

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE ADMINISTRAÇÃO, CONTABILIDADE E ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA DO DESENVOLVIMENTO

**PROGRAMA MAIS EDUCAÇÃO: AVALIAÇÃO DO IMPACTO DA EDUCAÇÃO
INTEGRAL NO DESEMPENHO DE ALUNOS NO RIO GRANDE DO SUL**

Lauren Lewis Xerxenevsky

Porto Alegre

2012

Lauren Lewis Xerxenevsky

**PROGRAMA MAIS EDUCAÇÃO: AVALIAÇÃO DO IMPACTO DA EDUCAÇÃO
INTEGRAL NO DESEMPENHO DE ALUNOS NO RIO GRANDE DO SUL**

Dissertação de mestrado apresentado à
Faculdade de Administração,
Contabilidade, e Economia da Pontifícia
Universidade Católica do Rio Grande do
Sul

Orientador: Prof. Dr. Paulo de Andrade Jacinto

Co-orientador: Prof. Dr. Adelar Fochezatto

Porto Alegre

2012

X6p

Xerxenevsky, Lauren Lewis

Programa Mais Educação: avaliação do impacto da educação integral no desempenho de alunos no Rio Grande do Sul. / Lauren Lewis Xerxenevsky. – Porto Alegre, 2012.

142 f.

Dissertação (Mestrado em Economia do Desenvolvimento) – Faculdade de Administração, Contabilidade e Economia, PUCRS.

Área de Concentração: Economia Regional.

Linha de Pesquisa: Desenvolvimento Local e Desigualdades Regionais.

Orientação: Prof. Dr. Paulo de Andrade Jacinto.

Co-orientação: Prof. Dr. Adelar Fochezatto.

1. Economia – Rio Grande do Sul. 2. Avaliação de Desempenho. 3. *Propensity Score Matching*. 4. Desempenho Escolar. 5. Método de Diferenças em Diferenças. 6. Política Educacional - Avaliação. I. Jacinto, Paulo de Andrade. II. Fochezatto, Adelar. III. Título.

CDD 330.98165

Ficha catalográfica elaborada pela bibliotecária:

Cíntia Borges Greff – CRB 10/1437

LAUREN LEWIS XERXENEVSKY

**“Programa mais educação: avaliação do impacto da
educação integral no desempenho de alunos no Rio
Grande do Sul”**

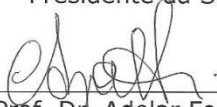
Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Economia do Desenvolvimento, pelo Programa de Pós-Graduação em Economia, da Faculdade de Administração, Contabilidade e Economia, da Pontifícia Universidade Católica do Rio Grande do Sul.

Aprovado em 28 de março de 2012.

BANCA EXAMINADORA:



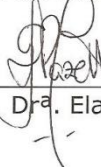
Prof. Dr. Paulo de Andrade Jacinto
Presidente da Sessão



Prof. Dr. Adelar Fochezatto
Co-orientador



Prof. Dr. Sabino Porfo Junior



Profª. Dra. Elaine Toldo Pazello

Porto Alegre
2012

DEDICATÓRIA

Ao meu esposo, *Diônifer Alan da Silveira*, pelo apoio incondicional, incentivo, força, amor, carinho e amizade sem igual.

AGRADECIMENTOS

Considerando que esta dissertação é resultado de uma caminhada que não começou na PUC-RS, a tarefa de agradecer pode não ser fácil. Para não correr o risco cometer injustiças, gostaria de agradecer de antemão a todos que passaram pela minha vida e que de alguma forma contribuíram para a construção do que sou hoje. Particularmente, agradeço algumas pessoas pela contribuição direta na construção deste pesquisa:

Agradeço aos amigos e colegas da Secretaria do Planejamento, Gestão e Participação Cidadã, pelo estímulo ao estudo sobre políticas públicas, em especial aos amigos do Departamento de Planejamento (DEPLAN).

À Secretaria Estadual do Planejamento e ao Governo Estadual pelo apoio à realização deste mestrado e pela liberação concedida para estudo em tempo integral. Em especial a duas ex-chefes e amigas que pelo incentivo ao estudo e à pesquisa científica: Maria Lúcia Leitão de Carvalho e Rejane Maria Alievi.

Agradeço ao meu orientador, professor Paulo de Andrade Jacinto, por ter acreditado no meu tema de pesquisa. Pela paciência na orientação, contribuições, incentivo e rigor científico e, principalmente, pela autonomia e confiança a mim conferida e pela amizade com a qual conduziu a nossa convivência ao longo desse período.

Ao Professor Adelar Fochezatto gostaria de agradecer a atenção e as conversas sobre o tema de pesquisa.

De um modo geral, gostaria de agradecer a todos os colegas e professores do pós-graduação pelo apoio, convívio e aprendizado nos últimos dois anos. E em especial, aos colegas Carla Albert, Diego Paiva, Flaviane de Oliveira, Guilherme de Oliveira, Isabel Silva, Isabelita Barboza, Kayline Gomes, Laura Vernier, Luana Collet, Mário Lima, Neuler de Almeida, Rodrigo Assis (Tuk), Rodrigo Scopel e Tatiele Lacerda, o meu muito obrigada. Aos professores Augusto Alvim, Carlos Eduardo Lobo e Silva, Gustavo de Moraes, Izete Pengo Bagolin, Osmar de Souza e Valter Stulp pelos ensinamentos e conhecimentos passados.

Aos colegas e bolsistas do Núcleo de Estudos e Pesquisa Econômica - NEPE pela convivência e troca de experiências nos últimos dois anos e as gurias da Secretaria do Pós, e em especial, a Janaína Vianna, pela ajuda e prontidão.

À CAPES e ao Governo Federal pelo apoio no financiamento do curso.

Agradeço também aos servidores da Secretaria Estadual da Educação do RS e do Ministério da Educação, pela atenção e solicitude que me atenderam durante a realização desta dissertação. Sem a ajuda deles nada disto seria possível.

Ao amigo e doutor pela UFPR, Marco Túlio de França, pelas dúvidas esclarecidas por e-mail e pela paciência em me escutar. E à professora Marília Patta Ramos da UFRGS pelas primeiras contribuições durante a defesa do projeto.

À Fundação Itaú por ter me proporcionado na reta final do mestrado a participação no Curso Avançado em Avaliação Econômica de Programas Sociais. Agradeço aos professores, aos colegas do curso e aos amigos que fiz em São Paulo, pela oportunidade de aprendizado e convívio.

E por fim, não poderia deixar de mencionar as pessoas mais importantes da minha vida: minha família. Pais, sogros, irmãos, cunhados e sobrinhos, pelo carinho, apoio e confiança que tanto eu precisava e pela compreensão da ausência do convívio familiar. Agradeço ainda esta dissertação ao meu esposo. Hoje, estamos colhendo, juntos, os frutos do nosso empenho. Muito Obrigada!

*...Se o homem, um governo ou uma instituição renuncia a conduzir e deixa-se conduzir, desiste de dominar e é arrastado pelos fatos, abdica então da principal liberdade humana, que é tentar decidir por nós e para nós onde queremos chegar e como lutar para alcançar nossos objetivos. O planejamento é assim uma ferramenta das lutas permanentes que o homem trava desde o início da humanidade para conquistar graus crescentes de liberdade... **Carlos Matus** (economista chileno, autor, entre outros, dos livros “Adeus, Senhor Presidente” e “Política, Planejamento e Governo”)*

RESUMO

O “Programa Mais Educação” é uma iniciativa do Governo Federal para a implementação da educação integral nas escolas públicas de todo o Brasil e está em funcionamento desde 2008. O programa, que contava com 274 escolas participantes no Rio Grande do Sul no ano de 2009, tem como uma das suas finalidades a implementação de ações pedagógicas e socioeducativas no contra-turno das escolas para melhorar o rendimento e o aproveitamento dos alunos. Nesse trabalho, avaliou-se o impacto do programa nas notas médias de português e matemática das escolas públicas do Rio Grande do Sul na Prova Brasil (Inep/MEC) para a 4ª e 8ª séries do ensino fundamental. Utilizou-se o método de estimação de diferenças em diferenças conjugado com o pareamento através do *propensity score matching*. Demonstrou-se que o programa tem um efeito positivo e estatisticamente significativo para as notas médias de português das escolas participantes do programa na 4ª série do ensino fundamental. Esse impacto é maior para as escolas que iniciaram o programa em 2008, dado o maior tempo de exposição ao programa. Por outro lado, para a proficiência em matemática (resolução de problemas) da 4ª série, o efeito do programa mostrou-se negativo. Além disso, foi encontrado efeito nulo sobre o desempenho escolar dos alunos da 8ª série tanto para português quanto para matemática.

Palavras-chaves: avaliação de impacto; *propensity score matching*; diferenças em diferenças; desempenho escolar.

ABSTRACT

The Program of the Brazilian Federal Government called "Mais Educação", in operation since 2008, was created to implement continuing education in public schools throughout the country. The Program, which included 274 participating schools in the state of Rio Grande do Sul in 2009, has as one of its purposes the implementation of educational and socio-educational activities to improve the educational performance and attainment. In this study, was evaluated the impact of the Program in the average scores of the disciplines of Portuguese language and mathematics in public schools of Rio Grande do Sul obtained in exams (Prova Brasil) for the 4th and 8th-graders. The method used was of estimation of *differences in differences* in conjunction with *propensity score matching*. The program has a positive effect and is statistically significant for the average scores of Portuguese language in the 4th grade of the participating schools. This impact is higher for schools that participated in the program since 2008, given the longer exposure to the program. On the other hand, for proficiency in mathematics (problem solving) in the 4th grade, the effect of the program was negative. Furthermore, it was found no effect on performance of 8th grade students for both Portuguese language and mathematics.

Keywords: impact evaluation, *propensity score matching*, differences in differences; school performance.

LISTA DE QUADROS

Quadro 1: Tipologias de avaliação de políticas: objetivos, critérios e modelos analíticos	28
Quadro 2: Revisão da literatura – Avaliação de programas e políticas de educação	42
Quadro 3: Variáveis indicadoras de resultado (dependentes)	60
Quadro 4: Variáveis de controle	62
Quadro 5: Variáveis de causa	63
Quadro 6: Macrocampos e atividades previstas no Mais Educação	67
Quadro 7: Revisão da literatura – Modelos hierárquicos	126
Quadro 8: Revisão da literatura – Indicadores de qualidade	129

LISTA DE TABELAS

Tabela 1: Escolas participantes do “Programa Mais Educação” no Rio Grande do Sul – 2008 e 2009	70
Tabela 2: Construção da amostra para o ano de 2007 e 2009 – Escolas do Rio Grande do Sul	71
Tabela 3: Amostras para o Rio Grande do Sul para a 4ª série e 8ª série	71
Tabela 4: Estatísticas descritivas e testes de médias da amostra 1 para a 4ª série	74
Tabela 5: Estatísticas descritivas e testes de médias da amostra 1 para a 8ª série	82
Tabela 6: Teste da propriedade de balanceamento do <i>propensity score</i> para 4ª série (Amostra1)	91
Tabela 7: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em língua portuguesa para a 4ª série (Amostra 1)	94
Tabela 8: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em matemática para a 4ª série (Amostra 1)	96
Tabela 9: Teste da propriedade de balanceamento do <i>propensity score</i> para 8ª série (Amostra1)	99
Tabela 10: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em língua portuguesa para a 8ª série (Amostra 1)	102
Tabela 11: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em matemática para a 8ª série (Amostra 1)	103
Tabela 12: Estatísticas descritivas e testes de médias da amostra 2 para a 4ª série	133
Tabela 13: Estatísticas descritivas e testes de médias da amostra 2 para a 8ª série	135
Tabela 14: Estimação do <i>propensity score</i> – amostra 1 - 4ª série	137
Tabela 15: Estimação do <i>propensity score</i> – amostra 1 - 8ª série	139
Tabela 16: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em língua portuguesa para a 4ª série (Amostra 2)	141
Tabela 17: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em matemática para a 4ª série (Amostra 2)	141
Tabela 18: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em língua portuguesa para a 8ª série (Amostra 2)	142
Tabela 19: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em matemática para a 8ª série (Amostra 2)	142

LISTA DE GRÁFICOS

Gráfico 1: Distribuição das notas de português na 4ª série da Prova Brasil por grupo, 2007 – Amostra 1	75
Gráfico 2: Distribuição das notas de matemática na 4ª série da Prova Brasil por grupo, 2007 – Amostra 1	76
Gráfico 3: Distribuição da média de matrículas por turma na 4ª série por grupo, 2007	78
Gráfico 4: Percentual de escolas por esfera administrativa e por grupo para a 4ª série, 2007	79
Gráfico 5: Distribuição do IDEB das escolas em 2007 para a 4ª série por grupo, 2007	80
Gráfico 6: Distribuição das notas de português na 8ª série da Prova Brasil por grupo, 2007 – Amostra 1	83
Gráfico 7: Distribuição das notas de matemática na 8ª série da Prova Brasil por grupo, 2007 – Amostra 1	84
Gráfico 8: Distribuição da média de matrículas por turma na 8ª série por grupo, 2007	86
Gráfico 9: Percentual de escolas por esfera administrativa e por grupo para a 8ª série, 2007	87
Gráfico 10: Distribuição do IDEB médio das escolas em 2007 para a 8ª série por grupo, 2007	88
Gráfico 11: Notas médias das escolas por grupo em matemática e português para a 4ª série, 2005 e 2007.....	105
Gráfico 12: Notas médias das escolas por grupo em matemática e português para a 8ª série, 2005 e 2007.....	106

LISTA DE SIGLAS

ASBRAN - Associação Brasileira de Nutrição
CEDEPLAR - Centro de Desenvolvimento e Planejamento Regional de Minas Gerais
CEM - *Coarsened Exact Matching*
CETA - *Comprehensive Employment and Training Act*
CPC - *Chicago Child-Parent Center and Expansion*
DD - Método Diferenças em Diferenças
DF - Distrito Federal
EDUDATABRASIL - Sistema de Estatísticas Educacionais
EF - Ensino Fundamental
EI - Ensino Infantil
EJA - Programa de Educação de Jovens e Adultos
EM - Ensino Médio
FIES - Fundo de Financiamento ao Estudante do Ensino Superior
FINBRA - Finanças Municipais do Brasil
FIRJAN - Federação das Indústrias do Rio de Janeiro
FUNDEB - Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização de Profissionais da Educação
GO - Goiás
IBGE - Instituto Brasileiro de Geografia e Estatística
IDEB - Índice de Desenvolvimento da Educação
INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IPEA - Instituto de Pesquisa Econômica Aplicada
LDB - Lei de Diretrizes e Bases da Educação Nacional
MDM - *Mahalanobis Distance Matching*
MDS - Ministério do Desenvolvimento Social e Combate à Fome do Brasil
MDTA - *Manpower Development and Training Act*
ME - Ministério dos Esportes
MEC - Ministério da Educação
MG - Minas Gerais
MPOG - Ministério do Planejamento, Orçamento e Gestão
MQO - Mínimos Quadrados Ordinários
MQP - Mínimos Quadrados Ponderados
MS - Mato Grosso do Sul
NSW - *National Supported Work Program*
OBMEP - Olimpíada Brasileira de Matemática das Escolas Públicas
PA - Pará
PDDE - Programa Dinheiro Direto na Escola
PE - Pernambuco
PIB - Produto Interno Bruto
PISA - Programa Internacional de Avaliação de Alunos
PLOA - Projeto de Lei Orçamentária Anual
PNAD - Pesquisa Nacional por Amostra de Domicílios
POF - Pesquisa de Orçamentos Familiares
PPA - Plano Plurianual
PPV/MS - Pesquisa sobre Padrões de Vida / Ministério da Saúde
PR - Paraná
PROEB - Programa de Avaliação da Rede Pública de Educação Básica de Minas Gerais
PROINFO - Programa Nacional de Tecnologia Educacional
PRONASCI - Programa Nacional de Segurança Pública com Cidadania
ProUni - Programa Universidade para Todos

PS - Propensity Score
PSM - *Propensity Score Matching*
PTRF - Programa de Transferência de Recursos Financeiros
RAIS- Relação Anual de Informações Sociais
RCM - Modelo Causal de Rubin
RM - Região Metropolitana
RO - Rondônia
RS - Rio Grande do Sul
SAEB - Sistema Nacional de Avaliação da Educação Básica
SAERS - Sistema de Avaliação do Rendimento Escolar do Rio Grande do Sul
SARESP - Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo
SE - Sergipe
SEB - Secretaria de Educação Básica do Ministério da Educação
SECAD - Secretaria de Educação Continuada, Alfabetização e Diversidade do Ministério da Educação
SEDUC - Secretaria da Educação do Estado do Rio Grande do Sul
SIMAVE - Sistema Mineiro de Avaliação da Educação Pública
SIMEC - Sistema Integrado de Monitoramento Execução e Controle do MEC
SP - São Paulo
STN - Secretaria do Tesouro Nacional do Ministério da Fazenda
TCU - Tribunal de Contas da União
TRI - Teoria de Resposta ao Item
UFMG - Universidade Federal de Minas Gerais
UnB - Universidade de Brasília
UNICEF - Fundo das Nações Unidas para a Infância

SUMÁRIO

1.	INTRODUÇÃO.....	16
2.	REFERENCIAL TEÓRICO.....	25
2.1	Avaliação de políticas públicas.....	25
2.1.1	<i>Modelos analíticos existentes</i>	26
2.1.2	<i>Avaliação do impacto de programas</i>	33
2.2	Estudos sobre educação na área de economia e suas aplicações empíricas.....	39
3.	METODOLOGIA E BANCO DE DADOS.....	45
3.1	Modelo econométrico para avaliação de impacto.....	45
3.1.1	<i>Estratégia de identificação</i>	46
a)	Metodologia do propensity score matching (PSM).....	47
b)	Metodologia de diferenças em diferenças.....	52
3.1.2	<i>Implementação da avaliação do programa Mais Educação</i>	55
a)	Estimando o propensity score.....	55
b)	Estimação por diferenças em diferenças.....	57
3.2	Base de dados.....	58
3.3	Variáveis.....	60
3.4	Descrição do “Programa Mais Educação”.....	64
3.5	A amostra para o Rio Grande do Sul.....	69
4.	RESULTADOS.....	73
4.1	Estatísticas descritivas e testes de igualdade de média e de distribuição.....	73
4.2	Resultados para a 4ª Série.....	89
4.2.1	<i>Resultados das estimativas do impacto na proficiência em língua portuguesa</i>	94
4.2.2	<i>Resultados das estimativas do impacto na proficiência em matemática</i>	96
4.3	Resultados para a 8ª Série.....	97
4.3.1	<i>Resultados das estimativas do impacto na proficiência em língua portuguesa</i>	101
4.3.2	<i>Resultados das estimativas do impacto na proficiência em matemática</i>	103
4.4	Análises de robustez.....	104
4.5	Discussão dos resultados.....	106
5.	CONCLUSÕES.....	112
6.	REFERÊNCIAS BIBLIOGRÁFICAS.....	114
7.	ANEXOS.....	126

1. INTRODUÇÃO

Esta dissertação apresenta os resultados da pesquisa acerca do impacto da educação em tempo integral¹ sobre o desempenho dos alunos das escolas da rede pública do Rio Grande do Sul a partir da avaliação do “Programa Mais Educação”.² O debate sobre a educação integral passou a ganhar força por meio da Lei de Diretrizes e Bases da Educação Nacional (LDB) de 1996, um instrumento legal que preconiza a ampliação progressiva da jornada escolar em direção a uma educação integral (artigo 34 e § 5º do artigo 87º da Lei nº 9.394 de 20 de dezembro de 1996).

É importante mencionar que a educação integral faz parte de uma discussão mais ampla sobre a qualidade da educação oferecida nas escolas brasileiras. Esse tema ganha destaque, pois, após 15 anos de esforços, a universalização da educação fundamental já está praticamente consumada (PACHECO, 2008). Estima-se³, para a faixa etária de 7 a 14 anos, que 98,0% dos jovens brasileiros estavam frequentando a escola em 2007. Além desse percentual, foram atingidos também outros objetivos nos últimos 15 anos, como por exemplo, o aumento da taxa de atendimento de jovens entre 15 a 17 anos (VELOSO, 2009) e a redução da taxa de analfabetismo no país. Contudo, apesar do significativo aumento de indicadores de “quantidade” da educação, os indicadores qualitativos estão ainda muito aquém do desejado.

A qualidade da educação tem sido considerada sobre vários pontos de vista e perspectivas distintas, tendo como elemento sinalizador do aprendizado, em muitos estudos, o desempenho dos alunos em testes padronizados como o SAEB, Prova Brasil e PISA (SOARES e MAROTTA, 2009). Entretanto, o que se observa para as escolas brasileiras é um desempenho dos alunos bem abaixo do padrão mínimo desejado pela sociedade. Ou seja, os níveis de qualidade e de aprendizado dos estudantes estão muito baixos comparativamente aos outros países e ao padrão mínimo necessário (VELOSO, 2009).

¹ Existe na literatura uma divisão entre educação integral e educação em tempo integral. A primeira se preocupa com a formação e o desenvolvimento humano mais amplo e múltiplo possível. Já a segunda trata da ampliação da jornada escolar. Aqui não faremos distinção entre os dois termos.

² O “Programa MAIS EDUCAÇÃO” foi instituído pela Portaria nº 17 de 24 de abril de 2007 e implementado posteriormente, no ano de 2008, nas escolas da rede pública municipal e estadual do Brasil.

³ Veloso (2009).

As avaliações do Sistema Nacional de Avaliação da Educação Básica (SAEB) para as disciplinas de língua portuguesa e matemática para as 4ª e 8ª séries (5º e 9º anos, respectivamente)⁴ da rede pública no Brasil são bons exemplos da situação da qualidade de educação do país.⁵ Em 2009, para os alunos da 4ª série da rede pública no Brasil, apenas 30,5% dos alunos tiveram um desempenho adequado para a sua série em língua portuguesa. Já para 8ª série, esse desempenho em língua portuguesa é ainda pior: 22,2% alcançaram o padrão mínimo. Em relação à matemática, os resultados caem ainda mais. Na 4ª série, 29,1% dos alunos se encontravam em um padrão adequado; na 8ª série, apenas 10,4% atingiram o mínimo ideal (TODOS PELA EDUCAÇÃO, 2010).

Outros indicadores que demonstram a baixa qualidade da educação no Brasil são as taxas de abandono, reprovação e distorção idade-série da rede pública. Dados do Instituto Nacional de Estudos e Pesquisas Educacionais (INEP)⁶ mostram que, no ano de 2005, para a rede pública, 14,1 % dos alunos foram reprovados no ensino fundamental e 12,3%, no ensino médio. Além disso, 32,7% dos alunos do ensino fundamental e 51,1 % dos alunos do ensino médio da rede pública possuíam distorção em relação à idade-série correspondente (2005). Como consequência⁷, as taxas de abandono do ensino fundamental e médio da rede pública no Brasil em 2005 foram de 8,2% e 17,1%, respectivamente.

O Estado do Rio Grande do Sul, objeto desta pesquisa, também apresenta indicadores ruins. Dados da Secretaria Estadual da Educação (SEDUC)⁸ mostram que, em 2008, 16,9% dos alunos no ensino fundamental da rede pública estadual foram reprovados e 20,7%, no ensino médio. Além disso, 21,9% dos alunos do ensino fundamental e 33,8% dos alunos do ensino médio possuíam distorção em relação à idade-série correspondente. Já as taxas de

⁴ A Lei 11.274, de 06 de fevereiro de 2006, amplia a duração do Ensino Fundamental para nove anos. Dessa forma, as 4ª e 8ª séries correspondem aos 5º e 9º anos. Doravante, a denominação utilizada será “série”.

⁵ Esses dados baseiam-se no conceito de “desempenho adequado” adotado pelo movimento Todos pela Educação (TODOS PELA EDUCAÇÃO, 2010). Os critérios adotados para 4ª série do ensino fundamental são os seguintes: para língua portuguesa acima de 200 pontos e para matemática acima de 225 pontos. Já para 8ª série do ensino fundamental, os critérios são: para língua portuguesa acima de 275 pontos e para matemática acima de 300 pontos.

⁶ As informações foram retiradas do Sistema de Estatísticas Educacionais – EDUDATABRASIL, do INEP.

⁷ Segundo o IPEA (p. 210, 2007), “além de ainda haver uma porcentagem residual de crianças e jovens fora da escola, proporção elevada dos que estão matriculados não aprende ou progride lentamente, repete o ano e termina por abandonar os estudos”.

