiro modelo de formação de habilidades.

Na sequência foi acrescentada uma terceira defasagem da proficiência. Os resultados encontrados para os coeficientes foram significativos indicando que a persistência do aprendizado permanece mesmo considerando três períodos de defasagem. Neste modelo o valor da estatística do teste de Sargan se reduziu novamente, embora tenha continuado a rejeitar o modelo.

A coluna 10 refaz a estimação com três defasagens da proficiência, mas considera que há uma persistência do efeito de insumos passados sobre a proficiência. Portanto, ao invés de incluir duas defasagens dos insumos endógenos, inclui apenas a terceira defasagem. Isso porque se houver uma correlação serial dos erros, pode haver uma correlação entre os insumos defasados com o termo de erro. Com esta estimação, o resultado do teste de Sargan não rejeitou o modelo conforme demonstra o valor da probabilidade do teste. Os coeficientes das defasagens da variável dependente estão entre os valores encontrados para a estimativa por MQO (coluna 7) e por EF (coluna 8)⁹. Portanto, considera-se que o uso de três defasagens consegue captar a dinâmica do processo de aprendizado. Apesar da segunda defasagem não apresentar coeficiente com valor significativo, a terceira defasagem volta a ser significativa.

Os resultados confirmam que o aprendizado é um processo dinâmico, ressaltando a importância que os níveis de habilidades passados têm no aprendizado presente. Este resultado indica que mesmo que as crianças entrassem todas em escolas iguais em suas características, a desigualdade de habilidades aumentaria durante os anos escolares, pois as crianças que iniciassem com níveis de habilidades mais elevados teriam um ritmo maior de aprendizado.

Quanto aos efeitos de insumos escolares e de características individuais sobre o aprendizado, embora a incorporação do componente dinâmico tenha reduzido tais efeitos, como apontado pelos testes F da tabela 9, os mesmos permanecem significativos a 1%.

O painel B da tabela 9 mostra os resultados para as estimativas de painel dinâmico para a proficiência em português. A inclusão das defasagens de forma gradativa apresenta o mesmo resultado das estimativas de matemática. Os resultados da coluna 10 são considerados consistentes de acordo com o

⁷ Na parte inferior da tabela, são apresentadas as defasagens utilizadas como instrumentos para as equações em diferença e em nível e os valores do teste de Sargan.

⁸ O valor da estimativa pontual da segunda defasagem é superior ao encontrado pela estimativa de MQO, mas esta dentro do intervalo de confiança.

⁹ Considerando o intervalo de confiança.

teste de Sargan. Os coeficientes das defasagens da variável dependente são positivos e significativos, indicando a persistência do aprendizado também em português.

7 - Considerações Finais

O artigo contribui para as políticas de avaliação educacional e de formulação de políticas de redução das desigualdades utilizando uma análise da dinâmica do aprendizado e dos efeitos de investimentos realizados anteriormente à entrada da criança na escola. Para tanto, a partir do modelo teórico de Cunha e Heckman (2007) para tecnologia da formação de habilidades e com o uso de informações longitudinais do GERES, um banco de dados educacional com ainda pouco explorado na literatura nacional, são estimados parâmetros de uma função de produção educacional.

Os resultados encontrados a partir de técnicas de estimação de painel dinâmico mostram evidências empíricas da persistência do aprendizado. Este resultado indica que a escola tem um papel limitado na redução de desigualdades de habilidades, dado a dependência do aprendizado a habilidades passadas. Ou seja, grande parte do que a criança consegue aprender no sistema de ensino regular é dependente dos investimentos em capital humano que foram realizados na primeira infância, antes da sua entrada na escola.

Os resultados dos efeitos dos insumos escolares não podem ser considerados conclusivos, em virtude da não estabilidade de seus coeficientes com alterações de especificações dos modelos. No entanto, existe uma indicação de efeitos positivos de recursos escolares, como biblioteca e laboratório de ciências, além do salário dos professores.

Como contribuição para a formulação de políticas públicas destaca-se a necessidade de intervenções em crianças de famílias de baixo nível socioeconômico antes mesmo da idade escolar. O baixo nível de habilidades medido quando as crianças entram na escola persiste durante o período escolar, principalmente para matemática.