⁸ Informações repassadas por e-mail. Título do e-mail: Evolução de indicadores de educação. E-mail institucional <gab-deplan@educ.rs.gov.br> 01/09/2010.

abandono do ensino fundamental e ensino médio da rede pública estadual em 2008 foram de 2,6% e 14,6%, respectivamente.

É importante mencionar ainda a forte relação existente entre desempenho escolar, repetência e abandono escolar. Como há uma porcentagem elevada de crianças e jovens que não aprendem ou que progridem em ritmo lento (vide os dados do SAEB), esses acabam repetindo e, posteriormente, abandonam os estudos⁹. Segundo um estudo realizado pelo Instituto de Pesquisa Econômica Aplicada (IPEA), no Brasil, apenas 53% dos alunos que ingressam no ensino fundamental concluem esse nível e apenas 37% desse universo concluem o ensino médio (IPEA, 2007). O estudo também afirma (p.214) que a universalização do ciclo educacional obrigatório no país não tem como limitador fundamental a oferta de vagas. O principal limitador diz respeito a “condições intra e extra-escolares” que acabam afetando o desempenho e a trajetória dos estudantes. É por essa razão que, segundo o estudo, o índice de conclusão no ensino fundamental é muito baixo no Brasil.

Cabe destacar que a educação é um dos desafios mais importantes a serem encarados nas próximas décadas no Brasil. Araújo et al. (2009) ressalta que essa questão se torna um grande problema, pois se não for resolvida é bem possível que a baixa qualidade da mão de obra nas próximas décadas se configure como um obstáculo ao crescimento de longo prazo da economia brasileira.

Ademais, Veloso (2009) destaca ainda que existem evidências de uma queda da qualidade da educação durante o período de maior expansão quantitativa da escola, isto é, após Constituição Federal de 1988. Diante da necessidade de o Estado brasileiro melhorar a qualidade de ensino, diversas políticas públicas e programas foram criados nos anos mais recentes pelos governos federal, estaduais e municipais, além de haver projetos em parceria com organizações não-governamentais. Ressaltam-se os programas e ações para formação e capacitação continuada de professores e gestores escolares, para apoio aos dirigentes municipais de educação, para implantação de salas de recursos multifuncionais, bem como para apoio à distribuição de materiais, livros e acervos bibliográficos, para auxílio à alimentação e ao transporte escolar. Além dessas políticas, cabe destacar o Programa Dinheiro Direto na Escola (PDDE), que é uma das ações complementares para a qualificação da educação. Todos esses projetos são financiados com recursos próprios do tesouro e/ou por

⁹ IPEA, 2007.

meio de fundos, como o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização de Profissionais da Educação – FUNDEB (MPOG, 2010a e MEC¹⁰).

Entre os programas criados com o objetivo de melhorar a qualidade da educação, estão as políticas de ampliação da jornada escolar ou de educação integral, como recomenda a LDB, entre as quais está o “Programa Mais Educação”. Criado em 2007 pelo governo federal, essa política, segundo o Ministério do Planejamento, Orçamento e Gestão (MPOG, 2010a), tem como base diversos estudos do Fundo das Nações Unidas para a Infância (UNICEF) que avaliam positivamente as iniciativas de ampliação da jornada escolar/educação integral. O “Programa Mais Educação”, objeto da presente dissertação, será detalhado no capítulo da metodologia. No momento, o que interessa é apresentar a possível relação entre ampliação da jornada escolar e a qualidade da educação.

Em um estudo recente realizado pela UNICEF (2010) com os 26 municípios brasileiros que apresentam boas práticas educacionais, a instituição aponta que a elevação do nível de aprendizado é compreendida como algo diretamente relacionada com o aumento da carga horária do aluno, seja em escolas com regime de tempo integral, seja em atividades de reforço. Cavaliere (2007), por sua vez, ressalta que alguns estudos no campo da educação e pedagogia consideram que a qualidade de ensino está alicerçada na relação entre o tempo e os espaços para educação. Isto é, a ampliação da educação integral pode contribuir para a qualidade da educação (KERSTENETZKY, 2006 *apud* MEC, 2009a). Já para outros pesquisadores, a extensão do horário escolar, por si só, não garante o incremento qualitativo do ensino, mas carrega essa possibilidade em potencial (CAVALIERE, 2002; COELHO, 1997, CASTRO, 2010).

Segundo Cavaliere (2007), existem diversas razões para a ampliação da jornada escolar. Entre elas, estão: (a) ampliação do tempo como forma de alcançar melhores resultados da ação escolar sobre os indivíduos, devido à maior exposição desses às práticas e rotinas escolares; (b) ampliação do tempo como adequação da escola às novas condições da vida urbana, das famílias e particularmente da mulher; (c) ampliação do tempo como parte integrante da mudança na própria concepção de educação escolar, ou seja, no papel da escola na vida e na formação dos indivíduos. Contudo, segundo dados do IPEA¹¹, o investimento necessário pelo Estado brasileiro para a implementação da educação integral em todas as

¹⁰ Informações sobre os programas em <http://portal.mec.gov.br/>. Acesso em 08/11/2011.

¹¹ <[http://www2.camara.gov.br/agencia/noticias/EDUCACAO-E-CULTURA/148651-IPEA:-INVESTIMENTO-PARA-ENSINO-DE-TEMPO-INTEGRAL-ULTRAPASSA-R\\$-20-BI.html](http://www2.camara.gov.br/agencia/noticias/EDUCACAO-E-CULTURA/148651-IPEA:-INVESTIMENTO-PARA-ENSINO-DE-TEMPO-INTEGRAL-ULTRAPASSA-R$-20-BI.html)>. Acesso em 19/02/2011.

escolas brasileiras ultrapassaria R\$ 20 bilhões, o que representaria 37,2 % da previsão de orçamento anual da área.¹² Além disso, o número de alunos em regime de educação integral é ainda modesto no país, passando de 155 mil em 2001 para 424 mil em 2009.

Diante do impasse das limitações financeiras e dos possíveis benefícios que as políticas de ampliação da jornada escolar podem ter provocado, é necessário analisar a efetividade¹³ de algum desses programas já implementados no Brasil. Partindo da ideia de que uma gestão adequada das políticas educacionais é um elemento fundamental para o desenvolvimento da qualidade da educação, um aspecto importante diz respeito às avaliações dos programas e das políticas educacionais, principalmente no que se refere às iniciativas de educação em tempo integral, que é o foco deste estudo. Segundo o MPOG (2010a), as avaliações têm como objetivo assegurar o aperfeiçoamento contínuo dos programas, permitindo verificar se os resultados desejados junto ao público-alvo, isto é, os indivíduos que receberam a política, ocorreram efetivamente.

Entretanto, no Brasil, apesar dos avanços, nos últimos anos, para a implementação dos programas e das políticas governamentais que contribuíssem com a melhoria educacional, poucas avaliações sobre seus impactos na qualidade da educação foram realizadas. Isto é, pouco se sabe sobre as contribuições que os programas governamentais de educação implementados recentemente, sobretudo os de educação integral, provocaram no desempenho escolar dos alunos e na qualidade da educação pública brasileira.

São raros os trabalhos, sejam acadêmicos ou institucionais, que avaliam a efetividade ou impacto desses programas.¹⁴ Em relação às políticas de educação integral, poucos estudos

¹² Consideram-se para o cálculo os valores previstos no PLOA 2011 da União para a manutenção do Ensino.

Fonte:

http://www9.senado.gov.br/portal/page/portal/orcamento_senado/PS_ORCEDUC/Elaboracao?p_ano=2011.

Acesso em 29/12/2011.

¹³ O termo efetividade está relacionado com o efeito desse resultado sobre a situação-problema. Em relação ao efeito, busca-se investigar a influência da intervenção governamental sobre os objetivos pretendidos de forma a indagar se os programas provocaram algum resultado sobre o público-alvo ou sobre a sociedade. Em outras palavras, a efetividade mede os efeitos (positivos ou negativos) na realidade que sofreu a intervenção. Ou seja, se houve mudanças socioeconômicas, ambientais ou institucionais decorrentes dos resultados obtidos pela política (Ministério do Planejamento, Orçamento e Gestão - MPOG, 2010b). Para esse tipo de avaliação, a literatura aponta vários métodos com o uso de técnicas estatísticas e econométricas. Entre os principais, estão: métodos de aleatorização de um experimento, método do *Propensity Score Matching*, diferenças em diferenças, variáveis instrumentais e regressão descontínua (KHANDKER; KOOLWAL; SAMAD, 2010).

¹⁴ A avaliação de efetividade de políticas e programas governamentais apresenta uma crescente evolução de trabalhos no meio acadêmico. No nível nacional, destacam-se os estudos da linha de avaliação de impacto de programas no campo do mercado de trabalho (FERNANDES; MENEZES FILHO; ZYLBERSTAJN, 2000), políticas de assistência social (RESENDE, 2006 e FERRO, 2003 e 2007) e políticas tributárias (CORSEUIL; MOURA, 2009). Já alguns estudos de avaliação de programas de educação foram realizados por Duenhas,

na área acadêmica foram realizados. Destacam-se as pesquisas realizadas no campo da educação e da pedagogia (GERMANI, 2006; PIRES, 2007; MOTA, 2008; GOMES, 2009), as quais são avaliações qualitativas sobre a percepção do professor quanto à ampliação da jornada escolar. Contudo, no campo da avaliação de impacto, há apenas dois estudos de âmbito regional: um estudo realizado em conjunto pelo CEDEPLAR/UFMG e pela Fundação Itaú (CEDEPLAR; FUNDAÇÃO ITAÚ, 2008) que avalia o programa “Escola Integrada” do município de Belo Horizonte sobre vários aspectos de motivação dos alunos para assistirem às aulas e de mudança de seus hábitos escolares. Além desse estudo, Aquino e Kassouf (2011) avaliaram recentemente o impacto do programa “Escola de Tempo Integral” da rede pública do estado de São Paulo sobre o desempenho escolar.

Ressalta-se também que a avaliação é pouco explorada no cotidiano do setor público brasileiro¹⁵, em especial a avaliação de efetividade (ou impacto) dos programas governamentais.¹⁶ Dos oito programas finalísticos¹⁷ do MEC no Plano Plurianual (PPA) 2008-2011, por exemplo, apenas dois programas contam com avaliações em andamento fora as avaliações formais obrigatórias do PPA (MPOG, 2010a). Porém, essas avaliações contratadas e realizadas por diversas universidades e centros de pesquisas¹⁸ estão relacionadas à avaliação de processos e não propriamente à avaliação de impacto.

França e Gonçalves (2010), Vasconcellos, Biondi e Menezes-Filho (2009), Biondi, Vasconcellos e Menezes-Filho (2009).

¹⁵ Para alguns autores, a dificuldade de aplicação pode ser explicada pela complexidade inerente das metodologias de avaliação (COTTA, 1998 e ARRETICHE, 2007). Já para outros autores, a dificuldade está na falta de relevância das informações da avaliação para os tomadores de decisão (SCHNEIDER, 2009). Arretche (2007) destaca algumas outras razões para a raridade de estudos de avaliação de impacto: dificuldades na obtenção informações e dificuldades operacionais relacionadas aos custos financeiros e organizativos. Independente do fator de desestímulo, a avaliação é conduzida, na maioria das vezes, por imposição de organismos financiadores ou como mero formalismo burocrático para o cumprimento das leis (MIRANDA e COSTA, 2007).

¹⁶ Uma das exceções é a experiência do Ministério do Desenvolvimento Social e Combate à Fome do Brasil (MDS), que desde 2004 já investiu mais de R\$ 15 milhões em pesquisas de avaliação (VAITSMAN; RODRIGUES; PAES-SOUSA, 2006).

¹⁷ Definição: são programas que resultam em bens e serviços ofertados diretamente à sociedade (Fonte: Lei nº 11.653, de 7 de ABRIL de 2008, que dispõe sobre o PPA federal de 2008-2011).

¹⁸ O Programa do PPA “Brasil Alfabetizado e Educação de Jovens e Adultos” possui avaliações em andamento que consideram aspectos de gestão, resultados e abrangência. Essas são realizadas pelo Centro de Seleção e de Promoção de Eventos (CESPE) da Universidade de Brasília (UnB); pelo Centro de Alfabetização, Leitura e Escrita (CEALE) da Universidade Federal de Minas Gerais (UFMG); pelo Instituto Paulo Montenegro/IBOPE e pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Já o programa do PPA “Brasil Escolarizado” conta com avaliações em andamento referente ao transporte escolar, realizado pelo Centro

Assim, constata-se que a avaliação de efetividade, que se constitui na verificação do impacto que os programas provocaram sobre a qualidade da educação, a partir dos problemas e objetivos inicialmente identificados ainda não foi instituída de forma sistemática no Brasil no campo da educação. Ressalta-se que sem uma avaliação do impacto que as ações governamentais provocaram na sociedade torna-se difícil a decisão para a manutenção, interrupção e aperfeiçoamento das ações governamentais, quando da revisão dos programas, dificultando as decisões dos gestores públicos.

Com a perspectiva de aperfeiçoamento das políticas é que se insere a avaliação do impacto do “Programa Mais Educação” no Rio Grande do Sul, o qual foi instituído em 2007 através da Portaria Interministerial nº 17 de 24 de abril de 2007 e implementado posteriormente no ano de 2008. Trata-se de uma contribuição para a formação em tempo integral de crianças, adolescentes e jovens, pela articulação de ações, projetos e programas do Governo Federal através de ações de diversos ministérios, entre os quais os Ministérios da Educação, Desenvolvimento Social e Combate à Fome, Cultura e Esportes, Ciência e Tecnologia e também Secretaria Nacional de Juventude e Assessoria Especial da Presidência da República (MEC 2009a; MEC, 2009b; Portaria nº 17 de 24/04/2007).

Para contribuir com a formação em tempo integral dos alunos através do Mais Educação, o MEC mobiliza os alunos visando: (a) a melhoria do desempenho educacional; (b) o cultivo de relações entre professores, alunos e suas comunidades; e (c) a garantia da proteção social da assistência social e à formação para a cidadania (Artigo 1º da Portaria nº 17 de 24/04/2007). Dessa forma, a realização desta dissertação está focada no primeiro objetivo do “Programa Mais Educação” que é a melhoria do desempenho educacional. Ou seja, busca-se avaliar a efetividade do programa a partir de um de três objetivos inicialmente identificados no programa, que é a promoção da melhoria do desempenho escolar dos estudantes.

Portanto, o objetivo geral da dissertação é avaliar o impacto do turno integral, expresso pelo “Programa Mais Educação”, no desempenho médio dos alunos em testes de proficiência das escolas da rede pública no Rio Grande do Sul na 4ª série e 8ª série. Os

Interdisciplinar de Estudos em Transportes (CFTRU/ Universidade de Brasília). Além disso, três programas do PPA contavam com avaliações de satisfação de beneficiários, o que pode se chamar, segundo o referencial teórico, de avaliação de efetividade subjetiva (MPOG, 2010a).

objetivos específicos da pesquisa são: (a) distinguir estatisticamente as escolas com características similares que não foram alvo do programa através da técnica de *propensity score matching*, a fim de comparar o desempenho em testes de proficiência nas notas de português e matemática de escolas alvo do “Programa Mais Educação”; (b) identificar as diferenças no desempenho escolar dos alunos “antes e após” a implementação do “Programa Mais Educação”.

Partindo do pressuposto de que uma pesquisa deva ter uma relevância social, pretende-se também promover uma discussão sobre a política pública de educação integral do Mais Educação e o resultado da avaliação do impacto desse programa no desempenho escolar dos alunos da rede pública do Rio Grande do Sul. Para tal, o presente estudo está estruturado em quatro capítulos, além da introdução.

O capítulo 2 trata do referencial teórico da pesquisa, em que é apresentada uma síntese da literatura sobre avaliação de políticas públicas e seus modelos analíticos. Também é mostrada uma revisão, com a apresentação dos principais estudos teóricos e empíricos, da subárea denominada de avaliação de impacto. Por fim, relatam-se os principais estudos empíricos sobre educação na área da economia, com ênfase nos estudos sobre educação integral.

O capítulo 3 mostra a descrição da metodologia do trabalho. Nessa seção, formalizam-se o procedimento de *propensity score matching* e o método de diferenças em diferenças, técnicas adotadas com o intuito de atingir os objetivos da pesquisa. Descrevem-se também a estratégia para a implementação dessas ferramentas na pesquisa de avaliação de impacto do “Programa Mais Educação”, bem como as informações sobre a política, os bancos de dados, as variáveis e sobre a amostra específica para o Rio Grande do Sul.

No capítulo 4 são apresentados os resultados da análise de impacto do “Programa Mais Educação” para a proficiência de matemática e língua portuguesa (4^a série e 8^a série). O programa teve um efeito positivo e estatisticamente significativo para as notas médias de português das escolas participantes do programa na 4^a série do ensino fundamental no Rio Grande do Sul. Esse impacto é maior para as escolas que iniciaram o programa em 2008, dado o maior tempo de exposição que elas tiveram ao programa. Por outro lado, para a proficiência em matemática (resolução de problemas) da 4^a série, o efeito do programa mostrou-se

negativo. Além disso, foi encontrado efeito nulo sobre o desempenho escolar dos alunos da 8ª série tanto para português quanto para matemática.

2. REFERENCIAL TEÓRICO

Nesta seção, é apresentada, inicialmente, uma síntese da literatura acerca dos diferentes tipos de avaliação de políticas públicas e seus modelos analíticos, visto que há inúmeros conceitos nesse complexo campo das políticas públicas. Após a contextualização, uma revisão da subárea de avaliação, chamada de *avaliação de impacto*, é detalhada. Nesse sentido, são mostradas as principais pesquisas teóricas e empíricas que contribuíram para o desenvolvimento da metodologia de avaliação de impacto de programas/políticas com o uso de técnicas estatísticas e econométricas.

Por fim, após a síntese sobre avaliação, é realizada também uma revisão das principais linhas de pesquisas empíricas sobre educação na área da economia. Além disso, apresentam-se os resultados dos estudos empíricos nacionais sobre os mais diversos métodos e ferramentas de pesquisa em educação. São mencionados também os estudos que analisam a educação integral ou a ampliação da jornada escolar, bem como os principais resultados encontrados na literatura do efeito dessa política.

2.1 Avaliação de políticas públicas

O processo de avaliação de um programa ou política constitui uma importante etapa do ciclo de políticas públicas¹⁹. A expressão políticas públicas²⁰ é originária das ciências políticas e pode designar tanto o conjunto de ações, que expressam a orientação estratégica e política do Estado em atividades diretas de produção de bens e serviços pelo próprio Estado, como também as atividades de regulação de outros agentes econômicos²¹, tendo em vista o

¹⁹ Ciclo de políticas públicas refere-se as fases de construção de uma política pública. O ciclo se divide em etapas sequenciais: identificação de problema, construção da agenda, implementação, monitoramento e avaliação.

²⁰ No campo das políticas públicas, há também uma subárea denominada de políticas sociais. Heidemann (2009) define políticas sociais como políticas consideradas sob o ponto de vista setorial, incluindo as áreas de educação, saúde, entre outras.

²¹ Secchi (2010) destaca, na literatura especializada de políticas públicas, a existência de duas abordagens: a abordagem estatista (*state-centered policy-making*), que considera as políticas públicas como sendo monopólio estatal, e a abordagem multicêntrica, que considera como protagonistas no enfrentamento de problemas públicos, assim como o estado, as organizações privadas, organizações não-governamentais, organismos multilaterais e

atendimento do bem comum (BERNARDONI; SOUZA; PEIXE, 1998). Para Dye (2009), a política pública é constituída por tudo o que os governos decidem fazer e ou deixar de fazer. Ou seja, o cientista político também inclui inação como uma política. Outro cientista político, Jones (1978, apud DYE, 2009), ao definir políticas públicas faz uma distinção entre várias propostas de políticas (definição de objetivos), os programas (autorização para atingir objetivos), as decisões (ações para implementação dos objetivos) e os efeitos (os impactos mensuráveis).

Em relação à avaliação de políticas públicas, essa pode ser entendida como a produção de informação para dar apoio à formulação e à tomada de decisões (SCHNEIDER, 2009). Araújo e Loureiro (2005) identificam que a avaliação extrapola sua capacidade de quantificar resultados. Em uma concepção mais ampla de avaliação, o autor considera que a avaliação de políticas públicas significa preocupar-se e comprometer-se com um melhor entendimento do Estado em ação. De forma complementar, Cano (2006) incluiu na avaliação a necessidade de averiguação das razões do sucesso ou fracasso de um programa.

No quadro analítico de Dye (2009), a avaliação constitui a última etapa do ciclo de políticas públicas. Nessa etapa, estudam-se os programas, relatam-se os *outputs* dos programas governamentais, avaliam-se os impactos sobre os públicos-alvos e propõem-se mudanças e ajustes. Na mesma linha, Secchi (2010) identifica a avaliação como a fase do ciclo de políticas públicas em que o desempenho e o processo de implementação são examinados, a fim de conhecer melhor o estado da política pública e de reduzir o nível de problema que a originou.

2.1.1 Modelos analíticos existentes

Segundo Bernardoni, Souza e Peixe (1998), o processo de avaliação de políticas públicas possui mecanismos e ferramentas próprias que cada pesquisador estabelece no momento de sua implementação. Em relação às tipologias de avaliação existentes, bem como os mecanismos e as técnicas escolhidas para análise, o que se observa é uma variada gama de conceitos dentro de um arcabouço teórico complexo. Schneider (2009) identifica que a evolução da área produziu uma massa confusa de tipos de avaliações ao invés de fornecer um quadro coerente como referencial. Assim, desde o início desta pesquisa, foi fundamental a

redes. Como Secchi (2010), a perspectiva adotada nesta dissertação é a segunda abordagem, isto é, acredita-se na pluralidade de atores protagonistas no enfrentamento dos problemas públicos.

busca por definições claras de avaliação a fim de que alguns conceitos pudessem servir de guia para as análises que seriam feitas sobre o “Programa Mais Educação”.

Nesta dissertação, considera-se a classificação conceitual adotada por Figueiredo e Figueiredo (1986) e MPOG (2010b). Nesse sentido, constata-se que as pesquisas de avaliação podem se enquadrar em dois tipos básicos: avaliação de processos²² e avaliação de impacto.

Cano (2006) também realiza essa divisão entre avaliação de processo ou de implementação e avaliação de impacto ou de resultados. Segundo o autor, o primeiro tipo de avaliação tenta averiguar em que medida um programa foi implementado conforme o plano previsto originalmente. O segundo tipo, por sua vez, tenta verificar se o impacto desejado foi atingido. É possível que essas duas dimensões tenham resultados distintos, ainda que elas sejam interdependentes. Por outro lado, as avaliações de impacto são quase sempre complementadas de avaliações de processo. Isso ocorre devido à necessidade de compreender se determinado resultado deficiente pode ser atribuído a dificuldades de implementação e/ou não eficácia na intervenção. A seguir, o modelo analítico de Figueiredo e Figueiredo (1986), com a divisão entre avaliação de processos e avaliação de impacto, é apresentado no Quadro 1.

²² Cabe ressaltar que alguns pesquisadores desconsideram a avaliação de processo como uma avaliação de fato. Entretanto, para Schneider (2009), esse tipo de avaliação (que poderia ser chamada de “monitoração”), embora não se possa atribuir uma relação de causa e efeito entre estratégia e nível de sucesso, já é por si só uma indicação de prestação de contas – que, muitas vezes, pode oferecer, aos tomadores de decisão, informações necessárias para saber se o desempenho de um programa foi satisfatório. Costa (2003, apud Sechhi, 2010) designa esse tipo de avaliação como avaliação *in itinere*. Para o autor, essa avaliação ocorre durante o processo de implementação, tendo a finalidade de promover ajustes imediatos.

Quadro1: Tipologias de avaliação de políticas: objetivos, critérios e modelos analíticos

Objetivo da avaliação	Tipo de pesquisa	Crítérios de aferição ou fracasso	Modelos analíticos para a mensuração do sucesso/ fracasso	Trabalhos na área avaliação em educação
Avaliação de processo: acompanhamento e avaliação da metodologia de implantação e execução (monitoramento)	Realização de metas ou resultados	Eficácia objetiva	Meta atingida, Meta proposta = Δ - Condições: $\Delta > 0$ (sucesso) e $\Delta < 0$ (fracasso)	Avaliação dos programas de educação¹ no PPA 2007-2011 federal (MPOG, 2010a)
	Meios: metodologia de implantação e execução	Eficácia administrativa e contábil Eficácia funcional	Auditoria Adequação entre meios, fins e objetivos	Auditorias operacionais de programas governamentais: Apoio à Formação Continuada de Professores do Ensino Fundamental (TCU, 2006), ProUni (TCU, 2009);
	Relação esforço/resultado, análise custo-benefício e custo-resultado	Eficiência instrumental, eficiência política ou social	Otimização da relação benefício máximo/custo mínimo	Avaliações da taxa interna de retorno dos programas: Olimpíada Brasileira de Matemática (BIONDI; VASCONCELLOS; MENEZES-FILHO, 2009)e Escrevendo o Futuro (VASCONCELLOS; BIONDI; MENEZES-FILHO, 2009).
Avaliação de impacto: acompanhamento, avaliação e controle dos efeitos do programa (intervenção controlada); Comparam o antes e o depois do programa, garantindo um controle analítico sobre as variáveis que atuam sobre a população	Impacto objetivo: mudanças quantitativas nas condições materiais	Efetividade objetiva	Modelo causal Inércia social: $Y' = f(Y_0)$ Intervenção Social: $Y'' = \Delta f(Y_0) + f(X_0)$	Avaliação do impacto do programa Comunidade Escola sobre a violência (DUENHAS; FRANÇA; GONÇALVES, 2010) e sobre a proficiência (FRANÇA; DUENHAS, 2011) Avaliações do impacto dos programas sobre o desempenho escolar: Olimpíada Brasileira de Matemática (BIONDI; VASCONCELLOS;

				MENEZES-FILHO, 2009) e Escrevendo o Futuro (VASCONCELLOS;BIONDI; MENEZES-FILHO, 2009).
	Impacto subjetivo: mudanças subjetivas na percepção da população sobre o seu bem-estar (nível de satisfação)	Efetividade subjetiva		Avaliação da satisfação dos beneficiários dos programas Brasil Escolarizado, Brasil Universitário e Desenvolvimento da Educação Especial (MPOG, 2010a)
	Impacto substantivo: mudança qualitativa nas condições de vida	Efetividade substantiva ²³		

Elaboração a partir de Figueiredo e Figueiredo (1986)

¹ Programas do PPA 2008-2011: Brasil Alfabetizado e Educação de Jovens e Adultos, Brasil Escolarizado, Desenvolvimento da Educação Profissional e Tecnológica, Brasil Universitário, Desenvolvimento da Educação Especial, Desenvolvimento do Ensino da Pós-Graduação e da Pesquisa Científica, Educação para Diversidade e Cidadania e Qualidade na Escola.

²³ Segundo Figueiredo e Figueiredo (1986, p. 118,) “a aferição da efetividade substantiva de uma política social transita constantemente entre a avaliação política da política implementada e a avaliação do impacto de seus resultados. Isto é, não basta aferir se o estado futuro ‘Y’ desejado ocorreu. É necessário avaliar se esta nova situação, à luz de alguns princípios de justiça social minimamente aceitos, é positiva. O exame de efetividade substantiva exige, pois, uma avaliação política da mudança”.

Conforme o Quadro 1, a avaliação de processos tem como objetivo a verificação da eficácia e eficiência dos programas. Em relação à eficácia, busca-se verificar se um programa atingiu as metas desejadas ou se está sendo desenvolvido de acordo com as diretrizes concebidas. Os critérios de aferição são divididos em eficácia objetiva (verificação de metas), administrativa ou contábil (auditoria) e funcional (análise da adequação dos meios aos objetivos de um programa).

Outro campo da avaliação de processos diz respeito à verificação da relação entre esforço e resultado, através análise de eficiência. Essa pode ser classificada, quanto aos critérios de aferição de sucesso ou fracasso, em eficiência instrumental ou eficiência social. A primeira, eficiência instrumental, é definida pela relação estrita entre custos econômicos e benefícios que são, em geral, tangíveis e divisíveis. Já a eficiência social é definida pela relação entre os custos sociais ou políticos e os benefícios derivados a partir de um programa (FIGUEIREDO; FIGUEIREDO, 1986).

A avaliação de impacto, por sua vez, tem a ambição de ser bem mais completa e complexa. Este tipo de avaliação diz respeito aos efeitos que um programa provocou sobre o público-alvo da intervenção e relaciona-se com o critério da efetividade. Além disso, a avaliação de impacto tem a intenção de estabelecer uma relação de causalidade entre um programa e as alterações na sociedade (FIGUEIREDO; FIGUEIREDO, 1986 e MPOG 2010b).

Segundo Figueiredo e Figueiredo (1986), pode-se dizer que as políticas ou programas obtiveram sucesso quando existe a possibilidade de se imputar a elas a condição de causa necessária. Isto é, a constatação de que ocorreram mudanças não é, por si só, suficiente para concluir pelo sucesso de um programa. Dessa forma, é necessário demonstrar que as mudanças não ocorreriam (de forma total ou parcialmente) sem sua implementação. Em outras palavras, é necessário demonstrar que a atuação de um programa é empiricamente importante para a determinação da mudança observada.

Além disso, três tipos de critérios de verificação de sucesso ou fracasso podem ser utilizados na avaliação de impacto: a efetividade objetiva, quando se investiga mudanças quantitativas na população-alvo do programa; a efetividade subjetiva, que avalia as alterações do estado de espírito da população-alvo; a efetividade substantiva, quando se verifica mudanças qualitativas nas condições de vida da população (FIGUEIREDO; FIGUEIREDO, 1986).

No campo educacional, considerando as definições apresentadas anteriormente, é possível classificar os diferentes tipos de avaliações realizadas, em âmbito nacional, sobre alguns programas já implementados. No campo da avaliação de processo destacam-se alguns trabalhos. A avaliação dos programas do Ministério da Educação no PPA 2007-2011, coordenado pelo MPOG (2010a), pode ser classificada como a avaliação de eficácia objetiva. Isso porque os relatórios de avaliação do PPA buscam investigar se as metas pré-estabelecidas (indicadores de programas e metas físicas e financeiras das ações) foram alcançadas ao longo de um ano específico de referência. Nesse sentido, foram avaliados pelos gestores do Governo Federal, anualmente, durante o período de 2007 a 2011²⁴ os macroprogramas: Brasil Alfabetizado e Educação de Jovens e Adultos, Brasil Escolarizado, Desenvolvimento da Educação Profissional e Tecnológica, Brasil Universitário, Desenvolvimento da Educação Especial, Desenvolvimento do Ensino da Pós-Graduação e da Pesquisa Científica, Educação para Diversidade e Cidadania e Qualidade na Escola.

Já as auditorias operacionais dos programas Universidade Para Todos - ProUni e do Fundo de Financiamento ao Estudante do Ensino Superior- FIES (TCU, 2009), no período de 2005-2008 e do Programa de Apoio à Formação Continuada de Professores do Ensino Fundamental (TCU, 2006), de 2000 a 2005, podem ser classificados tanto como avaliação de eficácia administrativa e contábil quanto avaliação funcional. Isso porque essas avaliações além de verificarem se a operacionalização dos programas dão margem para a ocorrência de algum tipo de impropriedade (auditoria tradicional), investigam também se a implementação está alinhada aos objetivos e normas previstos nos programas (eficácia funcional).

Por fim, na avaliação de processo, há também alguns estudos, no nível nacional, relacionados à avaliação de eficiência ou, então, de retorno econômico. Nesse tipo de avaliação, busca-se basicamente responder se os resultados dos programas diante dos seus custos estão razoáveis (PAZELLO, 2007). Nesse sentido, destacam-se as avaliações já realizadas em relação ao custo-benefício dos programas Olimpíada Brasileira de Matemática (BIONDI; VASCONCELLOS; MENEZES-FILHO, 2009) e Escrevendo o Futuro (VASCONCELLOS; BIONDI; MENEZES-FILHO (2009).

A avaliação de impacto, como já mencionado, é dividida em três segmentos: efetividade subjetiva, objetiva e substantiva; no entanto, somente para os dois primeiros critérios foram encontrados estudos nacionais. As avaliações desenvolvidas para os programas Brasil Escolarizado, Brasil Universitário e Desenvolvimento da Educação Especial (MPOG,

²⁴ MPOG (2009a, 2009b, 2010a, 2011a, 2011b).

2010a), são, segundo o Ministério do Planejamento, focados na satisfação dos beneficiários em relação ao programa. Dessa forma, essas avaliações podem ser classificadas como avaliações de efetividade subjetiva.

Em relação à avaliação de impacto de efetividade objetiva de políticas públicas, foco desta pesquisa, a literatura acadêmica nacional é crescente, porém incipiente.

Em âmbito regional, destacam-se dois trabalhos de avaliação de impacto do programa Comunidade Escola de Curitiba/PR. O primeiro, de Duenhas, França e Gonçalves (2010), investiga o impacto do programa sobre os diferentes tipos de violência; o segundo, desenvolvido por França e Duenhas (2011), avalia o desempenho médio das escolas municipais em língua portuguesa e matemática nos exames de proficiência da Prova Brasil. Silva (2010), por sua vez, analisa os efeitos do Projeto Jovem de Futuro²⁵ na proficiência de português e matemática nos estados de Minas Gerais e Rio Grande do Sul. Em outro estudo de avaliação de impacto, Rocha e Belluzo (2010) investigam o impacto do Programa de Transferência de Recursos Financeiros (PTRF) da Secretaria Municipal de Educação de São Paulo/SP sobre as notas da Prova Brasil das 4ª e 8ª séries das escolas municipais.

Mais recentemente, os programas de educação integral também começaram a ser avaliados. O estudo realizado em conjunto pelo CEDEPLAR/UFMG e pela Fundação Itaú (CEDEPLAR; FUNDAÇÃO ITAÚ, 2008) avaliou o programa Escola Integrada do município de Belo Horizonte sobre vários aspectos de motivação para assistir aula e mudança de hábitos escolares dos alunos. Aquino e Kassouf (2011) investigaram o impacto do programa Escola de Tempo Integral da rede pública do estado de São Paulo sobre o desempenho escolar nos anos de 2007 e 2008.

Em âmbito nacional, Loureiro et al. (2008) já avaliou a efetividade dos programas de alfabetização, entre eles o Programa de Educação de Jovens e Adultos – EJA sobre o analfabetismo no período de 2002 a 2006. Em relação ao desempenho escolar, já foram avaliados o impacto do programa Olimpíada Brasileira de Matemática (BIONDI; VASCONCELLOS; MENEZES-FILHO, 2009) e Programa Escrevendo o Futuro (VASCONCELLOS; BIONDI; MENEZES-FILHO, 2009) em que foram utilizadas, para a verificação de impacto, as notas da Prova Brasil da 8ª e 4ª série da rede pública no Brasil, respectivamente.

Na literatura internacional destacam-se as análises já realizadas para os programas voltados à educação infantil. O programa *Chicago Child-Parent Center and Expansion* (CPC)

²⁵ Programa da Fundação Itaú Social.

foi avaliado sobre os aspectos de proficiência, repetência e delinquência (FUERST; FUERST, 1993; TEMPLE; REYNOLDS; MIEDEL, 2000). Já o Programa *Head Start*, também de educação infantil e do governo norte-americano, recebeu avaliações de impacto sobre as perspectivas das notas nas proficiências, repetência, saúde, e até mesmo de impacto de mais longo prazo, como gravidez na adolescência, prisões e desempenho na universidade (GARCES; THOMAS; CURRIE, 2000; CURRIE; THOMAS, 1995 e 1999; LEE et al., 1990)²⁶.

Como o foco desta pesquisa é justamente a avaliação de impacto de efetividade objetiva de políticas públicas, ou apenas “avaliação de impacto”, a próxima seção apresenta uma síntese da literatura sobre o tema. Enfatiza-se o avanço dos estudos econométricos e estatísticos dos últimos 35 anos para avaliação do efeito causal de políticas e programas.

2.1.2 Avaliação do impacto de programas

Nas últimas décadas, muitas pesquisas foram realizadas sobre o efeito causal ou sobre o impacto de programas com a utilização de técnicas econométricas e de técnicas estatísticas. Segundo Imbens e Wooldridge (2008 e 2009), uma recente literatura teórica foi sendo construída no campo de pesquisa de avaliação com características combinadas de trabalhos anteriores desenvolvidos tanto no campo da estatística como no campo da econometria. Dessa forma, a pesquisa em avaliação de impacto atingiu um nível de maturidade que tornou uma ferramenta importante em muitas áreas de pesquisa empírica na economia, como, por exemplo, em economia de trabalho, finanças públicas, economia do desenvolvimento e economia industrial e em outras áreas da microeconomia aplicada.

Quanto à literatura desenvolvida no campo da econometria, o trabalho inicial de Ashenfelter (1978) e, posteriormente, os trabalhos desenvolvidos por Ashenfelter e Card (1985), Heckman e Robb (1985), Lalonde (1986), Fraker e Maynard (1987), Card e Sullivan (1988), e Manski (1990) podem ser considerados os pioneiros do campo de avaliação de impacto de programas. Esses trabalhos foram motivados principalmente pela implementação de avaliações de programas de treinamento de trabalhadores para a qualificação no mercado

²⁶ A revisão dessa literatura foi realizada por Currie (2001).

de trabalho em ambientes de observação ou não-aleatorizados. Destaca-se também que o foco dessa literatura econométrica está tradicionalmente ligado a questões de endogeneidade ou auto-seleção. Isso ocorre porque, pessoas que optarem por se inscrever em um programa de formação são diferentes daquelas que não optarem por se inscrever no programa, sendo que muitas vezes essas variáveis ou esses aspectos não são observáveis na prática. Essa diferença, caso exerça influência na resposta dos indivíduos frente a uma intervenção de interesse pode invalidar completamente as comparações dos resultados de causalidade, mesmo após adição de variáveis de controles. Como consequência desse problema, a literatura econométrica foi desenvolvida a partir da utilização de métodos tradicionais para lidar com a endogeneidade, como, por exemplo, os métodos de efeito fixo a partir de análises de dados em painel e métodos de variáveis instrumentais (IMBENS; WOOLDRIDGE, 2008 e 2009).

Em seu trabalho seminal, Ashenfelter (1978) analisou o impacto de um treinamento de 3 meses sobre o salários dos trabalhadores no ano de 1964 nos Estados Unidos. O programa *Manpower Development and Training Act* (MDTA) tinha como objetivo garantir o aperfeiçoamento na formação de trabalhadores nos Estados Unidos. A partir dos resultados, o autor demonstrou o problema do viés de seleção resultante das características dos indivíduos participantes desse programa de treinamento.

Posteriormente, Ashenfelter e Card (1985) mediram os ganhos salariais dos trabalhadores em um programa de treinamento (*Comprehensive Employment and Training Act* -CETA) implementado no ano de 1976 nos Estados Unidos, e compararam os resultados com um grupo de controle (indivíduos que não participaram do treinamento) ao longo do tempo. A sensibilidade dos resultados desse estudo não-experimental²⁷ levou os autores a concluir sobre a necessidade de uma seleção aleatória para a designação do tratamento.

No ano seguinte, Lalonde (1986) estimou o impacto do programa norte-americano *National Supported Work Program* (NSW), de 1975 a 1978, sobre a renda dos trabalhadores após a intervenção (treinamento). O autor analisou em que medida os estimadores não-experimentais podem reproduzir os resultados não-viesados dos estudos randomizados, a partir da comparação de diferentes bases de dados para a composição do grupo de controle. Assim como os autores anteriores, Lalonde (1986) concluiu que muitos dos estudos não-

²⁷ Refere-se a estudos em que a designação dos indivíduos tratados e não tratados não foi realizada de forma aleatória.

experimentais não fornecem estimativas confiáveis. Isso ocorre em função de erros de especificação dos modelos e da construção equivocada de grupos de controle, resultando em estimativas diferentes dos resultados experimentais. Dessa forma, o autor conclui que os pesquisadores devem estar cientes dos potenciais erros nas avaliações não-experimentais.

Na mesma linha dos artigos anteriores, Fraker e Maynard (1987) comparam os resultados de um estudo experimental com um estudo não-experimental a partir da base de dados *National Supported Work Demonstration*, de 1975 a 1979, nos EUA. Os autores também concluíram que os desenhos não-experimentais não são confiáveis para a estimativa da efetividade de programas. Isso ocorre porque as estimativas de impacto tendem a ser sensíveis à metodologia de construção do grupo de controle e ao modelo usado para estimação. Em outros termos, os resultados podem ser altamente viesados dependendo da população alvo, do grupo de comparação selecionado e/ou do modelo analítico utilizado.

Já Card e Sullivan (1988) mensuraram o impacto de um programa subsidiado de treinamento (o mesmo analisado por Ashenfelter e Card em 1985). Eles investigaram a probabilidade de adultos participantes do sexo masculino conseguirem emprego após a participação no programa de treinamento (*Comprehensive Employment and Training Act* (CETA), para o ano 1976. Os autores também ressaltam a necessidade de cautela na interpretação dos resultados. Segundo os autores, essa cautela se deve ao fato de os dados disponíveis serem não-experimentais. Contudo, muito do viés existente em estudos observacionais sobre o efeito de tratamento podem ser isolados através de procedimentos ecométricos, abstraindo o efeito efetivo do programa.

Em relação à literatura do campo da estatística, essa inicia em uma perspectiva diferente, isto é, ela parte da análise do efeito causal em estudos experimentais, cuja inspiração são os estudos da medicina (IMBENS; WOOLDRIDGE, 2008 e 2009). Os trabalhos pioneiros de Neyman (1923), Fischer (1925) e, posteriormente, na década de 1970, de Rubin (1973a, 1973b, 1974, 1977, 1978) são dominantes na literatura da análise de efeitos causais. Destaca-se que a formulação inicial de Rubin (Modelo Causal de Rubin - RCM) estabeleceu o efeito causal como comparação dos resultados potenciais para a mesma unidade ou indivíduo a partir de diferentes níveis de exposição ao tratamento (com tratamento e sem tratamento). Contudo, dada a impossibilidade dessa construção na prática, esse ficou

conhecido na literatura como o problema de inferência causal ou problema do contrafactual não observado.

O caso mais simples de análise apresentada por Rubin ocorre quando a atribuição do tratamento é aleatória, e por isso, os resultados potenciais dos grupo de tratamento e controle passam a ser independentes do tratamento, o que torna a tarefa de obter estimadores bastante simples. Ou seja, em função do caráter aleatório do experimento, qualquer diferença entre os resultados do grupo de tratamento e controle pode ser atribuída de forma exclusiva ao programa. Assim, o efeito médio do tratamento nada mais é do que a diferença de médias do *status* de tratamento (participação ou não no programa). Portanto, no caso dos experimentos aleatórios, é mais fácil obter estimadores para o efeito médio do tratamento, isto é, não há a necessidade da utilização de um ferramental econométrico sofisticado.

Contudo, apesar da utilização de experimentos aleatórios em alguns campos da economia²⁸, eles continuam relativamente raros na área (IMBENS; WOOLDRIDGE, 2008 e 2009). Dado que a maioria dos programas na prática não são aleatorizados, para a avaliação de impacto de programas é muito mais comum na literatura, a realização de estudos observacionais ou não-aleatorizados. Nesse caso, os pesquisadores devem, de alguma forma, ajustar a hipótese de homogeneidade da distribuição das características entre grupos de tratamento e controle pelas diferenças observadas entre as co-variáveis (variáveis de pré-tratamento), removendo todo o tipo de viés existente entre as unidades de tratamento e de controle. Imbens e Wooldridge (2008 e 2009) identificam na literatura sobre avaliação de impacto, considerando a junção das áreas de econometria e estatística, pelo menos cinco tipos de abordagens relacionadas à estimação dos efeitos de tratamento em estudos observacionais ou não-aleatorizados, as quais são descritas a seguir.

A primeira grande área diz respeito à *análise de sensibilidade*. Essa procura verificar a robustez das estimativas específicas diante de pequenos desvios. Ou seja, nesses trabalhos, a robustez das estimativas específicas é testada considerando a hipótese de independência

²⁸ Alguns dos experimentos mais interessantes foram realizados na economia do desenvolvimento, como os trabalhos de Miguel e Kremer (2004); Duflo (2001); Angrist, Bettinger e Kremer (2005); Banerjee et al. (2007) e na economia comportamental, com o trabalho de Bertrand e Mullainathan (2004). No caso da educação, Silva (2010) analisou os efeitos do Projeto Jovem de Futuro, pela diferença de médias uma vez que a participação no programa foi estabelecida por sorteio, após um pareamento inicial.

condicional ou *unconfoundedness*²⁹ (ROSENBAUM; RUBIN, 1983; ROSENBAUM; RUBIN, 1995). A segunda abordagem, *bounds analysis*, cujo desenvolvimento deve-se a Manski (1990, 2003), busca analisar intervalos de confiança consistentes que são derivados e estimados de acordo com a escolha de dados e com os pressupostos impostos pelos pesquisadores.

Já a terceira abordagem, *variáveis instrumentais*, baseia-se em tratamentos adicionais, isto é, os chamados instrumentos, que satisfazem a hipótese de exogeneidade e restrições de exclusão (IMBENS; ANGRIST, 1994 e ANGRIST; IMBENS; RUBIN, 1996). A quarta, por sua vez, é denominada *regressão descontínua*. Nessa abordagem, comparações podem ser feitas explorando a continuidade e a mudança de valores das variáveis de controle. Segundo Imbens e Wooldridge (2008 e 2009), essa definição tem uma longa tradição na estatística. Os principais trabalhos desenvolvidos na área são de Shadish, Campbell e Cook (2002), Cook (2008), Vanderklaauw (2002), Hahn, Todd, e Vanderklaauw (2000), Lee (2001) e Porter (2003).

Por fim, pesquisadores na área de avaliação de impacto também desenvolveram a *técnica de diferenças em diferenças*. A técnica é utilizada quando há dados adicionais na forma de amostras das unidades de tratamento e controle antes e após o tratamento. Uma aplicação inicial foi realizada por Ashenfelter e Card (1985) e trabalhos teóricos recentes incluem Abadie (2005), Bertrand, Duflo e Mullainathan (2004), Athey e Imbens (2006) e Donald e Lang (2007).

Destaca-se que, nos estudos de avaliação de impacto, muitos trabalhos acabaram se concentrando nas técnicas para pareamento. Isso ocorre porque existe a necessidade, em estudos observacionais ou não-aleatorizados, de encontrar um grupo de controle semelhante com o grupo que recebeu a política a partir de características observáveis. Ou seja, os procedimentos de pareamento consistem na construção de um grupo de controle (contrafactual) mais parecido possível com o grupo de tratamento (grupo que recebeu o programa ou tratamento).

²⁹ A hipótese de independência condicional ou ignorabilidade do tratamento será vista no capítulo de metodologia.

Entre as técnicas de pareamento, a ferramenta mais recorrente na literatura é o *propensity score matching* (PSM), desenvolvida inicialmente por Rosenbaum e Rubin (1983). A técnica baseia-se no método de escore de propensão ou *propensity score*, que é a probabilidade condicional de receber tratamento a partir das características observáveis (ROSENBAUM; RUBIN, 1983). Isto é, através do PSM é possível construir um grupo estatístico de comparação que é baseado no modelo de probabilidade de participação no tratamento (ou programa), utilizando co-variáveis (KHANDKER; KOOLWAL; SAMAD, 2010).

Além das técnicas de pareamento baseada no *propensity score*, constata-se que uma vasta literatura foi sendo desenvolvida a partir de outros métodos de pareamento para a inferência causal. Assim, na literatura, o que se observa é o amplo debate existente entre as diferentes técnicas de pareamento.³⁰ Imbens (2004), por exemplo, realiza uma revisão da literatura sobre os vários métodos de estimativa semiparamétrica que nos últimos anos têm sido propostos, incluindo as técnicas de pareamento e os métodos utilizando o *propensity score* como ponderação e “blocking”³¹, bem como a combinação dessas abordagens.

King et al. (2011), por sua vez, relata além das técnicas de pareamento baseados no *propensity score*, a técnica de pareamento *Mahalanobis Distance Matching* (MDM), mais antigo e popular e *Coarsened Exact Matching* (CEM), uma nova abordagem para o pareamento. Segundo King et al. (2011), essa última abordagem, apesar de simples, conta com muitas propriedades desejáveis. Contudo, apesar do amplo debate, não há consenso, segundo Imbens (2004), sobre quais são os melhores métodos para serem aplicados na prática. No entanto, o autor destaca a vantagem de ter surgido, nos últimos anos, uma quantidade de novos estimadores à disposição dos pesquisadores.