Tabela 9 – Estimações de Painel Dinâmico

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PAINEL A - MATEMÁTICA										
L.proficiência	0.840*** (0.005)	0.175*** (0.006)	0.740*** (0.011)	0.638*** (0.007)	-0.119*** (0.007)	0.398*** (0.028)	0.578*** (0.008)	-0.475*** (0.011)	0.644*** (0.086)	0.592*** (0.142)
L2.proficiência				0.390*** (0.009)	-0.022*** (0.008)	0.426*** (0.020)	0.321*** (0.010)	-0.036*** (0.010)	0.069 (0.044)	0.084 (0.057)
L3.proficiência							0.254*** (0.013)	0.115*** (0.015)	0.320*** (0.047)	0.333*** (0.064)
Teste F características individuais	85.40***	-	201.60***	53.74***	-	237.50***	20.44***	-	32.65***	18.05***
Teste F características escolares	83.26***	4.06***	243.79***	44.58***	2.14**	352.50***	12.08***	0.28	47.30***	28.47***
Teste F características dos professores	2.95***	4.88***	19.34***	2.79**	3.25***	46.62***	1.20	1.32	39.39***	40.67***
R-squared	0.792	0.767	-	0.783	0.750	-	0.779	0.674	-	-
e(sarganp)	-	-	2337	-	-	471.5	-	-	53.64	17.92
e(sargan)	-	-	0.0000	-	.-	0.0000	-	-	0.0036	0.528
PAINEL B - PORTUGUES										
L.proficiência	0.718*** (0.004)	0.036*** (0.006)	0.509*** (0.012)	0.608*** (0.006)	-0.166*** (0.007)	0.509*** (0.041)	0.552*** (0.007)	-0.466*** (0.009)	0.580*** (0.091)	0.520*** (0.117)
L2.proficiência	-	-	-	0.249*** (0.006)	-0.115*** (0.006)	0.247*** (0.022)	0.219*** (0.008)	-0.196*** (0.009)	0.235*** (0.058)	0.226*** (0.067)
L3.proficiência	-	-	-	-	-	-	0.099*** (0.007)	-0.000 (0.008)	0.124*** (0.029)	0.151*** (0.031)
Teste F características individuais	114.05***	-	413.81***	59.17***	-	167.97***	27.93***	-	23.44***	16.97***
Teste F características escolares	48.38***	4.43***	274.26***	13.34***	1.24	66.59***	4.54***	2.82**	29.81***	23.31***
Teste F características dos professores	4.37***	6.91***	24.26***	3.74***	14.64***	8.52***	1.98*	12.95***	12.46***	18.84***
R-squared	0.773	0.725	-	0.762	0.626	-	0.764	0.552	-	-
e(sarganp)	-	-	1314	-	-	281.3	-	-	55.08	23.29
e(sargan)	-	-	0.0000	-	-	0.0000	-	-	0.0024	0.2250
Observations	34,040	34,040	34,040	25,460	25,460	25,460	16,956	16,956	16,956	16,956
Number of idaluno	-	8,605	8,605	-	8,605	8,605	-	8,605	8,605	8,605
Instrumentos Eq. em diferença	-	-	Y_{t-2}, X_{t-2}	-	-	Y_{t-3}, X_{t-2}	-	-	Y_{t-4}, X_{t-2}	Y_{t-4}, X_{t-3}
Instrumentos Eq. em nível	-	-	$\Delta Y_{t-2}, \Delta X_{t-2}$	-	-	$\Delta Y_{t-3}, \Delta X_{t-2}$	-	-	$\Delta Y_{t-4}, \Delta X_{t-2}$	$\Delta Y_{t-4}, \Delta X_{t-3}$

Fonte: elaborado pelos autores.