Ressalta-se que, para finalidade deste trabalho e considerando que a presente pesquisa é um estudo observacional ou não-aleatorizado, será utilizada a técnica *propensity score matching* (PSM) como metodologia para pareamento. A seguir é mostrada uma síntese dos estudos na área da economia da educação, incluindo os estudos de avaliação de programas educacionais.

³⁰ As mais recentes contribuições sobre as técnicas de pareamento são de: Hahn (1998), Imbens (2004), Angrist e Hahn (2004), Austin (2008a, 2008b), Iacus, King e Porro (2011) e King et al. (2011)

³¹ Categorização dos sujeitos no pré-teste (blocking).

2.2 Estudos sobre educação na área de economia e suas aplicações empíricas

Nas últimas duas décadas no Brasil, a implementação de sistemas avaliativos de larga escala possibilitou uma grande difusão de estudos empíricos educacionais. Isto é, a introdução, na década de 90, do Sistema Nacional de Avaliação da Educação Básica (SAEB) e, posteriormente, a implementação de sistemas estaduais de avaliação de educação básica³², serviram de instrumentos para o desenvolvimento de uma literatura na área de economia recente sobre os determinantes da qualidade da educação no país.

Nesse sentido, pode se destacar três grandes linhas de estudos que estão sendo desenvolvidas: (a) pesquisas que utilizam como ferramenta os modelos hierárquicos (ou multiníveis) com a finalidade de analisar, em grande parte, os determinantes do desempenho escolar, expressos pela proficiência dos alunos; (b) os estudos que focalizam algum aspecto da qualidade da educação, traduzido em um indicador específico de qualidade, com as mais variadas técnicas de estimação; (c) os estudos específicos de avaliação de impacto de programas educacionais.

Conforme esta classificação, a primeira linha de pesquisa na área da economia da educação utiliza os modelos hierárquicos (multiníveis) para investigar os fatores determinantes da qualidade da educação, expressa pelo desempenho escolar dos estudantes. Segundo Natis (2001), a criação desses modelos foi motivada, pela própria estrutura de dados. Isto é, os dados relacionados à educação possuem uma estrutura de agrupamentos em função das próprias características de suas observações: os alunos estão sempre agrupados em turmas, as turmas em escolas e essas, por sua vez, estão agrupadas nos municípios. Dessa forma, os modelos hierárquicos (multiníveis) têm a capacidade de incorporar essa estrutura hierárquica que existe naturalmente nos dados de educação.

Várias são as vantagens apontadas pela literatura para a utilização dos modelos hierárquicos (multiníveis) em relação ao modelo clássico de regressão. Soares e Mendonça (2003) destacam que esse método de estimação baseia-se em modelos mais flexíveis e que utilizam melhor a informação presente na amostra. Além disso, o modelo hierárquico

³² Destacam-se os sistemas de avaliação dos estados de São Paulo, Minas Gerais, Pernambuco e Rio Grande do Sul.

(multinível) fornece uma equação para cada escola, permitindo análises individuais. Ressalta-se também que o uso das informações pelos dados agrupados possibilita formular e testar hipóteses relativas a efeitos de interação entre elementos dos diferentes níveis e também permite a realização de análises quanto à partição da variabilidade da variável resposta para os diversos níveis (SOARES; MENDONÇA, 2003).

Quanto aos resultados dos estudos empíricos com o uso de modelos hierárquicos (quadro 7, em anexo), observa-se que as bases de dados mais utilizadas são as do Sistema de Avaliação da Educação Básica- SAEB e as dos sistemas estaduais de avaliação dos estados de São Paulo e de Minas Gerais.³³ Entre os mais variados aspectos, destacam-se os fatores socioeconômicos dos alunos e a escolaridade dos pais como determinantes ao desempenho escolar nas proficiências em português e matemática. Entre os principais os fatores negativos que influenciam o desempenho estão a defasagem idade-série, o abandono e alunos do sexo masculino. (SOARES; MENDONÇA, 2003, FERRÃO; BELTRÃO; SANTOS, 2002 e 2003, ALBERNAZ; FERREIRA; FRANCO, 2003, SOARES, 2003, DE JESUS e LAROS 2004, SOARES, 2005, GONÇALVES; RIOS-NETO; CÉSAR, 2011).

Além dos modelos hierárquicos (multiníveis), uma segunda linha de pesquisa foi desenvolvida a partir das mais variadas técnicas e questões a serem respondidas sobre os determinantes da qualidade da educação (quadro 8, em anexo). Essas pesquisas se concentram que algum aspecto da qualidade da educação expressa geralmente em termos de um indicador de qualidade.

Ainda sobre a investigação dos determinantes de desempenho escolar, alguns estudos dessa segunda linha de pesquisa mostram que, além dos fatores socioeconômicos identificados nos modelos hierárquicos (multiníveis), outros aspectos relacionados à escola também possuem impacto sobre a proficiência dos alunos. Esse é o caso do efeito positivo da ausência de alta rotatividade dos professores ao longo do ano, da experiência média dos professores superior a dois anos em sala de aula e da existência, na escola, de conexão com a internet (BIONDI; FELÍCIO, 2007). Por outro lado, os estudos demonstram o efeito negativo

³³ Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - SARESP e Sistema Mineiro de Avaliação da Educação Pública/ Programa de Avaliação da Rede Pública de Educação Básica de Minas Gerais- SIMAVE/PROEB.

da formação heterogênea da turma e da alta rotatividade dos professores, este último aspecto semelhante como no estudo anterior (MENEZES; SOARES, 2010).

Além disso, destacam-se também pesquisas que avaliam o efeito positivo do número de horas-aula sobre o desempenho escolar, verificado nos trabalhos de Menezes-Filho (2007), Oliveira (2008 e 2010) e Menezes e Soares (2010). Entretanto, Franco (2009) só encontrou o efeito positivo da ampliação de mais de 5 horas diária de aula para a rede privada de ensino. Há também um estudo que encontrou influência positiva da educação infantil (FELÍCIO; VASCONCELLOS, 2007) sobre desempenho escolar e outros que não encontraram significância estatística do gasto público em educação sobre desempenho escolar (AMARAL; MENEZES-FILHO, 2008 e FIGUEIREDO; HADDAD; FREGUGLIA, 2011)

Cabe mencionar também outros aspectos qualitativos que essas pesquisas tentam responder além da proficiência dos alunos. Alguns estudos investigam, por exemplo, os diferenciais de qualidade existentes entre a rede pública e privada (FRANCO, 2009, CURI; MENEZES-FILHO, 2010 e COSTA et al. 2011) e também um estudo que avalia o impacto da carência nutricional sobre a defasagem idade-série do aluno (GOMES e MAGALHÃES, 2011). Destacam-se que estes estudos empíricos utilizam as mais variadas técnicas para estimação, que vão desde procedimentos estatísticos de regressão linear até as técnicas de análise de escolha discreta, dados em painel, além de técnicas de avaliação de impacto, como pareamento.

Por fim, em uma terceira linha de pesquisa, estão os trabalhos de avaliação de impacto de programas de educação. Ressalta-se, conforme comentado anteriormente, que, embora bastante incipiente, essa área de pesquisa começa a ganhar destaque na literatura nacional. As avaliações vão desde projetos-pilotos de prefeituras, como a abertura de escolas nos finais de semana, até programas e políticas de grande vulto, como a Olimpíada de Matemática, a municipalização de escolas, a educação de Jovens e Adultos e até a própria política de educação integral, como se observa no Quadro 2.

Quadro 2: Revisão da literatura – Avaliação de programas e políticas de educação

REFERÊNCIA	BASE DE DADOS	TIPO DE ESTIMAÇÃO	ADMIST./REGIÃO	SÉRIE	VARIÁVEL DE RESULTADO	VARIÁVEIS DE CAUSA/ IMPACTO
D'atri (2007)	Censo Escolar de 1998 e 2004	Diferenças em diferenças	Pública municipal e estadual / Brasil	EF	Desempenho escolar (medido por taxas de aprovação, reprovação, abandono e a distorção idade-série)	<u>Positivo</u> : política de municipalização (obs.: magnitude pequena)
Loureiro et al. (2008)	PNADs ¹ de 2002 a 2006	Variáveis Instrumentais	Pessoas que responderam ao PNAD	-	Condição de saber ler e escrever	<u>Positivo</u> : participação em cursos de Educação de Jovens e Adultos
Cedeplar e Fundação Itaú (2008)	Pesquisa de Campo em 2007	Regressão; grupo de controle	Pública municipal /Belo Horizonte-MG	-	Indicadores educacionais e culturais; indicadores gerais de bem-estar; socialização e convivência; alocação do tempo da criança e de seu responsável	<u>Positivo</u> : participação no programa Escola Integrada (obs.: impactos positivos do programa em quase todas as dimensões analisadas)
Biondi, Vasconcellos e Menezes-Filho, 2009	Prova Brasil de 2005 e 2007; Censo Escolar de 2007 e IBGE ²	PSM e Mínimos quadrados ponderados (MQP) pelo propensity score	Pública/ Brasil	8ª série	Proficiência em matemática	<u>Positivo</u> : participação no programa Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP)
Vasconcellos, Biondi e Menezes-Filho (2009)	Prova Brasil de 2005 e 2007; Censo Escolar de 2006; Censo Populacional (2000) e Cadastro do programa	PSM e Diferenças em Diferenças	Pública/ Brasil	4ª série	Proficiência em português	<u>Positivo</u> : participação no programa Escrevendo o Futuro
Duenhas, França e Gonçalves (2010)	Censo Escolar e Prova Brasil de 2007	PSM MQP ponderado pelo propensity Score	Pública municipal /Curitiba-PR	4ª e 8ª séries	Percepção do diretor sobre diversas formas de violência: a) diversas violências internas; b) diversas violências externas	Participação no programa Comunidade Escola; Impacto do programa: <u>positivo</u> para b e <u>negativo</u> para a

Silva (2010)	Sistema de avaliação / Instituto Itaú	Mínimos Quadrados Ordinário (MQO)	Pública estadual/ RS e MG		Proficiência em matemática e português	<u>Positivo</u> : participação no programa Projeto Jovem de Futuro
França e Duenhas (2011)	Censo Escolar e Prova Brasil de 2005 e 2007	Diferenças em Diferenças	Pública municipal /Curitiba-PR	4ª e 8ª séries	Proficiência em matemática e português	Programa: Comunidade Escola. Impacto do programa: <u>Negativo</u> e não sig. para a e <u>negativo</u> e sig. para b
Aquino e Kassouf (2011)	SARESP ³ de 2007 e 2008	Modelo de efeito fixo e PSM	Pública estadual / SP	8ª série	Proficiência em matemática e português	Participação no programa Escola de Tempo Integral; Desempenho em matemática (não significativo); <u>Positivo</u> : desempenho em português (obs.: magnitude pequena)

Notas: 1. PNAD - Pesquisa Nacional por Amostragem de Domicílio; 2. IBGE - Instituto Brasileiro de Geografia e Estatística; 3. SARESP -Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo.

Verifica-se, conforme os dados do Quadro 2, que as pesquisas de avaliação de impacto utilizam, em grande medida, as técnicas de diferenças em diferenças e PSM. Um dos aspectos mais importante em alguns estudos é a identificação de impacto negativo ou estatisticamente não significativo de alguns programas (D'ATRI, 2007; DUENHAS; FRANÇA; GONÇALVES, 2010; FRANÇA; DUENHAS, 2011; AQUINO; KASSOUF, 2011). A constatação de que o programa não gerou impacto sobre a variável de interesse está condizente com a função da avaliação de impacto, que é a produção da informação para subsidiar a tomada de decisão dos gestores públicos.

Em relação aos programas de educação integral, um estudo do CEDEPLAR em parceria com a Fundação Itaú (CEDEPLAR; FUNDAÇÃO ITAÚ, 2008) encontrou um impacto positivo do programa Escola Integrada de Belo Horizonte/MG em quase todas as dimensões analisadas. Contudo, Aquino e Kassouf (2011) analisaram o programa Escola de Tempo Integral do estado de São Paulo. Para os anos analisados (2007 e 2008), o programa não impactou o desempenho em matemática (não significativo estatisticamente). Já para o desempenho em português, o programa teve impacto, porém de magnitude muito pequena.

Em síntese, buscou-se com a revisão da literatura a identificação e a conceituação das diferentes técnicas de avaliação de programas e políticas. Dentre as mencionados, estão a avaliação de impacto, ou avaliação de efetividade objetiva, que foi utilizada para avaliação do “Programa Mais Educação” nessa dissertação. A avaliação de impacto apresenta um crescente desenvolvimento nos últimos 35 anos, tendo o seu início deflagrado com as avaliações de programas de treinamento nos Estados Unidos. Por fim, buscou-se também apresentar os resultados dos principais trabalhos empíricos desenvolvidos no país sobre os determinantes da qualidade da educação. Entre as três linhas de pesquisa apresentadas, estão os trabalhos de avaliação de políticas e programas educacionais, que embora incipientes no Brasil, começam se multiplicar.

A seguir, no próximo capítulo, apresenta-se a metodologia utilizada para a avaliação do impacto do “Programa Mais Educação” sobre o desempenho escolar dos estudantes do Rio Grande do Sul. Descrevem-se informações sobre o banco de dados, as variáveis e os procedimentos utilizados na pesquisa.

3. METODOLOGIA E BANCO DE DADOS

Nessa seção, apresenta-se a metodologia utilizada para a avaliação do impacto do “Programa Mais Educação”. Primeiramente, formaliza-se o procedimento de *propensity score matching* (PSM) e o método de diferenças em diferenças. Após essa apresentação, descreve-se a estratégia utilizada para a implementação dessas ferramentas na pesquisa de avaliação de impacto do “Programa Mais Educação”.

Na parte final do capítulo, as informações sobre os bancos de dados e as variáveis utilizadas na pesquisa, além da amostra específica para o Rio Grande do Sul são apresentados. Descreve-se também um esboço da estrutura e do funcionamento do programa estudado.

3.1 Modelo econométrico para avaliação de impacto

O procedimento de *propensity score matching* e o método de diferenças em diferenças são as técnicas mais utilizadas na literatura para a estimação do efeito tratamento de um programa ou política educacional (D’ATRI, 2007; BIONDI; VASCONCELLOS; MENEZES-FILHO, 2009; VASCONCELLOS; BIONDI; MENEZES-FILHO, 2009; DUENHAS; FRANÇA; GONÇALVES, 2010; AQUINO; KASSOUF, 2011). Assim como na literatura, também foi utilizada na pesquisa as duas ferramentas para avaliação do impacto do “Programa Mais Educação” sobre o desempenho médio nas notas de proficiência das escolas do Rio Grande do Sul³⁴. Portanto, a seguir, são descritas as duas técnicas, bem como a estratégia analítica que foi seguida para a avaliação do programa.

³⁴ O uso do PSM é justificado pelo fato da seleção para a participação no programa não ser aleatória entre os grupos tratamento e controle. Já o método de diferenças em diferenças é utilizado quando se tem dados disponíveis antes e depois do programa. O método de diferenças em diferenças tem a vantagem de eliminar diferenças não observáveis entre os grupos de tratamento e controle que são invariantes no tempo dentro de cada grupo.

3.1.1 Estratégia de identificação

Para inferir o impacto de um programa de educação integral sobre o desempenho médio das escolas que participaram do programa, é necessário saber o que teria acontecido caso as escolas que participaram dele não tivessem participado. Dada a impossibilidade dessa observação na prática, essa questão ficou conhecida na literatura como o problema do contrafactual não observado (CALIENDO; KOPEINING, 2005) ou da inferência causal (HOLLAND, 1986 *apud* ANGRIST; PISCHKE, 2009).

Em termos de resultados potenciais, o problema do contrafactual não observado ou da inferência causal pode ser expresso a partir das seguintes nomenclaturas: Y_0 é o resultado potencial caso uma escola não participe do programa, Y_1 é o resultado potencial caso a escola tivesse participado do programa, $D = 1$ quando a escola participou do programa e $D = 0$ quando a escola não participou do programa. A partir disso, é possível observar $Y_1 | D = 1$ e $Y_0 | D = 0$, porém nunca $Y_0 | D = 1$.

Assim, como só conseguimos observar na prática para um conjunto de escolas $E[Y_1 | D = 1]$ e $E[Y_0 | D = 0]$, um viés pode surgir de uma comparação ingênua entre diferenças de médias entre as escolas participantes e as escolas não participantes do “Programa Mais Educação”, caso esses dois grupos sejam muito diferentes. Isto é, ao compararmos o que deveria ser observado $A = E[Y_1 | D = 1] - E[Y_0 | D = 1]$ com o que é observado de fato $B = E[Y_1 | D = 1] - E[Y_0 | D = 0]$, verifica-se que um viés de seleção surge da diferença entre esses dois termos: $B - A = E[Y_0 | D = 1] - E[Y_0 | D = 0]$. Para Angrist e Pischke (2009) este último termo, o viés de seleção, é a diferença média em Y_0 daqueles que foram e daqueles que não foram tratados. Segundo os autores, este viés pode ser tão grande em termos absolutos que pode mascarar completamente um efeito positivo ou negativo do tratamento.

Como forma de contornar o problema de viés de seleção, uma das alternativas descritas pela literatura é obter um grupo de controle mais parecido possível com o grupo de tratamento a partir da seleção de características observáveis. Utiliza-se a adição de controles de características observáveis que tornam os grupos de controle e tratamento diferentes e que afetam a variável de interesse (neste caso, o desempenho escolar). Uma das principais

técnicas utilizadas é o *propensity score matching*. A técnica pode ser utilizada através de um procedimento simples de *matching*, quando há um conjunto pequeno de variáveis de controle discretas, ou então, no caso de existir um conjunto grande de variáveis, através de um escore de propensão (*propensity score*), que sintetizaria as informações em um único número. Tal procedimento, segundo Rosenbaum e Rubin (1983), quando atendidas as hipóteses de identificação³⁵ geram estimadores muito próximos aos procedimentos de aleatorização. O *propensity score* reduz, portanto, a dimensionalidade das variáveis. A seguir descreve-se a técnica criada por Rosenbaum e Rubin (1983).

a) Metodologia do propensity score matching (PSM)

Conforme dito anteriormente, o método *propensity score matching* (PSM), descrito inicialmente por Rosenbaum e Rubin (1983) surge como alternativa para amenizar o problema de viés de seleção na medida em que é utilizado para construir um grupo de comparação estatístico (grupo de controle) composto por escolas com características semelhantes ao grupo de escolas tratadas (que participaram do programa), mas que não participaram do programa. Segundo Rosenbaum e Rubin (1983), o ajuste nas diferenças entre os grupos é realizado a partir de um conjunto grande de variáveis que determinariam a probabilidade condicional de receber o tratamento.

Em outros termos, a construção desse grupo de controle baseia-se em um modelo de probabilidade de participar do tratamento, utilizando características observáveis ou seleção em observáveis. Rosenbaum e Rubin (1983) definem *propensity score* $p(x)$ como a probabilidade condicional de atribuição para um particular tratamento dado um vetor de variáveis observadas (X). Os participantes são então combinados em função dessa probabilidade, ou da pontuação de *propensity score*, com os não participantes.

³⁵ As duas hipóteses são: independência condicional e suporte comum. A seguir as duas serão descritas.

Dessa forma, o efeito médio do tratamento do programa é calculado como a diferença na média dos resultados entre esses dois grupos. O efeito tratamento sobre os tratados (ATT), baseado em seleção em observáveis (X), pode ser escrito da seguinte forma:

$$E(Y_{1i} - Y_{0i} | X, D = 1) \tag{1}$$

Contudo, a validade do modelo vai depender da eliminação desse viés, sendo duas hipóteses são necessárias para que isso aconteça. Estas são conhecidas como hipóteses de ignorabilidade forte, e são descritas a seguir:

Teorema 1: independência condicional ou ignorabilidade do tratamento:

Isto é, dado um conjunto de covariáveis observáveis (X) que não são afetadas pelo tratamento, o resultado potencial de $Y_{1,0}$ é independente do tratamento designado. Em outras palavras $Y_{1,0} \perp D | X$, ou seja, $Y_{1,0}$ é independente em relação ao tratamento condicional às características observáveis. Essa suposição é também conhecida por *unconfoundedness* (ROSENBAUM; RUBIN, 1983)³⁶.

Teorema 2: apoio de suporte comum ou hipótese implícita de suporte comum:

Não existe valor de características observáveis (X) que se possa dizer com certeza a que grupo pertence a unidade observada. Em termos matemáticos:

$$0 < P(D = 1 | X) < 1 \tag{2}$$

A hipótese de suporte comum (teorema 2) nos diz que as observações do grupo tratamento têm uma distribuição de *propensity score* “semelhante” as observações do grupo de controle. (HECKMAN; LALONDE; SMITH, 1999). Segundo Khandker, Koolwal e Samad (2010) essa condição garante que as observações do grupo de tratamento tenham uma distribuição de *propensity score* “próxima” ao grupo de controle.

³⁶ Como solução é necessário ter um conjunto de dados pré-tratamento para apoiar a hipótese de independência condicional, permitindo um controle de características observáveis (X) que possam estar afetando a participação no programa (ROSENBAUM; RUBIN, 1983).

Dado a ignorabilidade forte, isto é, as hipóteses 1 e 2 são válidas³⁷, Rosenbaum e Rubin (1983) mostram um terceiro teorema:

Teorema 3: se o tratamento designado tem ignorabilidade forte dado X, então ele também tem ignorabilidade forte para qualquer propensity score, ou seja:

$$Y_{1,0} \perp D \mid p(x) \quad (3)$$

onde $p(x)$ é a probabilidade de ser tratado dado X (*propensity score*).

Supondo que a condição de ignorabilidade forte seja válida e que $p(x)$ seja o *propensity score*, Rosenbaum e Rubin (1983) mostram que a diferença esperada nas respostas observadas para dois grupos dado $p(x)$ é igual ao efeito de tratamento médio (ATE) dado $p(x)$, ou seja:

$$E[Y_1 \mid p(x), D = 1] - [Y_0 \mid p(x), D = 0] = E[Y_1 - Y_0 \mid p(x)] \quad (4)$$

Além disso, é possível estimar o efeito médio de tratamento sobre os tratados (ATT), como segue:

$$E_{p(x)}\{E[Y_1 \mid p(x), D = 1] - [Y_0 \mid p(x), D = 0] \mid D = 1\} = \{E[Y_1 - Y_0 \mid p(x)] \mid D = 1\} \quad (5)$$

O uso do $p(X)$ ao invés do (X) , descrito por Rosenbaum e Rubin (1983), é conhecido na literatura como de método de *propensity score*. Em suma esta metodologia permite a realização do pareamento (*matching*) entre participantes e não participantes, e a partir dela, a geração de estimativas confiáveis do efeito de tratamento médio.

Contudo, para proceder com o pareamento, a partir dessas suposições acima é necessário a realização de três procedimentos: (a) a estimação do modelo de participação do programa; (b) a definição da região de suporte comum e (c) o pareamento (*matching*) entre

³⁷ Essas duas hipóteses são conhecidas como ignorabilidade forte

participantes e não participantes (CALIENDO; KOPEING, 2005; KHANDKER; KOOLWAL; SAMAD, 2010).

O primeiro procedimento refere-se a estimação da probabilidade de participação sobre todas as covariáveis (X) observadas nos dados que são susceptíveis de determinar a participação. Quando se está interessado apenas em comparar os resultados para os participantes ($D = 1$) com aqueles que não participam ($D = 0$), isto é, um modelo de resposta binária, esta estimativa pode ser construída a partir de um modelo *logit* ou *probit* da participação no programa. Esse modelo pode ser representado da seguinte forma:

$$\hat{p}(x) = pr(D_i | X_i = x) \quad (6)$$

Em que a matriz (X) é o conjunto de características observáveis que determinam a participação no tratamento e também de seu resultado; $\hat{p}(x)$ é a probabilidade estimada de ser tratado dado (X) ou *propensity score* e D_i o tratamento, sendo $D = 1$ para aqueles que participaram do programa e $D = 0$, caso contrário. Nos modelos *logit* ou *probit*, a variável dependente será uma variável binária (*dummy*) para a participação ao programa (1 = para participantes do programa Mais Educação e 0 = caso contrário).

O segundo passo, diz respeito a definição da região de suporte comum, conforme o teorema 2. A área de suporte comum é definida através da sobreposição da distribuição de *propensity score* para os grupos de tratamento e controle. Dessa forma, testes de balanceamento devem ser realizados para verificar se os grupos de tratamento e comparação têm *propensity scores* similares para cada covariável (X) inserida no modelo (BECKER; ICHINO, 2002 e KHANDKER; KOOLWAL; SAMAD, 2010).

Uma revisão dos principais testes foi realizado por Lee (2006), sendo os principais testes descritos na literatura, os seguintes: teste DW (Dehejia e Wahba), teste padronizado de diferenças, teste para igualdade de médias antes e após o matching (t-testes) e teste de igualdade conjunta de médias em amostras *matching* (Teste de *Hotelling*). Em outras palavras, é preciso verificar se as distribuições entre os grupos são semelhantes, o que implica em equilíbrio ou balanceamento. Além disso, a eficácia do PSM também depende de uma grande e quase igual número de observações de participantes e de não participantes de modo que uma região importante de apoio comum possa ser encontrada.

Por fim, procede-se com o pareamento (*matching*) entre participantes e não participantes, isto é, constrói-se um grupo de comparação composto por unidades com características semelhantes ao grupo de tratamento. Diferentes procedimentos podem ser adotados para a realização do pareamento (*matching*). Os principais procedimentos descritos na literatura são os seguintes: *Nearest-neighbor*, *Caliper ou radius*, *Stratification ou interval*, *Kernel e local linear* (CAMERON; TRIVEDI, 1995; BECKER; ICHINO, 2002; CALIENDO; KOPEING, 2005; KHANDKER; KOOLWAL; SAMAD, 2010).

O primeiro deles, *Nearest-neighbor matching*, é um dos métodos mais utilizados e consiste no pareamento de cada unidade de tratamento com a unidade de comparação do grupo de controle de acordo com o *propensity score* mais próximo. Já no procedimento *Caliper ou radius matching*, cada unidade de tratamento só é comparável com as unidades de controle, cujos *propensity scores* situam-se em um limite pré-definido segundo o *propensity score* da unidade de tratamento.

O método de *Stratification ou interval matching*, por sua vez, consiste em dividir o intervalo da variação do *propensity score* em intervalos de tal forma que as unidades dentro de cada intervalo, tratado e controle, em média possuem o mesmo *propensity score*. Por fim, o método mais robusto é o método de *Kernel e local linear matching* em que todas as unidades tratadas são comparadas com a média ponderada de todas as unidades de controle com pesos definidos a partir da distância inversamente proporcionais entre os *propensity scores* das unidades tratadas e controle.

A literatura aponta (CALIENDO; KOPEING, 2005; HECKMAN; ICHIMURA, TODOD, 1998) que a técnica de Kernel leva vantagem em relação às outras, pois em relação a consistência e eficiência dos estimadores³⁸, os algoritmos de PSM por essa técnica tendem a ser mais eficientes que os outros em função do aproveitamento de toda a amostra. Nesse método, a escolha das escolas do grupo de controle não será dada apenas pela semelhança às escolas do grupo de tratamento, mas também serão inseridas todas as escolas pertencentes ao suporte comum (hipótese 2 do modelo). Isso significa que as escolas tratadas serão pareadas com uma média ponderada de todas as escolas do grupo de comparação, sendo os pesos alocados de forma inversamente proporcional à distância entre os escores de propensão dos grupos de tratamento e controle.

³⁸ Caliendo e Kopeing (2005) destacam a menor variância obtida através desse algoritmo.

Escolhido o grupo de controle, e após a realização de testes de balanceamento e a realização do pareamento, as médias dos resultados dos dois grupos podem ser comparadas a partir de um modelo de regressão.

Cabe ressaltar as vantagens e desvantagens do método PSM. As vantagens estão relacionadas ao fato da implementação da abordagem ser simples, pois há uma redução grande da dimensionalidade do conjunto das variáveis. Além disso, o uso do PSM permite a redução das diferenças entre os grupos a partir de covariáveis observáveis. Por outro lado, as desvantagens do método estão relacionadas à necessidade de se trabalhar com grandes amostras e da necessidade de haver sobreposição dos grupos de tratamento e controle. Biondi, Vasconcellos e Menezes-Filho (2009) destacam também que a principal crítica em relação ao método diz respeito ao fato de que o $p(x)$ ser desconhecido, e com isso, ao ser estimado, a variância do estimador do pareamento poderá ser afetada³⁹. Além disso, mesmo com a aplicação do método, o viés pode ainda persistir, pois o pareamento só ocorre para as variáveis observadas. Dessa forma, para minimizar esse viés que pode estar relacionado com características não observadas que afetam os resultados, um outro método, diferenças em diferenças, também foi utilizado na pesquisa.

b) Metodologia de diferenças em diferenças

O método de diferenças em diferenças (DD) se baseia na comparação dos participantes e não participantes antes e depois da implementação de um programa. O método é utilizado quando há características não observáveis heterogêneas entre os grupos que influenciam a participação do programa, mas que não variam ao longo do tempo⁴⁰, como por exemplo, as diferentes habilidades ou personalidade entre alunos, professores e diretores de escolas participantes do programa e escolas não participantes.

³⁹ Segundo os autores: “Dessa forma, não é possível conhecer a distribuição assintótica do pareamento fazendo com que os erros-padrão dos estimadores não sejam confiáveis” (BIONDI, VASCONCELLOS e MANEZES-FILHO, 2009, p.7)

⁴⁰ Admite-se também que as variáveis não observadas são não correlacionadas com o tratamento ao longo do tempo.

O método DD é descrito a seguir. Dados dois períodos no tempo, em que $t = 0$ é o período antes do programa e $t = 1$ é o período após a execução do programa, sendo que o Y_t^T é o resultado para os participantes do programa e Y_t^C é o resultado para os não participantes do programa, no tempo t , a partir do método DD é possível estimar o impacto médio do programa da seguinte forma:

$$DD = E(Y_1^T - Y_0^T | T_1 = 1) - E(Y_1^C - Y_0^C | T_1 = 0) \quad (7)$$

O método DD também pode ser expresso por um modelo de regressão linear. O modelo de regressão linear pelo MQO é descrito da seguinte maneira:

$$Y_i = \alpha + \beta D_i + \gamma t_i + \delta (D_i \circ t_i) + \varepsilon_i \quad (8)$$

onde Y_i = desempenho médio na proficiência em português ou matemática da 4ª ou 8ª série, $i = 1, \dots, N$ é o número de escolas; $D = 0$ indica que a escola i não recebeu o tratamento (grupo de controle) e $D = 1$ indica que a escola i recebeu o tratamento (grupo tratamento); α , β , δ são parâmetros desconhecidos a serem estimados; ε erro aleatório (não observados). Esses coeficientes podem ser interpretados da seguinte forma⁴¹:

α = termo constante;

β = efeito específico do grupo de tratamento (contabiliza diferenças fixas nos tempos entre o grupo de tratamento e controle);

γ = tendência temporal comum aos dois grupos;

δ = efeito médio do tratamento, isto é, o parâmetro de interesse no presente estudo.

O efeito médio do tratamento também pode ser escrito como segue abaixo:

$$\hat{\delta} = (\bar{Y}_1^T - \bar{Y}_0^T) - (\bar{Y}_1^C - \bar{Y}_0^C) \quad (9)$$

Onde C é o grupo de controle e T corresponde ao grupo de tratamento. A estimativa do efeito do tratamento $\hat{\delta}$ não será viesada, isto é, $E[\hat{\delta}] = \delta$, sob as seguintes hipóteses: a) a equação (equação 8) for corretamente especificada; b) erros aleatórios têm média zero; e b)

⁴¹ Vasconcellos, Biondi e Menezes-Filho (2009)

os termos de erros não são correlacionados com as outras variáveis da equação⁴² (VASCONCELLOS; BIONDI; MENEZES-FILHO, 2009 e KHANDKER; KOOLWAL; SAMAD, 2010).

A equação (8) para uma determinada escola em $t = 1$ e em $t = 0$ pode ser expressa da seguinte forma⁴³:

$$Y_1 = \alpha + \beta D + \gamma t_1 + \delta(D^\circ t_1) + \varepsilon_1 \quad (10)$$

$$Y_0 = \alpha + \beta D + \gamma t_0 + \delta(D^\circ t_0) + \varepsilon_0 \quad (11)$$

Assim, com a subtração da equação (11) de (10), temos que:

$$Y_1 - Y_0 = \gamma(t_1 - t_0) + \delta D(t_1 - t_0) + (\varepsilon_1 - \varepsilon_0) \quad (12)$$

$$Y_1 - Y_0 = \gamma + \delta D + \mu \quad (13)$$

Onde $\mu = (\varepsilon_1 - \varepsilon_0)$

A estimação da equação 13 é equivalente a estimação da equação 8.

Contudo, Vasconcellos, Biondi e Menezes-Filho (2009) destacam que para a recuperação da tendência paralela de Y entre o grupo de tratados e controle, é necessário incluir as diferenças ou variações de médias entre $t = 0$ e $t = 1$ das características socioeconômicas média dos alunos das escolas, isto é, é necessário incluir uma matriz de características observáveis X disponíveis em $t = 1$ e $t = 0$. Dessa forma, com a inclusão das características socioeconômicas médias dos alunos das escolas em $t = 0$ e $t = 1$, a equação a ser estimada é apresentada a seguir:

⁴² H1: $cov(\varepsilon_{it}, D_{i1}) = 0$,

H2: $cov(\varepsilon_{it}, t) = 0$

H3: $cov(\varepsilon_{it}, D_{i1}t) = 0$

Segundo Vasconcellos, Biondi e Menezes-Filho (2009) essa última hipótese é conhecida como hipótese tendência paralela. Isso significa que a variável Y e as características não observáveis devem seguir a mesma tendência no tempo para o grupo de tratamento e controle, o que garante que a diferença entre dois grupos esteja refletindo somente o efeito médio do tratamento. Já Khandker, Koolwal e Samad (2010) destacam que essa suposição é a mais crítica de todas e significa que características não observáveis que afetam a participação no programa não variam ao longo do tempo com o *status* tratamento.

⁴³ Vasconcellos, Biondi e Menezes-Filho (2009).

$$Y_1 - Y_0 = \gamma + \delta D + \mu(X_1 - X_0) + u \quad (14)$$

O método DD se destaca, pois é possível com ele, relaxar a suposição de exogeneidade condicional ou a seleção unicamente em características observadas, fornecendo uma forma adicional para dar conta de características não observadas. Além disso, Khandker, Koolwal e Samad (2010) ressaltam que o uso conjunto do DD com o PSM pode ajudar a resolver o problema de viés de seleção. Ademais, o uso em conjunto com o PSM possibilita também a estimação através de uma regressão ponderada (pelo PSM), o que auxilia na correção de eventuais distorções no modelo. Para Hirano, Imbens e Ridder (2003), a ponderação das observações do grupo de controle de acordo com seu *propensity score* pode produzir um estimador plenamente eficiente. O ponderador pode ser calculado da seguinte forma⁴⁴:

$$W_{ATE_i} = \frac{D_i}{\hat{p}(x_i)} + \left(\frac{1-D_i}{1-\hat{p}(x_i)} \right) \quad (15)$$

$$W_{ATT_i} = W_{ATE_i} * \hat{p}(x)_i \quad (16)$$

3.1.2 Implementação da avaliação do programa Mais Educação

A implementação da análise de impacto do “Programa Mais Educação” sobre as notas de 4ª e 8ª série foi realizada em duas etapas. Primeiramente, para cada amostra, foi estimada a probabilidade de participação no programa (*propensity score*) com um modelo *logit* (equação 6). Posteriormente, em uma segunda etapa, procedeu-se com a estimação do método de diferenças em diferenças. O modelo estimado compreende a equação 8, com a adição das variáveis de controles, antes e depois do programa.

a) Estimando o propensity score

Para o desenvolvimento do primeiro passo, ou seja, a aplicação do método de estimação do *propensity score*, foi estimada uma equação da probabilidade de participação do

⁴⁴ Disponível em : <http://www.undp-povertycentre.org/evaluation/praticas/do/> . Acesso em 04/05/2010.

programa através do modelo *logit*. Para a estimação dessa equação, uma amostra de participantes e não participantes foi reunida, e então, a probabilidade de participação foi estimada a partir do conjunto das variáveis (*X*) observadas.

No modelo *logit*, a variável dependente é uma variável *dummy* para a participação ao programa (1 = para escola participante do programa e 0 = caso contrário). Ao estimar a probabilidade da participação no programa, utilizou-se como variáveis explicativas as condições iniciais (pré-tratamento) das escolas, alunos, diretores, municípios e dos critérios pré-estabelecidos pelo MEC, isto é, todas as variáveis do ano de 2007 (ver quadro 4: variáveis de controle, na seção 3.3). Isto é, as variáveis que foram incluídas no modelo *logit* estão relacionadas às características socioeconômicas média dos alunos, características dos professores, dos diretores e da infraestrutura, dos municípios e aos pré-requisitos para a adesão ao programa para o ano de 2007, ou seja, variáveis pré-tratamento⁴⁵. Essas variáveis foram selecionadas devido ao fato de que elas poderiam influir tanto na participação no “Programa Mais Educação” quanto na variável de resultado (conforme a revisão da literatura no capítulo 2), ou seja, no desempenho de alunos em testes de proficiência.

Ressalta-se que a escolha do modelo de probabilidade de participação no programa, com o conjunto das variáveis (*X*) observadas mais adequadas foi realizado a partir do teste LR e dos critérios de informações de critérios de informações de Akaike - AIC e Schwarz - SIC. Dessa forma, estimou-se vários modelos com a adição de blocos de características (bloco socioeconômico dos alunos, características das escolas, dos municípios) e comparou-se com o modelo mais geral.

Após a estimação do modelo de *logit*, aplicou-se também testes de diferenças de médias⁴⁶ para verificar se o conjunto de variáveis *X* estavam balanceadas entre os grupos. Para isso, dividiu-se a amostra em estratos ou blocos de valores estimados do *propensity score*. Este procedimento permitiu uma maior eficiência na estimativa do *propensity score*, tornando o balanceamento entre os grupos mais homogêneo.

⁴⁵ Caso uma variável que determina a participação não seja incluída na equação de participação, o PSM será viesado (IMBENS; WOOLDRIDGE, 2008 e 2009). Assim, durante a construção do modelo deve-se dar atenção às características que determinam a participação no programa específico a ser avaliado, para preservar a validade do modelo.

⁴⁶ Os testes de balanceamento utilizados para a verificação do equilíbrio existente entre cada componente ou variável independente nos grupos de tratamento e controle foram: o t-teste de médias e o teste de igualdade de distribuições de Kolmogorov-Smirnov antes do pareamento e o t-teste após a estimação do *propensity score* e a definição dos blocos.

Ademais, para contornar o viés que poderia surgir quando o grupo de controle é escolhido de forma inadequada, foram realizados testes de balanceamento⁴⁷ em que algumas escolas caíram fora da região de suporte comum e, portanto, acabaram sendo excluídas das amostras. Para definição da região de suporte comum, foi utilizado o critério de “comparação mínima e máxima” (CALIENDO; KOPEING, 2005).

Em suma, após a estimação do modelo de probabilidade de participação do modelo e com a obtenção do *propensity score*, foi possível realizar testes de balanceamento, o que resultou em grupos de controle e tratamento balanceados e dentro da região de suporte comum.

b) Estimação por diferenças em diferenças

Após a estimação do *propensity score*, utilizou-se o método das diferenças-em-diferenças (DD). O método tem a vantagem de eliminar efeitos fixos de variáveis não observáveis e que não variam ao longo do tempo dentro dos grupos de controle e tratamento. Ou seja, através do método, eliminam-se as diferenças heterogêneas de variáveis não observadas nos dois grupos. Ressalta-se que sem a consideração dessas variáveis não observáveis, a estimação poderá conter viés (KHANDKER; KOOLWAL; SAMAD, 2010).

As variáveis dependentes utilizadas no método de DD, foram os resultados nos testes de proficiência do antes (2007) e pós tratamento (2009) da Prova Brasil em língua portuguesa e matemática. Além disso, como variável de impacto, para cada escola foi observada a participação no programa, em que $D = 1$ indica a participação e $D = 0$ caso contrário. Como variáveis de controle, incluiu-se as variáveis socioeconômicas dos alunos, das escolas e dos municípios dessas escolas. Assim, a lista de controles utilizados no modelo de diferenças em diferenças compreende as características de 2007 e 2009 (ver quadro 4: variáveis de controle, na seção 3.3). O modelo estimado compreende a equação 8, com a adição das variáveis de controle. Entretanto para dar maior robustez, a equação foi ponderada pelo inverso do peso dado na equação 16.

Com esses procedimentos foi possível obter as estimativas de impacto do “Programa Mais Educação” no desempenho em língua portuguesa e matemática dos alunos da 4ª série e

⁴⁷ Uma revisão completa dos testes de balanceamento pode ser encontrada em Lee (2006).

da 8ª série das escolas estaduais e municipais do Rio Grande do Sul. Ressalta-se que houve a necessidade de controlar o número de participações (há escolas que participaram das duas edições - 2008 e 2009 e escolas que participaram somente em 2009). Ou seja, a hipótese de que o maior tempo de exposição ao programa das escolas traz um maior impacto também foi testada.

Em relação a análise de robustez, realizou-se dois procedimentos. Primeiramente foi testado, a partir da construção de um grupo de tratamento “placebo”, se o programa apresenta impacto sobre a proficiência em português e matemática para a 4ª e 8ª séries. Isto é, simulou-se se uma falsa participação no programa seria capaz de impactar o desempenho escolar. O grupo de tratamento “placebo” foi constituído a partir de escolas que entraram no programa somente em 2010, ou seja, posterior aos desempenhos escolares de 2009. Assim, a expectativa era que o impacto do programa não fosse significativo estatisticamente quando esse grupo de tratamento fosse utilizado para a avaliação de impacto do “Programa Mais Educação”.

Por fim, verificou-se visualmente se as escolas do grupo de tratamento e controle utilizadas na pesquisa seguiam tendências paralelas antes da implementação do programa nos desempenhos da prova Brasil para a 4ª e 8ª séries em língua portuguesa e matemática. Ou seja, construiu-se gráficos dos desempenhos médios das escolas em 2005 e 2007 a fim de se verificar as trajetórias das notas dessas escolas antes do “Programa Mais Educação”. Caso essas trajetórias sejam bem distintas, é bem provável que outros fatores estejam influenciando as escolas de forma diferenciada, e sendo assim, não seria possível isolar o efeito do programa.

3.2 Base de dados

As bases de dados utilizadas na pesquisa foram os microdados da Prova Brasil 2007 e 2009 e Censo Escolar 2007 e 2009, coordenados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP⁴⁸. As informações sobre a participação das escolas no programa Mais Educação nos anos de 2008 e 2009 foram obtidas junto ao Ministério da Educação – MEC e Secretaria Estadual da Educação do Rio Grande do Sul-

⁴⁸ Do site do INEP, obteve-se também o IDEB e a taxa de aprovação na série para os anos de 2007 e 2009.

SEDUC. Adicionalmente, contou-se também com informações sobre os municípios (2007 e 2009). Esta última base de dados foi obtida no sítio do IBGE.

A Prova Brasil é um exame padronizado de língua portuguesa (leitura) e matemática (resolução de problemas) organizado a cada dois anos pelo INEP/MEC. A primeira edição ocorreu em 2005, e depois disso, mais duas edições foram realizadas: em 2007 e 2009. O exame é aplicado a estudantes de 4ª e 8ª séries⁴⁹ escolas públicas do Ensino Fundamental (EF) com pelo menos 20 alunos matriculados na série (MEC, 2009e; 2011).

Os testes de desempenho da Prova Brasil são elaborados a partir da Teoria de Resposta ao Item (TRI)⁵⁰. Esta metodologia tem a vantagem de permitir a comparação de notas entre os estudantes ao longo do tempo e de diferentes séries (BIONDI; VASCONCELLOS; MENEZES-FILHO, 2009). Paralelamente aos testes de desempenho em português e matemática, são aplicados também questionários socioeconômicos e a fatores associados a esses resultados (MEC, 2009e). A investigação concentra-se nas informações relativas aos alunos, escola, diretores e docentes.

O Censo Escolar, coordenado também pelo INEP, é realizado anualmente nas escolas de educação básica de todo o país (privadas ou públicas). O levantamento tem como objetivo a promoção de uma ampla coleta de informações sobre a situação das escolas nos quesitos escola (infraestrutura física, dependências e equipamentos), docentes (escolaridade, turma, disciplinas) e turmas (tipo e turma, horários e disciplinas).

Cabe destacar que a partir dos dados do Censo Escolar, o MEC gerencia uma série de programas como o FUNDEB, a Alimentação Escolar, a distribuição de Livros Didáticos, o Dinheiro Direto na Escola - PDDE, e inclusive, o “Programa Mais Educação”. Além disso, as informações também servem para o cálculo de vários indicadores como o IDEB, as taxas de rendimento e de fluxo escolar, distorção idade-série, entre outros. O Censo Escolar, segundo o MEC⁵¹, é uma ferramenta importante para o conhecimento sobre a situação da educação no Brasil bem como para o acompanhamento da efetividade das políticas públicas.

⁴⁹ Em 2009, as provas foram aplicadas nas 4ª e 8ª séries do ensino fundamental regular de 08 anos e, nas escolas organizadas no regime de 09 anos para o ensino fundamental regular, em turmas de 5º e 9º anos (MEC, 2011).

⁵⁰ Mais informações sobre essa metodologia pode ser encontrada em TAVARES (2001) e KLEIN (2003).

⁵¹ Fonte: <http://inep.gov.br/web/educacenso/situacao-do-aluno>. Acesso em 14/12/2011.

3.3 Variáveis

A unidade de análise será as escolas estaduais e municipais do Rio Grande do Sul. Como o objetivo desse trabalho foi analisar o impacto do “Programa Mais Educação”, implantado a partir de 2008, no desempenho escolar médio dos alunos das escolas da rede pública municipal e estadual do Rio Grande do Sul, as variáveis indicadoras de resultado selecionadas para a pesquisa foram as notas da proficiência média das escolas em Língua Portuguesa e em Matemática das 4ª séries e 8ª séries da Prova Brasil, em 2007 e 2009, conforme o quadro 3.

Quadro 3: Variáveis indicadoras de resultado (dependentes)

Variável	Descrição
Informação da Prova Brasil 2007 e 2009	
Nota_4a_MAT	Nota média da escola em Matemática (4ª série)
Nota_4a_PORT	Nota média da escola em Português (4ª série)
Nota_8a_MAT	Nota média da escola em Matemática (8ª série)
Nota_8a_PORT	Nota média da escola em Português (8ª série)

Fonte: Elaboração própria

Ressalta-se que não foi possível obter informações sobre a participação individual e a intensidade desta participação de cada aluno no programa. Sabe-se apenas que o MEC (2009b) recomenda, para o atendimento prioritário, a escolha por parte das escolas de estudantes das séries finais da 1ª fase do ensino fundamental (4º e 5º anos) e de estudantes das séries finais da 2ª fase do ensino fundamental (8º e 9º anos) e que no Rio Grande do Sul, a adesão dos alunos de todas as séries das escolas participantes ocorreu em massa⁵². Diante da dificuldade de avaliação individual e da suposição de que as escolas tenham no mínimo seguido a orientação do MEC, utilizou-se para análise o desempenho médio dos alunos das escolas, expressos pelas notas médias na Prova Brasil para a 4ª e 8ª séries.

As variáveis de controle utilizadas para a pesquisa foram obtidas através dos microdados da Prova Brasil 2007 e 2009, do Censo Escolar 2007 e 2009, do IBGE (2007 e 2009) e do INEP/MEC. Estas variáveis ajudaram a compor tanto o modelo *logit* (com a

⁵² Conforme relato de gestores da SEDUC-RS em 24/02/2012.

geração do *propensity score*) quanto o modelo de diferenças em diferenças para a análise do impacto. Em relação à Prova Brasil, contou-se com informações referentes aos questionários dos diretores e alunos. Já do Censo Escolar, algumas variáveis dos questionários da turma, docentes e escola também foram selecionadas. Adicionalmente, as informações do IBGE no ano de 2007 e 2009 referentes a características demográficas, de localização e de PIB per capita dos município foram incluídas para controle. Do sítio do INEP/MEC também foi possível obter o IDEB⁵³ e taxa de aprovação na série para cada escola (2007 e 2009) e o IDEB do município correspondente a escola (2007 e 2009).