Nota 1: *** significância a 1%, ** significância a 5%, * significância a 10%

Nota 2: desvio padrão entre parênteses

8 - Referências

- Andrabi, T., Das, J., Khwaja, A. I. e Zajonc, T. (2001). **Do Value Added Estimates Add Value?** Accounting for Learning Dynamics. American Economic Journal: Applied Economics
- Andrade, D. F e Miranda, E. C. M. (2004). **GERES: Relatório do Plano Amostral**. GERES-2005. Disponível em: < <http://www.geres.ufmg.br>>. Acesso em: 22/05/2011.
- Arellano, M. e Bond, S. (1991). **Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations**. Review of Economic Studies 58: 277-297.
- Barros, R. P. e Mendonça, R.S.P. (1996). **Os Determinantes da Desigualdade no Brasil**. Economia brasileira em perspectiva. Rio de Janeiro: IPEA, p. 421-474.
- Blundell, R., and S. Bond. (1998). **Initial conditions and moment restrictions in dynamic panel data models**. Journal of Econometrics 87: 115-143.
- Brooke, N. e Soares, J.F. (Orgs.) (2008). **Pesquisa em Eficácia Escolar: Origens e Trajetórias**. Belo Horizonte: Editora UFMG.
- Cameron, S. V. e Heckman, J.J. (1998). **Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males**. Journal of Political Economy, University of Chicago Press, vol. 106(2), pp 262-333, Abril.
- Coleman, J. S. (1966). Equality of Educational Opportunity (COLEMAN) Study (EEOS), [Computer file]. ICPSR06389-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007-04-27. doi:10.3886/ICPSR06389
- Cunha, F., Heckman, J. J., Lochner, L. e Masterov, D. (2006). **Interpreting the Evidence on Life Cycle Skill Formation**. In Handbook of the Economics of Education, edited by E. A. Hanushek and F. Welch. Amsterdam: North-Holland.
- Cunha, F. e Heckman, J. (2007). **The Technology of Skill Formation**. American Economic Review, 97(2): 31-47.
- Cunha, F. e Heckman, J. (2009). **The Economics and Psychology of Inequality and Human Development**. NBER Working Paper n. 14695.
- Ferreira, F. H. G. (2000). **Os Determinantes da Desigualdade de Renda no Brasil: Luta de Classes ou Heterogeneidade Educacional?** Texto para Discussão n. 415, Departamento de Economia PUC-Rio.
- Geres, (2005). **Estudo Longitudinal sobre a Qualidade e Equidade no Ensino Fundamental Brasileiro**.
- Hanushek, E. A. e Rivkin, S. G. (2006). **School Quality and the Black-White Achievement Gap**. Working Paper no. 12651, NBER, Cambridge, MA
- IBGE - Instituto Brasileiro de Geografia e Estatística. **Censo demográfico, 1991, 2000 e 2010**. Rio de Janeiro.
- INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Censo Escolar, 2005-2008**.
- Keane, M. P. e Wolpin, K. I. (1997). **The Career Decisions of Young Men**. Journal of Political Economy, University of Chicago Press, vol. 105(3), pp 473-522, Junho.
- Langoni, C. G. (1973). **Distribuição da Renda e Desenvolvimento Econômico do Brasil**. Rio de Janeiro, Editora Expressão e Cultura.

- Leibowitz, Arleen. 1974. "Home Investments in Children." *J.P.E.* 82, no. 2, pt.2 (March/April): S111–S131.
- Ministério do Trabalho e Emprego. **Raismigra**. Brasília: MTE, 2005/2008.
- Murnane, R. J., J. B. Willett, and F. Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." *Rev. Econ. and Statis.* 77 (2): 251–66.
- Neal, D. A. e Johnson, W. R. (1996). **The Role of Premarket Factors in Black-White Wage Differences**. *Journal of Political Economy*, University of Chicago Press, vol. 104(5), pp 869-95, Outubro.
- PISA – Programme for International Student Assessment. (2009).
- Roemer, J. E. (1998). **Equality of Opportunity**. Harvard University Press, Cambridge.
- Roodman, D. (2009). **How to do xtabond2: An introduction to difference and system GMM in Stata**. *The Stata Journal*. Vol 9. N°1.
- Todd, P. e Wolpin, K. I. (2003). **On the Specification and Estimation of the Production Function for Cognitive Achievement**. *Econ. J.* 113 (485):F3–F33.
- Todd, P. e Wolpin, K. (2007). **The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps**. *Journal of Human Capital*, University of Chicago Press, vol. 1(1), pages 91-136.