Destaca-se também que as variáveis de controle selecionadas foram aquelas que apresentaram alguma magnitude nos testes de proficiência no Brasil. Em outros termos, utilizou-se como referência para a seleção das variáveis no modelo, aquelas em que a literatura (capítulo 2) apontou como relevante no desempenho escolar.

Paralelamente a isso, buscou-se também selecionar variáveis que o MEC utilizou para a indicação das escolas a que participariam do “Programa Mais Educação” nos anos de 2008 e 2009. Como será visto mais adiante, o programa não se configura como um experimento aleatório. Isso porque o MEC estabeleceu alguns critérios de seleção das escolas que poderiam ser atendidas pelo programa e as escolas que atenderam aos critérios de elegibilidade optaram ou não por participar do programa. Dessa forma, dadas as especificidades do “Programa Mais Educação”, essas variáveis utilizadas como critério de seleção pelo MEC também foram incluídas nos modelos estimados. Cabe destacar que estas variáveis afetam fortemente a probabilidade de participação no programa⁵⁴. A seguir, são descritas todas as variáveis de controle utilizadas nos modelos.

⁵³ O Índice de Desenvolvimento da Educação - IDEB foi criado em 2007 para medir a qualidade de cada escola pública do país. O indicador é calculado com base no desempenho do estudante em avaliações da Prova Brasil e em taxas de aprovação. Dessa forma, o indicador de desenvolvimento da educação mostra que para se avançar na qualidade da educação é preciso que o aluno aprenda, não repita o ano e frequente a sala de aula. O índice é medido de dois em dois anos. A meta estabelecida pelo MEC para o país, a partir do alcance das metas municipais e estaduais, é alcançar a nota 6 em 2022 – correspondente à qualidade do ensino em países desenvolvidos.

Fonte: http://portal.mec.gov.br/index.php?option=com_content&view=article&id=180&Itemid=286. Acesso em 03/03/2011.

⁵⁴ Essas variáveis são fundamentais para não violação das hipóteses de independência condicional e de suporte comum.

Quadro 4: Variáveis de controle

Variável	Descrição
1. Informação do Censo Escolar 2007 e 2009	
1.1 Questionário da turma	
MÉDIA_AULA	Duração média de funcionamento das turmas (minutos)
MED_MAT_SERIE	Média de matrículas por turma na série
TOTAL_MAT	Total de matrículas na série
1.2 Questionário Escola	
BIBLIO	Escola possui biblioteca/sala de leitura (1=sim e 0= não)
INTER	Escola possui acesso a internet (1=sim e 0= não)
LAB_INFO	Escola possui Laboratório de informática (1=sim e 0= não)
1.3 Questionário Docentes	
PROF_SUPER	% de professores na série com ensino superior (1=sim e 0= não) na série
PROF_POS	% de professores na série com pós-graduação na série
PROF_FUND	Número de professores no ensino fundamental
2. Informação da Prova Brasil 2007 e 2009	
2.1 Questionário dos alunos (4ª série ou 8ª série)	
SEXO	% alunos homens na série
RACA	% de alunos brancos na série
MAQ_LAV	% de alunos que têm máquina de lavar na série
CAR	Numero médio de automóveis que possui na série
INTER_CASA	% de alunos que têm computador com internet em casa na série
DEVER	% de alunos que têm pais que incentivam fazer o dever na série
LER	% de alunos que têm pais que incentivam a leitura na série
NÃO_FALTAR	% de alunos que têm pais que incentivam a não faltar na série
MAE_MEDIO	% de alunos com mães que completaram o ensino médio na série
MAE_SUPER	% de alunos com mães que completaram o ensino superior na série
MAE_NAORES	% de alunos que não responderam sobre a escolaridade da mãe na série
TRAB	% de alunos que trabalham fora de casa na série
REUN	% com pais que compareceram à reunião na série
PRE_ESC	% alunos que iniciaram os estudos na pré-escola na série
2.2 Questionário dos diretores	
DIR_29	Diretor com 29 anos de idade ou menos(1=sim e 0= não)
DIR_30_a_39	Diretor com 30 a 39 anos de idade (1=sim e 0= não)
DIR_40	Diretor com mais de 40 anos de idade (1=sim e 0= não)
DIR_POS	Diretor possui pós-graduação completa (1=sim e 0= não)
DIR_11_a_15	Diretor com 11 a 15 anos na direção da escola (1=sim e 0= não)
DIR15	Diretor com mais de 15 anos na direção da escola (1=sim e 0= não)
PRO_SEL	Escola tem prova para seleção de alunos
INT_ATI	Escola com alta interrupção das atividades escolares por problema grave
ROT_PROF	Escola com alta rotatividade do corpo docente
REF_ESC	Escola possui programa de apoio ou reforço de aprendizagem (1=sim e 0= não)
DIR_SEL_ELE	Diretor passou por seleção e eleição (1=sim e 0= não)
PROG_FED	Escola participa de programa de fin. do Gov. Federal? (1=sim e 0= não)
PROG_EST	Escola participa de programa de fin. do Gov. Estadual? (1=sim e 0= não)

PROG_MUN	Escola participa de programa de fin. do Gov Municipal? (1=sim e 0= não)
3. Informações adicionais do IBGE –	
POP	População no município da Escola
PIB_PER	PIB <i>per capita</i> no município (R\$) da Escola
RM_OU_POLO	Escola está inserida na região metropolitana (1=sim e 0= não)
4. Informações adicionais do MEC e INEP	
IDEB_esc	IDEB da escola (2007 e 2009) na série
IDEB_mun	IDEB do município em 2007
Dif_IDEB	Diferença entre o IDEB (das escolas públicas) do município em 2007 e IDEB da escola
PRONASCI	Município da escola era integrante do PRONASCI em 2007 (1=sim e 0= não)
TAXA_APRO	Taxa de aprovação (2007 e 2009) na série
MUNIC	Escola pertence a administração pública municipal (1=sim e 0= não)

Fonte: Elaboração própria

A variável de impacto é uma variável binária referente à participação ou não da escola no “Programa Mais Educação”. A investigação da pesquisa está centrada nas escolas que participaram em algum momento do programa (amostra 1). Isto é, independente do ano de início⁵⁵. Contudo, uma amostra contendo somente as escolas que participaram do programa em 2009 (amostra 2) também foi analisada a fim de comparar se o tempo de exposição gera impacto diferenciado nas notas médias das escolas.

Quadro 5: Variáveis de causa

Amostra	Variável	Descrição
1	TRAT	Escolas que participaram do programa em algum ano (2008 ou 2009) (1=sim e 0= não)
2	TRAT	Escolas que participaram do programa somente em 2009 (1=sim e 0= não)

Fonte: Elaboração própria

⁵⁵ Excluiu-se da amostra, aquelas que iniciaram em 2008 e interromperam o programa. Foram encontradas quatro escolas para esse caso no Rio Grande do Sul.

3.4 Descrição do “Programa Mais Educação”

O programa federal Mais Educação foi instituído em 2007 através da Portaria Interministerial n° 17 de 24 de abril de 2007 e implementado, posteriormente, no ano de 2008 em escolas da rede pública municipal e estadual de todo o Brasil. O programa é uma contribuição para a formação integral de crianças, adolescentes e jovens. Nesse sentido, ele constitui-se como uma estratégia de iniciativa do governo federal para a ampliação da jornada escolar nas escolas públicas brasileiras, com a organização dos currículos escolares na perspectiva da educação integral e de uma carga horária mínima de sete horas diárias⁵⁶.

Para fornecer a formação integral, o MEC mobiliza os alunos das escolas integrantes do “Mais Educação” tendo em vista: (a) a melhoria do desempenho educacional; (b) o cultivo de relações entre professores, alunos e suas comunidades; e (c) a garantia da proteção social da assistência social e à formação para a cidadania (artigo 1º da Portaria n° 17 de 24/04/2007). Assim, a contribuição dessa dissertação está na avaliação do impacto do programa quanto ao seu primeiro objetivo. Ou seja, a melhoria do rendimento escolar é o que está sendo avaliada nesse estudo a partir das análises dos testes de proficiência do Rio Grande do Sul.

Cabe destacar, além dos objetivos mais gerais, as principais finalidades propostas pelo “Programa Mais Educação”. Estas são as seguintes: (a) apoiar com recursos financeiros, a ampliação do tempo e do espaço educativo e a extensão do ambiente escolar nas redes públicas de educação básica, mediante a realização de atividades no contra-turno escolar, articulando ações desenvolvidas pelos ministérios integrantes do programa; (b) prevenir e combater o trabalho infantil, a exploração sexual e outras formas de violência contra crianças; (c) contribuir para a redução da evasão, da reprovação, da distorção idade/série, mediante a implementação de ações pedagógicas para melhoria de condições para o rendimento e o aproveitamento escolar (artigo 2º da Portaria n° 17 de 24/04/2007).

Para atingir esses objetivos e essas finalidades, o Programa “Mais Educação” é operacionalizado através da Secretaria de Educação Continuada, Alfabetização e Diversidade (SECAD/MEC) e em parceria com a Secretaria de Educação Básica (SEB/MEC), por meio do

⁵⁶ Segundo o MEC (2009d) nessa carga hora não se contabiliza o intervalo de almoço.

programa Dinheiro Direto na Escola (PDDE) e do Fundo Nacional de Desenvolvimento da Educação (FNDEB). Além disso, o “Mais Educação” é composto de ações desenvolvidas pelos ministérios da Educação, Desenvolvimento Social e Combate a Fome, Cultura e Esporte.

Como critério de seleção das escolas estaduais e municipais que inicialmente seriam atendidas pelo programa, a SECAD indicou escolas com baixo IDEB, situadas em capitais, regiões metropolitanas e grandes cidades em territórios caracterizados por situações de vulnerabilidade social e que necessitavam de uma convergência prioritária de políticas públicas e educacionais (MEC, 2009b). Assim, a SECAD selecionou as escolas e as secretarias, em conjunto com as escolas, confirmam a escolha, de acordo com alguns critérios pré-estabelecidos. Segundo o MEC (2008, 2009d), a seleção ao programa segue os seguintes critérios⁵⁷:

- a) Ter assinado o termo de compromisso “Todos pela Educação”;
- b) Escolas estaduais e municipais localizada em capitais ou cidades de regiões metropolitanas com mais de 200 mil habitantes;
- c) Escolas estaduais e municipais com mais de 99 matrículas registradas, conforme os dados do Censo Escolar de 2007;
- d) Escolas estaduais e municipais com Índice de desenvolvimento da educação básica (IDEB) inferior ao IDEB de seu município em 2007;
- e) Unidades escolares estaduais e municipais localizadas em municípios atendidos pelo Programa Nacional de Segurança Pública com Cidadania – PRONASCI, do Ministério da Justiça.

Segundo o MEC (2008; 2009d), o “Programa Mais Educação” é executado por meio do apoio à realização de atividades socioeducativas no contra-turno escolar, através da aquisição de *kits* para o desenvolvimento das atividades dos macrocampos do programa, do ressarcimento de despesas de alimentação e transporte de monitores e da contratação de

⁵⁷ A partir desses critérios, o MEC indica as escolas passíveis de atendimento. Entretanto, as escolas que não constarem nessa listagem preliminar do MEC, mas que estejam contempladas com recursos do PDDE-Escola poderiam implementar o programa (MEC, 2009d).

pequenos serviços e aquisição de materiais de consumo. Ou seja, além da entrega de *Kits*⁵⁸ às escolas, o MEC repassa recursos para o transporte e alimentação de monitores, transporte de alunos para outros locais de atividades, manutenção de quadra esportiva, merenda escolar, entre outros serviços (MEC, 2009d).

Em relação às atividades, estas se distribuem em dez macrocampos: educação, artes, cultura, esporte, lazer, direitos humanos, cultura digital, prevenção e promoção da saúde, educomunicação, educação científica e educação econômica. As escolas podem selecionar no mínimo de cinco e o máximo de dez atividades, distribuídas em no mínimo três macrocampos. Contudo, as escolas devem obrigatoriamente escolher pelo menos uma atividade do macrocampo “acompanhamento pedagógico”. Já as demais atividades selecionadas podem ser de outros macrocampos, desde que distribuídas em três macrocampos diferentes, conforme o quadro abaixo (2009d).

⁵⁸ Segundo MEC (2010), “os kits são compostos por materiais pedagógicos e de apoio indicados para o desenvolvimento de cada uma das atividades escolhidas pelas escolas que integram o “Programa Mais Educação”. Para cada uma das atividades do Programa, há um kit de material específico” (p. 13).

Quadro 6: Macrocampos e atividades previstas no Mais Educação

MACROCAMPO	ATIVIDADES
ACOMPANHAMENTO PEDAGÓGICO	Matemática; Letramento; Línguas Estrangeiras; Ciências; História e Geografia; Filosofia e Sociologia.
MEIO AMBIENTE	Com-Vidas – Agenda 21 na Escola – Educação para Sustentabilidade; Horta escolar e/ou comunitária.
ESPORTE E LAZER	Atletismo; Ginástica rítmica; Corrida de orientação; Ciclismo; Tênis de campo; Recreação/lazer; Voleibol; Basquete; Basquete de rua; Futebol; Futsal; Handebol; Tênis de mesa; Judô; Karatê; Taekwondo; Ioga; Natação; Xadrez tradicional; Xadrez virtual; Programa Segundo Tempo (ME).
DIREITOS HUMANOS EM EDUCAÇÃO	Direitos humanos e ambiente escolar
CULTURA E ARTES	Leitura; Banda fanfarra; Canto coral; Hip hop; Danças; Teatro; Pintura; Grafite; Desenho; Escultura; Percussão; Capoeira; Flauta doce; Cineclube; Prática circense; Mosaico.
INCLUSÃO DIGITAL	Software educacional; Informática e tecnologia da informação (PROINFO); Ambiente de Redes Sociais.
PROMOÇÃO DA SAÚDE	Alimentação saudável/alimentação escolar saudável; Saúde bucal; práticas corporais e educação do movimento; Educação para a saúde sexual; Saúde reprodutiva e prevenção das DST/Aids; • Prevenção ao uso de álcool, tabaco e outras drogas; Saúde ambiental; Promoção da cultura de paz e prevenção em saúde a partir do estudo dos principais problemas de saúde da região (dengue, febre amarela, malária, hanseníase, doença falciforme, e outras).
EDUCOMUNICAÇÃO	Jornal escolar; Rádio escolar; Histórias em quadrinhos; Fotografia; Vídeo.
INICIAÇÃO À INVESTIGAÇÃO DAS CIÊNCIAS DA NATUREZA	Laboratório, feiras de ciências e projetos científicos
EDUCAÇÃO ECONÔMICA E CIDADANIA	Educação econômica e empreendedorismo; Controle social e cidadania.

Fonte: Elaboração própria a partir de MEC (2009b)

Como alunos prioritários para o atendimento do programa o MEC recomendou inicialmente que as escolas selecionassem os estudantes em defasagem série/idade e os estudantes que estão em situação de risco e vulnerabilidade social e sem assistência. O atendimento prioritário previsto inicialmente eram as séries finais da 1ª fase do ensino fundamental e 2ª fase do ensino fundamental (MEC, 2009d), que são as séries estudadas no presente estudo⁵⁹.

O Programa teve início efetivo em 2008. Participaram nesse ano 1.409 escolas públicas estaduais e municipais de ensino fundamental localizadas em 54 municípios de 25 estados mais o Distrito Federal. Em 2009 ocorreu a ampliação do programa para 5.003 escolas em todos os estados e no Distrito Federal (MPOG, 2010b e MEC, 2011)⁶⁰.

Como mecanismo de transmissão do programa em relação do desempenho escolar, espera-se que as atividades socioculturais realizadas dentro e fora das escolas atuem como fator um motivacional. Isto é, dado que os alunos são de classe baixa de renda, essas atividades podem ser consideradas como oportunidades desses alunos vivenciarem experiências que sem o programa dificilmente conseguiriam. Além da redução das taxas de abandono em função da maior atratividade por parte do novo ambiente escolar, a suposição que é feita aqui é de que o programa atue como estímulo aos alunos no desenvolvimento de suas capacidades e habilidades, melhorando a auto-estima e a concentração deles nos estudos.⁶¹

Um outro fator importante que pode influenciar o melhor aproveitamento escolar, são as atividades de acompanhamento pedagógico das disciplinas, que a escola deve desenvolver no contra-turno. Isto é, é razoável supor que as notas dos alunos possam melhorar em função de uma maior carga horária de matemática e português. Além disso, um outro fator de estímulo é que essas matérias são ensinadas de maneira mais lúdica e menos formal, através de jogos e atividades fornecidas pelos *Kits*, o que pode facilitar a compreensão dos conteúdos.

Cabe ressaltar também que o “Programa Mais Educação” não foi criado a partir de um projeto piloto. Isto é, ele já foi pensado para ser executado em larga escala nas escolas públicas de todo o país. Nesse sentido, o questionamento sobre a validade externa é um fator fundamental a ser estudado, pois a tendência do governo brasileiro é universalizar a educação

⁵⁹ Contudo, conforme informação da SEDUC a adesão dos alunos ao Mais Educação no Rio Grande do Sul das escolas participantes ocorreu em massa.

⁶⁰ Em 2010, 10.042 escolas já estavam sendo atendidas pelo programa (MEC, 2011).

⁶¹ Baseado no DVD do Mais Educação divulgado pelo MEC.

integral nos próximos anos. Assim, é necessário promover o questionamento sobre o impacto dele em outras populações, como é o caso das escolas com alto IDEB e que não foram alvo inicialmente dessa política: essas escolas reagiriam de forma semelhante ao programa ou reagiriam de uma forma diferente. Entretanto, essa análise foge do escopo desta pesquisa⁶².

3.5 A amostra para o Rio Grande do Sul

Considerando as 1.409 escolas brasileiras que participaram do “Programa Mais Educação” no ano de 2008, 76 escolas eram do Rio Grande do Sul, sendo que a adesão foi maior nas escolas estaduais (49 escolas) do que nas escolas municipais (27 escolas). Contudo, no ano de 2009, a situação se inverteu, isto é, a adesão foi maior nas escolas municipais (139 escolas) do que nas escolas estaduais (63 escolas). O número total de escolas no Rio Grande do Sul que aderiram ao programa em 2009 foi de 202, um acréscimo de 165,8% em relação ao primeiro ano.

Observa-se que essa expansão também ocorreu no Brasil inteiro. Ou seja, as escolas participantes do programa em 2009 já ultrapassavam as 5 mil em todos os estados e no Distrito Federal, contra 1,4 mil escolas no ano anterior. No Rio Grande do Sul, apesar da elevada expansão de um ano para o outro, o número total de escolas que aderiram ao programa no ano de 2009 ainda era modesto: cerca de 274 escolas estavam participando do programa federal, ou seja, pouco mais de 5% do total de escolas participantes no Brasil.

⁶² Uma forma de tentar responder esse questionamento é realizar análises através de outras técnicas de avaliação de impacto descritas na literatura, como é o caso da regressão descontínua, em que uma faixa de escolas com mais alto IDEB entre as escolhidas a participar do programa poderiam ser comparadas com as de mais baixo IDEB das que possuem um IDEB mais elevado. Ou seja, em uma faixa de intervalo na variável de elegibilidade do programa.

Tabela 1: Escolas participantes do “Programa Mais Educação” no Rio Grande do Sul – 2008 e 2009⁶³

Escolas	Início em 2008	Início em 2009	Participantes em 2008 e 2009
Nº de escolas estaduais participantes	49	63	108 ⁶⁴
Nº de escolas municipais participantes	27	139	166
Total de escolas participantes	76	202	274⁶⁵

Fonte: elaboração própria a partir dos dados fornecidos pelo MEC e SEDUC-RS

Entre as 9.888 escolas da rede pública municipal e estadual ativas do Rio Grande do Sul do censo escolar de 2007, 3.101 escolas contavam com informações na Provas Brasil 2007. Em relação a 2009, das 10.058 escolas das redes estadual e municipal do Rio Grande do Sul com informações no censo escolar, 3.420 contavam com informações na Provas Brasil. Após a exclusão de informações *missing*⁶⁶, o total de escolas com informações na Prova Brasil 2007-2009 e do Censo Escolar 2007-2009 foram de 1.909 e 1.532, respectivamente para a 4ª série e 8ª série, conforme o quadro abaixo. Cabe destacar apenas que inicialmente o grupo de controle das amostras foi constituído a partir de todas as demais escolas públicas do RS que não participaram do “Programa Mais Educação”.

⁶³ Informações obtidas no MEC e Secretaria de Educação do RS através do SIMEC (Sistema Integrado de Monitoramento Execução e Controle do MEC). Segundo a SEDUC/RS, o número de alunos participantes no estado do RS chega a 149.663 em 2010 < http://www.educacao.rs.gov.br/pse/html/mais_educacao.jsp?ACAO=acao1>. Acesso em 04/01/2012.

⁶⁴ Quatro escolas estaduais deixaram de participar do programa em 2009.

⁶⁵ Quatro escolas estaduais deixaram de participar do programa em 2009.

⁶⁶ Para formar um painel balanceado, foram excluídas observações (escolas) que possuíam dados em branco em qualquer uma das variáveis selecionadas. Entretanto, antes da exclusão de observações foram realizados testes de médias para comparar o dados faltantes entre os dois grupos (tratamento e controle). Em nenhuma das variáveis há diferenças significativas nos dois grupos. Assim, procedeu-se com a exclusão de observações que possuíam dados faltantes, sem a necessidade da utilização de técnicas de imputação ou reponderação das variáveis.

Tabela 2: Construção da amostra para o ano de 2007 e 2009
– Escolas do Rio Grande do Sul

Amostra para o ano de 2007	Nº escolas
Encontradas no Censo Escolar 2007	9.888
Encontradas no na Prova Brasil 2007	3.101
Amostra para o ano de 2009	
Encontradas no Censo Escolar 2009	10.058
Encontradas no na Prova Brasil 2009	3.420
Filtros	
Escolas com turmas de 4ª série encontradas na Prova Brasil 2007/2009 e Censo Escolar 2007/2009 após filtros	1.909
Escolas com turmas de 8ª série encontradas na Prova Brasil 2007/2009 e Censo Escolar 2007/2009 após filtros	1.532

Fonte: Elaboração própria

Com as informações sobre a participação das escolas do Rio Grande do Sul obtidas junto ao MEC e a Secretaria Estadual de Educação – SEDUC (tabela 1) e com os filtros realizados junto a Prova Brasil e Censo Escolar de 2007 e 2009 (tabela 2), construiu-se as amostras para a 4ª série e 8ª série como segue abaixo.

Tabela 3: Amostras para o Rio Grande do Sul para a 4ª série e 8ª série

Quarta série					
	Grupo Tratamento		Grupo Controle		Total
	Descrição	número escolas	Descrição	número escolas	
Amostra 1	escolas participantes da edição 2008 ou 2009	189	Demais escolas públicas do RS que nunca participaram do programa	1.720	1.909
Amostra 2	escolas participantes somente da edição 2009	140	Demais escolas públicas do RS que nunca participaram do programa	1.720	1.860
Oitava série					
	Grupo Tratamento		Grupo Controle		Total
	Descrição	número escolas	Descrição	número escolas	
Amostra 1	escolas participantes da edição 2008 ou 2009	137	Demais escolas públicas do RS que nunca participaram do programa	1.395	1.532
Amostra 2	escolas participantes somente da edição 2009	93	Demais escolas públicas do RS que nunca participaram do programa	1.395	1.488

Fonte: Elaboração própria

Na 4ª série, a amostra contém as escolas participantes do programa independente do ano de início. Foram identificadas 189 escolas como integrantes do programa, após a realização dos filtros. O grupo de controle contém 1.720, o que totaliza 1.909 escolas na amostra 1. Já as escolas da amostra 2 para a 4ª série (escolas participantes somente da edição 2009), 140 escolas foram identificadas como participantes do programa e 1.710 escolas como pertencentes ao grupo de controle. Essa última amostra contém 1.860 escolas.

Em relação a 8ª série, na amostra 1, que contém as escolas participantes do programa independente no ano de início, 137 escolas pertencem ao grupo de tratamento e 1.395 ao grupo de controle, no total de 1.532 escolas. Para a amostra 2 na 8ª série, cerca de 93 escolas fazem parte do grupo de tratamento e 1.395 do grupo de controle. O total de escolas pertencentes a amostra 2 é de 1.488.

Após a apresentação da metodologia e das bases de dados e amostra utilizadas, apresentam-se no próximo capítulo os resultados da pesquisa.

4. RESULTADOS

A seguir são mostradas as estatísticas descritivas para as amostras para a 4ª série e 8ª série. Após a apresentação sintética dos dados das amostras, os resultados da análise de impacto do “Programa Mais Educação” para a proficiência de matemática e língua portuguesa (4ª série e 8ª série) são mostrados. Por fim, ao final do capítulo realiza-se uma discussão sobre os resultados da pesquisa.

4.1 Estatísticas descritivas e testes de igualdade de média e de distribuição

Nessa seção apresenta-se a análise das estatísticas descritivas das amostras para a 4ª e 8ª séries. Para comparar as diferenças existentes entre o grupo de tratamento (escolas participantes do programa) e o grupo de controle (escolas não participantes) antes do pareamento, foram realizados dois testes em cada variável: o teste de igualdade de médias (t teste) e o teste igualdade entre duas distribuições (teste *Kolmogorov-Smirnov*).

As estatísticas descritivas e resultados dos testes de médias e de distribuição para a amostra 1 para a 4ª série são apresentadas na tabela 4. O objetivo da apresentação dessas informações foi mostrar a existência de diferenças nas médias e nas distribuições das variáveis entre o grupo de controle (1.720 escolas) e grupo de tratamento (189 escolas). Como há diferenças e essas se mostraram estatisticamente significativas (verificadas nos testes) na maioria das co-variáveis, houve a necessidade de se proceder como o pareamento⁶⁷. Isso porque, ao selecionar um grupo de controle que corresponde a todas as demais escolas públicas do Rio Grande do Sul e que não participaram do programa, outros fatores que afetam o desempenho dos alunos e conseqüentemente o desempenho médio das escolas não estão sendo isolados na análise, o que gera um viés nos resultados.

⁶⁷ Nesse sentido, o pareamento tem o propósito de selecionar no grupo de controle escolas mais parecidas possíveis com o grupo de tratamento.

Tabela 4: Estatísticas descritivas e testes de médias da amostra 1 para a 4ª série

	Escolas Controles = 1720		Escolas tratadas = 189		Teste de igualdade de médias	Teste Kolmogorov- Smirnov
	Média	Desvio-padrão	Média	Desvio-padrão		
Prova Brasil						
Nota_4a_MAT_2007	199.61	16.49	188.52	13.18	***	***
Nota_4a_PORT_2007	181.73	15.30	170.47	12.35	***	***
Nota_4a_MAT_2009	210.42	19.07	195.92	13.09	***	***
Nota_4a_PORT_2009	189.14	16.47	177.95	11.81	***	***
SEXO	0.49	0.11	0.50	0.09	n	n
RACA	0.59	0.16	0.52	0.12	***	***
MAQ_LAV	0.86	0.13	0.85	0.08	n	***
CAR	0.70	0.27	0.58	0.20	***	***
INTER_CASA	0.24	0.16	0.25	0.13	n	**
DEVER	0.98	0.03	0.97	0.03	**	***
LER	0.97	0.04	0.95	0.04	***	***
NÃO_FALTAR	0.98	0.03	0.97	0.03	***	***
MAE_MEDIO	0.12	0.08	0.11	0.07	n	***
MAE_SUPER	0.11	0.09	0.10	0.06	*	***
MAE_NAORES	0.33	0.12	0.34	0.10	n	***
TRAB	0.12	0.09	0.11	0.06	n	n
REUN	0.92	0.06	0.89	0.06	***	***
PRE_ESC	0.74	0.16	0.54	0.19	***	***
DIR_29	0.03	0.16	0.01	0.10	n	n
DIR_30_a_39	0.21	0.40	0.24	0.43	n	n
DIR_40	0.77	0.42	0.75	0.44	n	n
DIR_POS	0.69	0.46	0.68	0.47	n	n
DIR_11_a_15	0.09	0.29	0.08	0.28	n	n
DIR15	0.05	0.22	0.05	0.22	n	n
PRO_SEL	0.00	0.05	0.00	0.00	n	n
INT_ATI	0.10	0.29	0.19	0.39	***	n
ROT_PROF	0.07	0.25	0.10	0.29	n	n
REF_ESC	0.75	0.43	0.72	0.45	n	n
DIR_SEL_ELE	0.08	0.27	0.10	0.29	n	n
PROG_FED	0.73	0.44	0.84	0.37	***	**
PROG_EST	0.41	0.49	0.32	0.47	**	n
PROG_MUN	0.36	0.48	0.47	0.50	***	**
Censo Escolar 2007						
MÉDIA_AULA	244.17	22.56	247.13	25.37	*	***
MED_MAT_SERIE	24	5	28	5	***	***
TOTAL_MAT	50	26	85	41	***	***
BIBLIO	0.96	0.20	0.94	0.24	n	n
INTER	0.66	0.47	0.59	0.49	*	n
LAB_INFO	0.55	0.50	0.65	0.48	**	*
PROF_SUPER	0.83	0.18	0.79	0.20	***	***
PROF_POS	0.28	0.22	0.25	0.24	n	***
PROF_FUND	78	48	113	89	***	***
Informações do IBGE 2007						
POP	177.733	336.197	692.894	581.055	***	***
PIB_PER	16,223.97	11,558.05	19,431.03	8,475.68	***	***
RM_OU_POLO	0.57	0.50	1.00	0.00	***	***
Informações do MEC e INEP 2007						
IDEB_mun	4.54	0.46	4.19	0.24	***	***
IDEB_esc	4.58	0.74	3.87	0.57	***	***
Dif_IDEB	0.03	0.61	-0.33	0.49	***	***
PRONASCI	0.29	0.45	1.00	0.00	***	***
TAXA_APRO	89.71	9.57	83.65	9.76	***	***

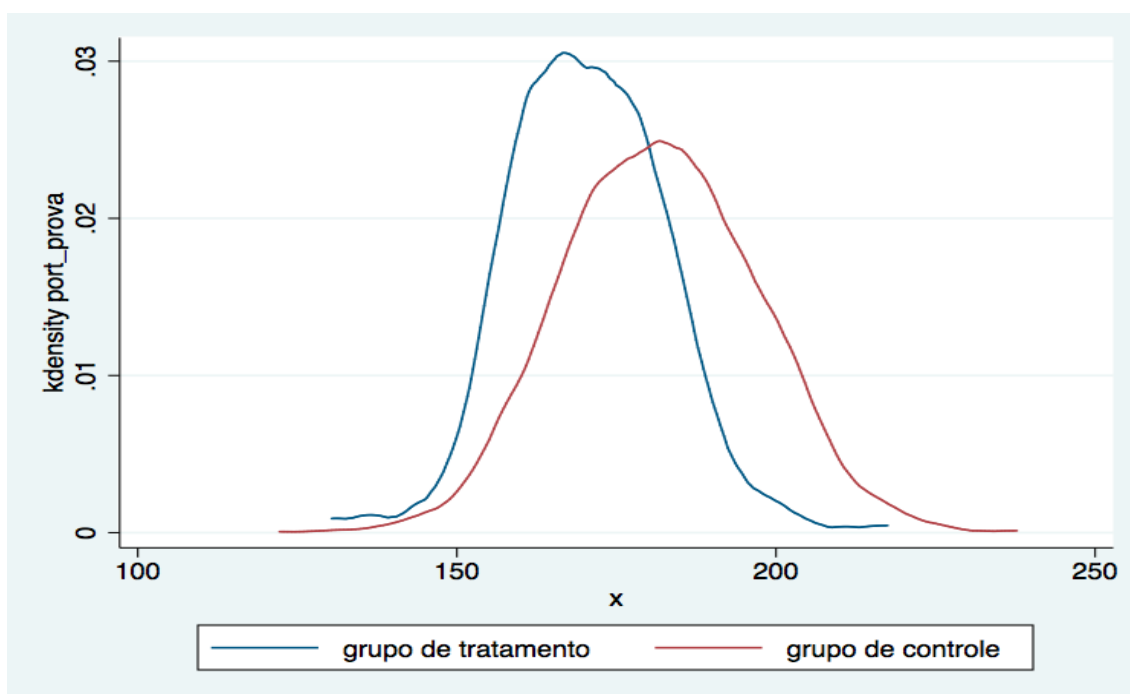
MUNIC	0.46	0.50	0.61	0.49	***	***
-------	------	------	------	------	-----	-----

Nota: (***) Nível de significância de 1%; (**) Nível de significância de 5%; (*) Nível de significância de 10%; (n) não significativo.

Fonte: elaboração própria a partir dos dados da Prova Brasil 2007 e 2009, Censo Escolar 2007 e Informações do IBGE e INEP/MEC para o ano de 2007.

A tabela 4 mostra uma diferença acentuada na maioria das variáveis entre os grupos de tratamento e de controle, com grande vantagem para esse último grupo. Em relação às variáveis dependentes (proficiência em português e matemática), percebe-se que o grupo controle parte de uma condição inicial bem mais favorável.

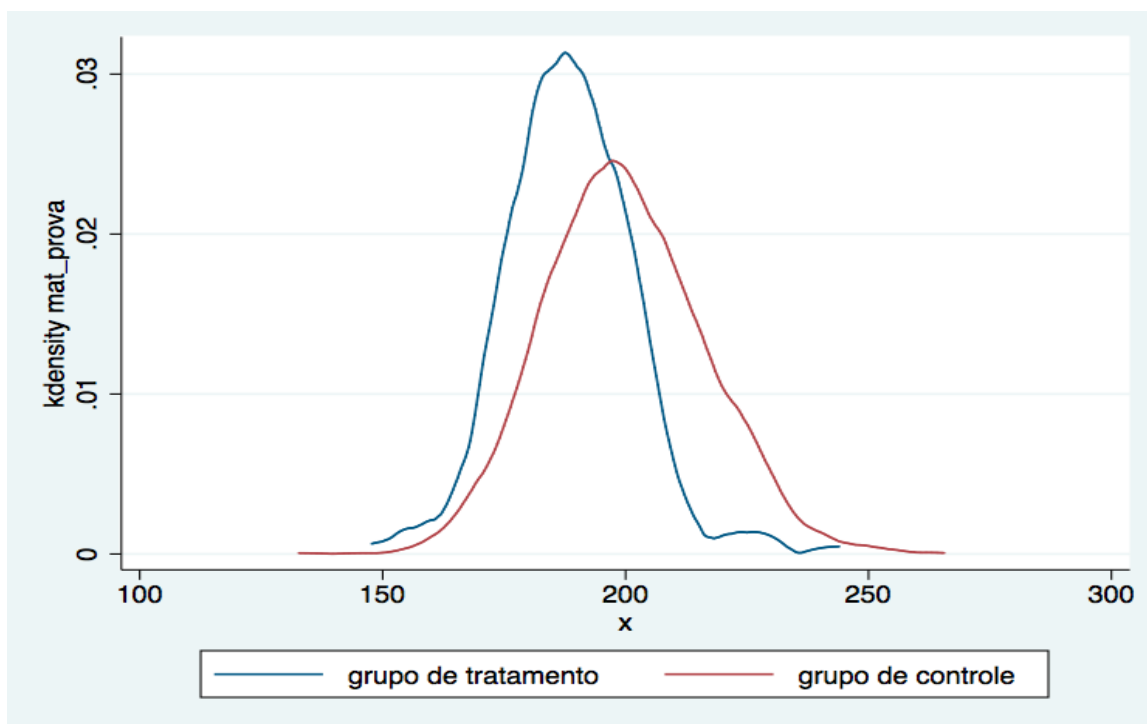
Gráfico 1: Distribuição das notas de português na 4ª série da Prova Brasil por grupo, 2007 – Amostra 1



Fonte: Elaboração própria

Pelo gráfico 1, é possível verificar a diferença nas distribuições das notas de português no ano de 2007 entre o grupo de tratamento e grupo de controle. Ou seja, o grupo de controle sem qualquer procedimento de pareamento, apresenta melhores desempenhos em relação ao grupo de tratamento. A nota média das escolas em português é de 170,47 para o grupo de tratamento e 181,73 para o grupo de controle. Cabe ressaltar que, pelos testes t e *Kolmogorov-Smirnov*, essas diferenças são estatisticamente significativas (a 1% de significância).

Gráfico 2: Distribuição das notas de matemática na 4ª série da Prova Brasil por grupo, 2007 – Amostra 1



Fonte: Elaboração própria

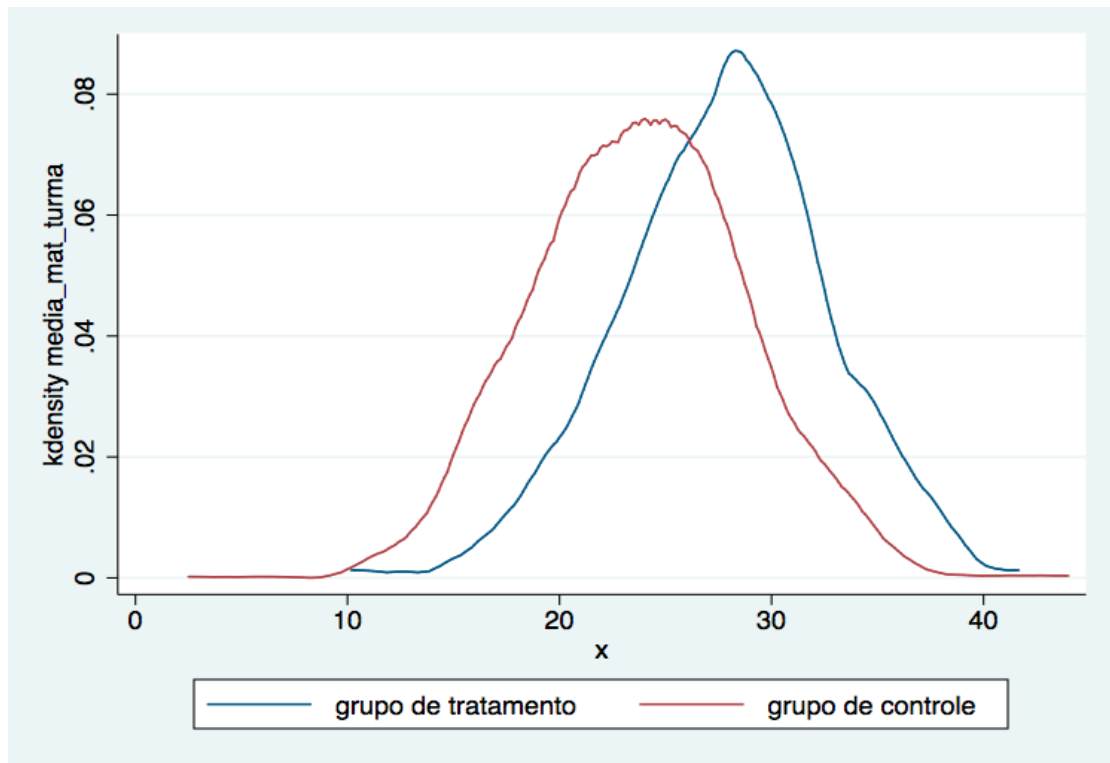
Em relação à proficiência em matemática para o ano de 2007, há também diferenças para os dois grupos. O grupo de controle também apresenta resultados relativamente melhores. A nota média das escolas em matemática é de 188,52 para o grupo de tratamento e 199,61 para o grupo de controle. Os testes média e de distribuição também mostram que essa diferença é estatisticamente significativa ao nível de 1%.

A diferença entre as médias das variáveis de controle também é significativa entre os dois grupos para grande maioria dessas variáveis. Para as variáveis que retratam as características socioeconômicas médias dos alunos das escolas, o grupo de controle apresenta superioridade no percentual de alunos da 4ª série que possuem máquina de lavar, são alunos brancos, são incentivados pelos pais a ler, a fazer o dever e a não faltar, possuem pais que comparecem as reuniões da escola, mães com ensino médio e superior. Além disso, são significativas, com vantagem para o grupo de controle, as diferenças entre os dois grupos em relação ao número médio de automóveis para os alunos de 4ª série das escolas e no percentual de alunos que frequentaram a pré-escola. As diferenças mais significativas (em termos de pontos percentuais e estatísticas) referem-se ao número médio de carros dos alunos da 4ª

séries das escolas (diferença de 0,12 carros), no percentual de pais que freqüentam reuniões da escola (diferença de 3 pontos percentuais - p.p), no percentual de alunos da raça branca (7 p.p) e no percentual de alunos que frequentaram a pré-escola (20 p.p.). Estas diferenças alcançam nos dois testes um nível de significância de 1%.

Em relação às características das escolas, não há diferenças estatisticamente significativas para as características dos diretores (idade, processo seletivo) e nem para a alta rotatividade dos professores. Também não diferem as médias e as distribuições das variáveis se a escola possui atividade de reforço escolar ou se possui biblioteca. Contudo aspectos relacionados aos professores, turma e matrículas e alguns aspectos relacionado à infraestrutura se mostram diferentes entre os dois grupos. O grupo de controle apresenta um maior percentual de professores da 4ª série com ensino superior (diferença de 4 p.p) e pós-graduação (diferença de 3 p.p.). Destaca-se também que essas escolas apresentam o número de matrículas na série (35 matrículas a menos) e o número de professores do ensino fundamental (35 professores a menos) inferiores que o grupo de tratamento. O gráfico 3 mostra a distribuição da média de matrículas por turma na 4ª série para o tratamento e para o grupo de controle. Constata-se que o grupo de controle possui o número médio de alunos por turma inferior ao grupo de tratamento. Em termos de valores absolutos essa diferença corresponde a 4 alunos a menos em média nas turmas de 4ª série.

Gráfico 3: Distribuição da média de matrículas por turma na 4ª série por grupo, 2007

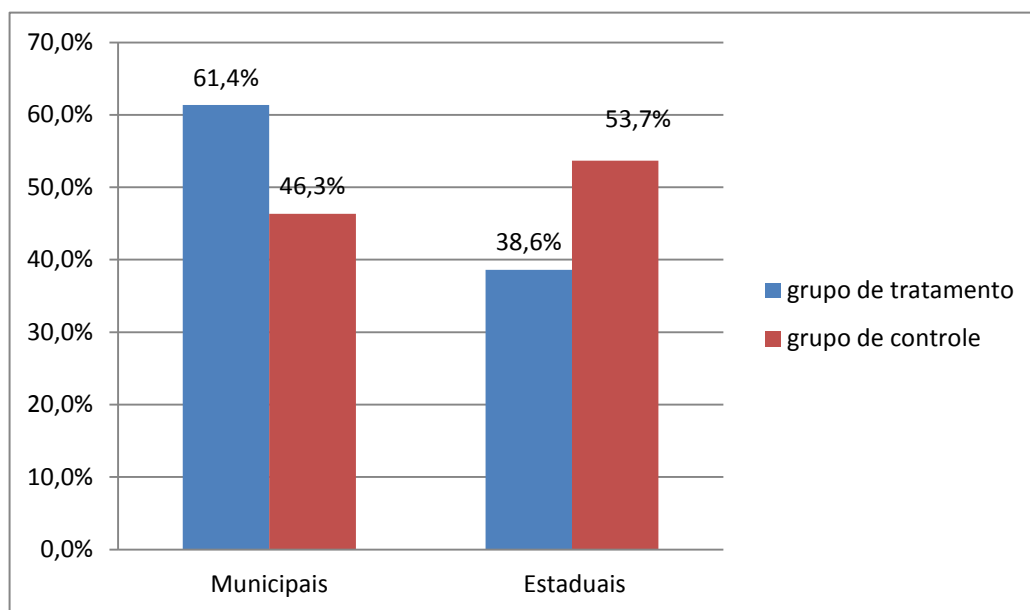


Fonte: Elaboração própria

Ressalta-se que o grupo tratamento só leva vantagem no que diz respeito ao número de escolas com laboratório de informática (65,00% conta 55,00%) e no percentual de escolas com ajuda financeira de outros programas dos municípios e do governo federal (diferença de 11 p.p). Além de apresentar uma média de duração de aula (em minutos) maior do que o grupo de controle, isto é, 247,13 minutos de aula conta 244,17 minutos do grupo de controle.

Quanto à estrutura municipal em que as escolas dos dois grupos estão inseridas, o grupo de tratamento apresenta escolas que estão localizadas em municípios mais populosos e que possuem maior PIB per capita. Além disso, o grupo de tratamento conta com maior percentual de escolas situadas na região metropolitana (100,00% contra 57,00%). Há também um maior percentual de escolas desse grupo pertencente à esfera administrativa municipal do que o grupo de controle (61,00% contra 46,00%) e a municípios pertencentes ao PRONASCI (100,00% conta 29,00% do grupo de controle). O gráfico 4 apresenta o percentual de escolas do grupo de controle e tratamento por esfera administrativa.

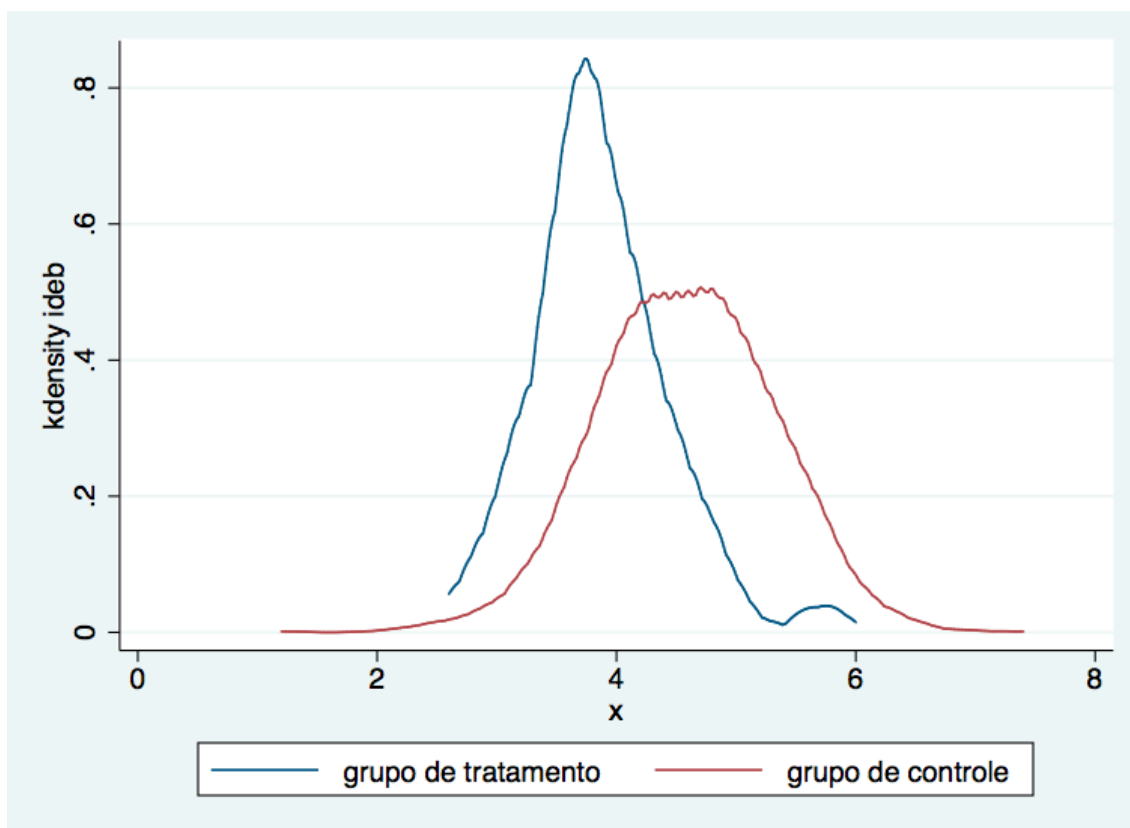
Gráfico 4: Percentual de escolas por esfera administrativa e por grupo para a 4ª série, 2007



Fonte: Elaboração própria

Outras diferenças estatisticamente significativas estão relacionadas a média do IDEB das escolas e taxa de aprovação em 2007 para a 4ª série. O grupo de controle possui maior percentual de taxa de aprovação de alunos da 4ª série (89,71%) do que o grupo de tratamento (83,65%), uma diferença de 6,06 p.p. Em relação ao IDEB das escolas, o grupo de controle tem escolas com um IDEB médio de 4,58, enquanto que as escolas do grupo de tratamento possuem um IDEB médio de 3,87 (gráfico 5). Além disso, as escolas do grupo de controle estão inseridas em municípios com maior IDEB do que as escolas do grupo de tratamento e possuem uma menor diferença média do IDEB das escolas em relação ao IDEB municipal (enquanto que a diferença é positiva para o grupo de controle, isto é 0,30 pontos, para o grupo de tratamento esta diferença é negativa, ou seja, -0,33 pontos em média).

Gráfico 5: Distribuição do IDEB das escolas em 2007 para a 4ª série por grupo, 2007



Fonte: Elaboração própria

Ressalta-se, portanto, as desvantagens iniciais que existem nas escolas participantes do programa frente às escolas não participantes do programa. Sugere-se que esses resultados se devam a própria seleção do programa, que priorizou a adesão de escolas em pior situação⁶⁸. Conforme visto na literatura, as características socioeconômicas dos alunos e as características das escolas afetam o desempenho médio dos estudantes e se não forem isolados, os resultados podem conter um forte viés. Nesse sentido, torna-se fundamental a realização de um bom pareamento. Em termos práticos, isso significa que após o pareamento, os testes de igualdades de médias devem ser realizado novamente e essas diferenças não poderão ser significativas. Em relação a amostra 2, isto é, com somente as escolas que

⁶⁸ Cabe lembrar também que nos outros estudos de avaliação de impacto de projetos e programas realizados com variáveis semelhantes (BIONDI; VASCONCELLOS; MENEZES-FILHO, 2009; VASCONCELLOS; BIONDI; MENEZES-FILHO, 2009), o grupo de tratamento apresentava maiores vantagens que o grupo de controle. Este fenômeno pode ser explicado pelo fato da adesão aos projetos ser voluntária. Além disso, os grupos se mostram mais homogêneos do que o presente estudo, o que facilita o trabalho de pareamento.

iniciaram o programa em 2009, os resultados não diferem muito do que o encontrado para a amostra 1, conforme a tabela 12 em anexo.

Para a 8ª série, as estatísticas descritivas e resultados dos testes para a amostra 1 são apresentadas na tabela a seguir (tabela 5). Assim como na análise para a 4ª série, o objetivo é mostrar as diferenças nas médias das variáveis existentes entre o grupo de controle (1.395 escolas) e grupo de tratamento (137 escolas) antes do pareamento. Como a amostra para a 8ª série apresentou diferenças entre o grupo de tratamento e controle e estão são estatisticamente significativas (verificadas nos testes de médias e distribuição) existe a necessidade de se proceder com algum tipo de pareamento.

Tabela 5: Estatísticas descritivas e testes de médias da amostra 1 para a 8ª série

	Escolas Controles = 1395		Escolas tratadas = 137		Teste de igualdade de médias	Teste Kolmogorov- Smirnov
	média	desvio- padrão	média	desvio- padrão		
Prova Brasil						
Nota_4a_MAT_2007	252.24	17.71	241.04	15.97	***	***
Nota_4a_PORT_2007	239.57	15.73	232.27	15.49	***	***
Nota_4a_MAT_2009	258.56	18.47	245.64	15.48	***	***
Nota_4a_PORT_2009	250.58	16.26	241.18	14.66	***	***
SEXO	0.46	0.11	0.47	0.10	n	n
RACA	0.67	0.15	0.56	0.13	***	***
MAQ_LAV	0.84	0.14	0.83	0.10	n	***
CAR	0.75	0.26	0.58	0.19	***	***
INTER_CASA	0.30	0.19	0.30	0.15	n	n
DEVER	0.96	0.04	0.95	0.04	**	***
LER	0.90	0.07	0.89	0.06	**	**
NÃO_FALTAR	0.99	0.02	0.99	0.02	n	n
MAE_MEDIO	0.21	0.12	0.20	0.10	n	n
MAE_SUPER	0.08	0.08	0.05	0.05	***	***
MAE_NAORES	0.14	0.08	0.14	0.07	n	n
TRAB	0.21	0.12	0.16	0.08	***	***
REUN	0.89	0.09	0.82	0.10	***	***
PRE_ESC	0.80	0.16	0.65	0.18	***	***
DIR_29	0.02	0.15	0.01	0.09	n	n
DIR_30_a_39	0.19	0.39	0.23	0.42	n	n
DIR_40	0.79	0.41	0.76	0.43	n	n
DIR_POS	0.71	0.45	0.69	0.46	n	n
DIR_11_a_15	0.10	0.29	0.04	0.21	**	n
DIR15	0.04	0.20	0.07	0.25	n	n
PRO_SEL	0.01	0.07	0.00	0.00	n	n
INT_ATI	0.11	0.31	0.20	0.40	***	n
ROT_PROF	0.09	0.28	0.11	0.31	n	n
REF_ESC	0.73	0.44	0.67	0.47	n	n
DIR_SEL_ELE	0.07	0.26	0.06	0.24	n	n
PROG_FED	0.71	0.45	0.82	0.38	***	*
PROG_EST	0.48	0.50	0.38	0.49	**	n
PROG_MUN	0.29	0.45	0.40	0.49	***	*
Censo Escolar 2007						
MÉDIA_AULA	252.45	18.63	254.52	13.55	n	***
MED_MAT_SERIE	25	6	27	5	***	***
TOTAL_MAT	51	30	69	39	***	***
BIBLIO	0.98	0.13	0.96	0.19	n	n
INTER	0.74	0.44	0.68	0.47	n	n
LAB_INFO	0.63	0.48	0.69	0.47	n	n
PROF_SUPER	0.87	0.15	0.87	0.11	n	**
PROF_POS	0.29	0.21	0.28	0.25	n	***
PROF_FUND	91	44	140	88	***	***
Informações do IBGE 2007						
POP	169.048	333,763	753.291	589,445	***	***
PIB_PER	16,013.92	10,359.39	19,933.42	8,590.23	***	***
RM_OU_POLO	0.54	0.50	1.00	0.00	***	***
Informações do MEC e INEP 2007						
IDEB_mun	3.84	0.48	3.38	0.17	***	***
IDEB_esc	3.81	0.72	3.17	0.60	***	***
Dif_IDEB	-0.03	0.56	-0.21	0.56	***	***
PRONASCI	0.27	0.45	1.00	0.00	***	***
TAXA_APRO	84.29	12.20	77.26	14.84	***	***

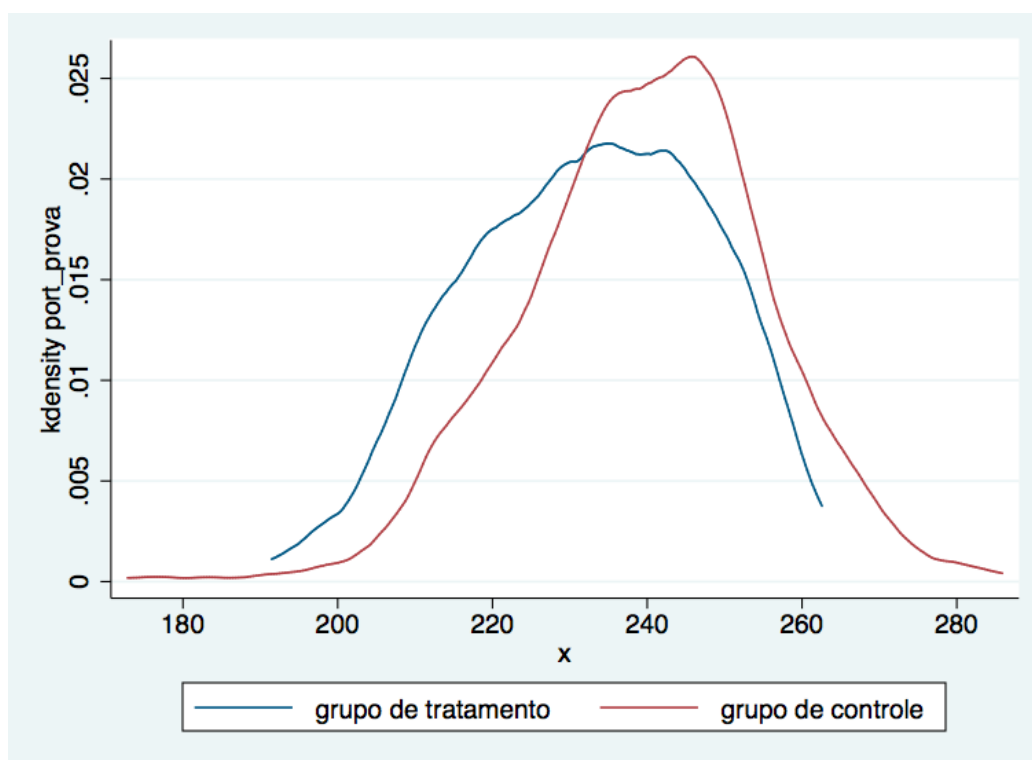
MUNIC	0.37	0.48	0.55	0.50	***	***
-------	------	------	------	------	-----	-----

Nota: (***) Nível de significância de 1%; (**) Nível de significância de 5%; (*) Nível de significância de 10% ; (n) não significativo.

Fonte: elaboração própria a partir dos dados da Prova Brasil 2007 e 2009, Censo Escolar 2007 e Informações do IBGE e INEP/MEC para o ano de 2007.

Observa-se para a amostra 1 da 8ª série, que existe uma diferença expressiva na maioria das variáveis entre os grupos de tratamento e de controle, com grande vantagem para esse último grupo. Em relação às variáveis dependentes, constata-se, assim como na 4ª série, que o grupo controle parte de uma condição inicial bem mais favorável.

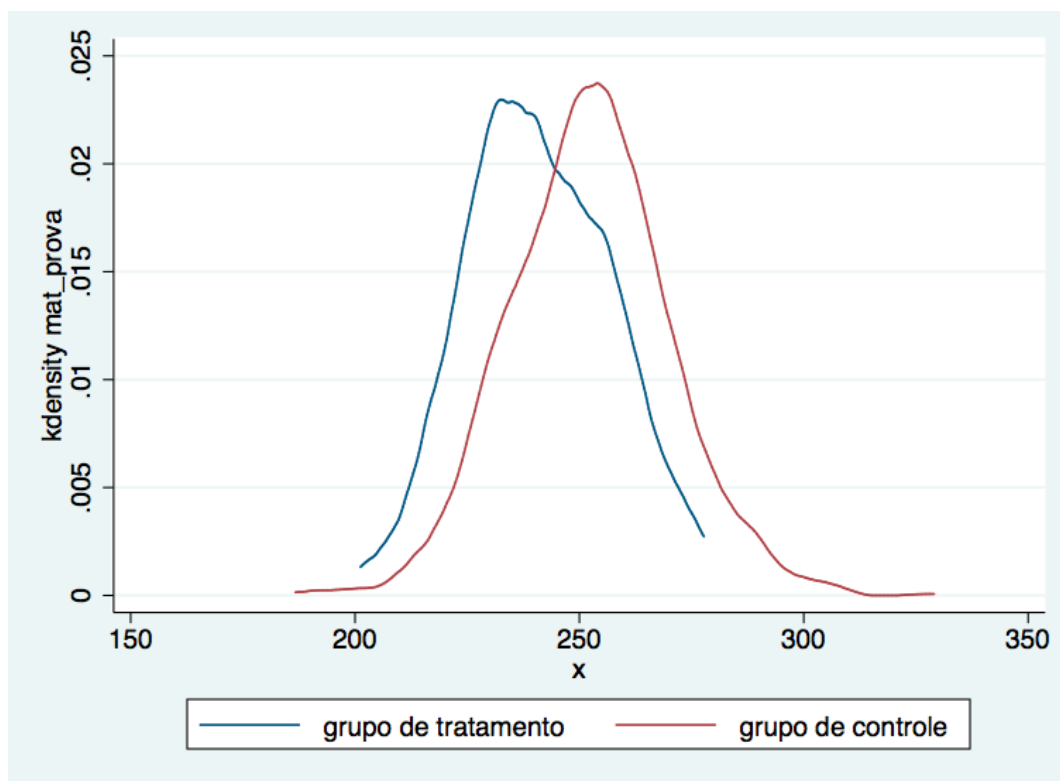
Gráfico 6: Distribuição das notas de português na 8ª série da Prova Brasil por grupo, 2007 – Amostra 1



Fonte: Elaboração própria

O Gráfico 6 mostra a distribuição das notas em português no ano de 2007 para o grupo de controle e o grupo de tratamento. O grupo de controle sem qualquer procedimento de pareamento, apresenta melhores desempenhos em relação ao grupo de tratamento. A nota média das escolas em português para a 8ª série é de 232,27 para o grupo de tratamento e 239,57 para o grupo de controle. Pelos testes *t* e *Kolmogorov-Smirnov*, essas diferenças são estatisticamente significativas (a 1% de significância).

Gráfico 7: Distribuição das notas de matemática na 8ª série da Prova Brasil por grupo, 2007 – Amostra 1



Fonte: Elaboração própria

Em relação às notas de matemática em 2007, há também uma diferença entre o grupo de tratamento e o grupo de controle, com vantagem para esse último, conforme pode ser verificado no gráfico 7. As notas médias das escolas do grupo de controle para matemática são de 252,24 e do grupo de tratamento são de 241,04. Cabe ressaltar que, pelos testes de médias e distribuição, essas diferenças são estatisticamente significativas.

Para a 8ª série, as variáveis de controle também se mostram diferentes entre os dois grupos. Em relação às variáveis que refletem às características socioeconômicas médias dos alunos das escolas, o grupo de controle apresenta superioridade no percentual de alunos de 8ª série que são incentivados pelos pais a ler e fazer o dever e ao percentuais de pais de alunos que comparecem a reuniões de escola. Além disso, as diferenças entre o grupo de controle e tratamento, com vantagem para o primeiro, são acentuadas para o percentual de alunos da 8ª série de raça branca (diferença de 11 p.p), mãe com ensino superior (diferença de 3 p.p) alunos que freqüentaram a pré-escola (diferença de 15 p.p) e no número médio de carros (diferença de 0,17 carros). A exceção fica por conta do percentual de alunos que trabalham,

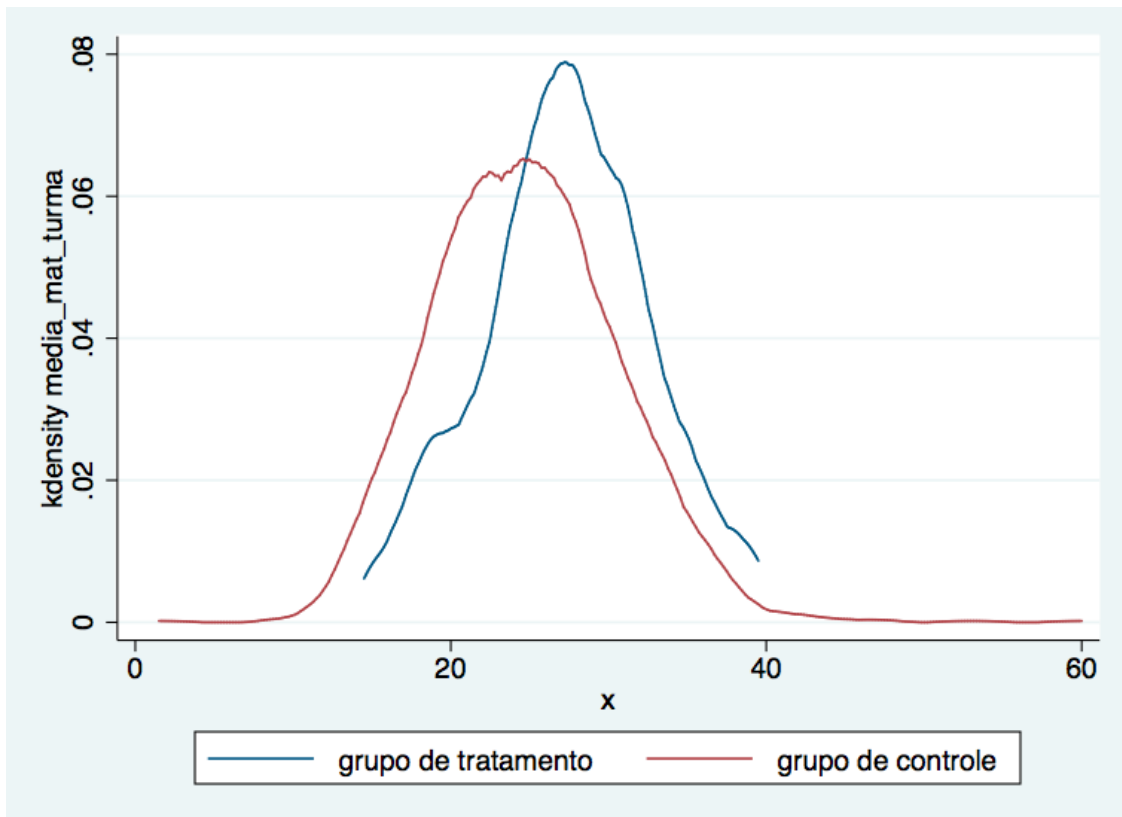
em que esse percentual representa 21,00% para o grupo de controle e 16,00% para o grupo de tratamento. Ou seja, o grupo de tratamento para a 8ª série apresenta vantagem nesse aspecto.

Em relação às características das escolas, não há diferenças estatisticamente significativas para a maioria das características dos diretores (idade, processo seletivo), para o percentual de escolas com alta rotatividade dos professores e em relação à infraestrutura (biblioteca, laboratório de informática, internet na escola). Adicionalmente, as diferenças na média de duração de aula (em minutos) e professor com ensino superior não se mostraram significativas para a 8ª série⁶⁹, diferentemente da amostra para a 4ª série.

Na dimensão escola, as principais diferenças estão relacionadas ao percentual de escolas com interrupção por problema grave (diferença de 9 p.p a mais para o grupo de tratamento), total de matrícula na série (média de 18 matrículas a mais para o grupo de tratamento) e professores no ensino fundamental (média de 41 professores a mais no grupo de tratamento). O Gráfico 8 mostra a distribuição da média de matrículas na série. Em termos de pontos percentuais essa diferença corresponde a 2 matrículas a mais para o grupo de tratamento. Ou seja, a média de alunos por turma na 8ª série para o grupo de controle é de 25 alunos, enquanto que essa média para o grupo de tratamento corresponde a 27 alunos por turma.

⁶⁹ Somente para o teste de igualdade de médias

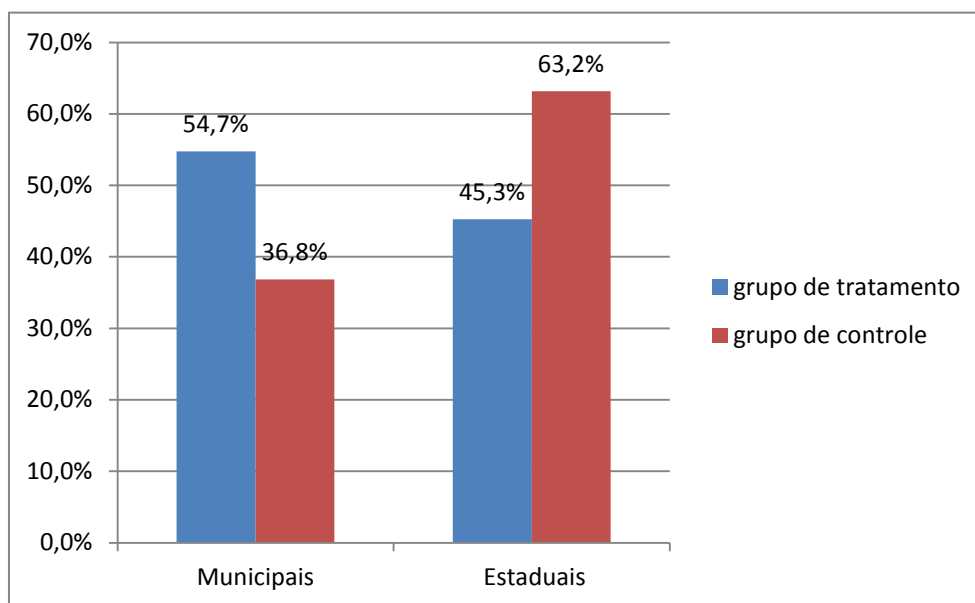
Gráfico 8: Distribuição da média de matrículas por turma na 8ª série por grupo, 2007



Fonte: Elaboração própria

Assim como para a 4ª série, em relação à estrutura municipal em que as escolas dos dois grupos estão localizadas, o grupo de tratamento apresenta escolas que estão situadas em municípios mais populosos e com maior PIB per capita e com maior percentual das escolas situadas na região metropolitana. Além disso, conforme o gráfico 9, há também um maior percentual de escolas desse grupo que pertencente à esfera administrativa municipal do que o grupo de tratamento (55,00% contra 37,00%) e a municípios pertencentes ao PRONASCI (100,00% contra 27,00% do grupo de controle).

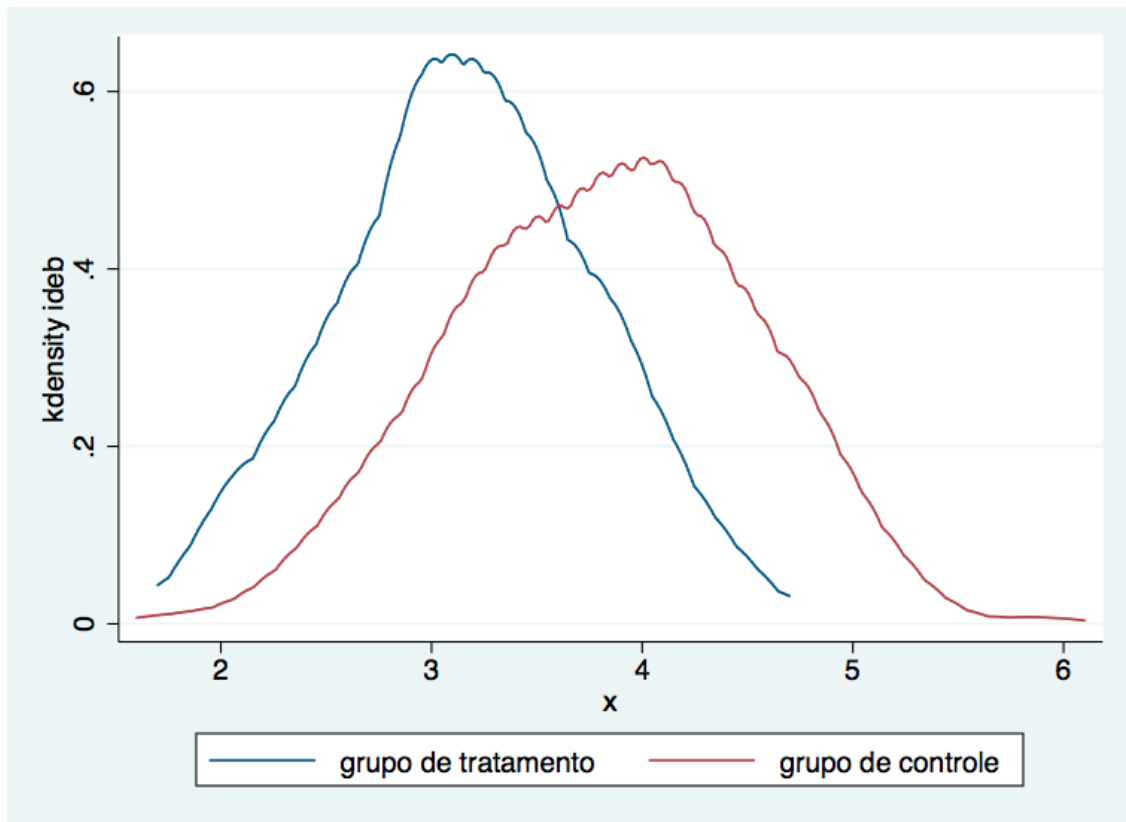
Gráfico 9: Percentual de escolas por esfera administrativa e por grupo para a 8ª série, 2007



Fonte: Elaboração própria

Para a 8ª série, o IDEB da escola e taxa de aprovação em 2007 também apresentam significativas diferenças nos dois grupos. O grupo de controle mostra um maior percentual de taxa de aprovação de alunos da 8ª série (84,29 %) do que o grupo de tratamento (77,26%). Além disso, as escolas do grupo de controle estavam inseridas em municípios com maior IDEB do que as escolas do grupo de tratamento. Pelo gráfico 10, observa-se também que o IDEB médio das escolas do grupo de controle é superior que o grupo de tratamento.

Gráfico 10: Distribuição do IDEB médio das escolas em 2007 para a 8ª série por grupo, 2007



Fonte: Elaboração própria

Ressalta-se, portanto, que as desvantagens iniciais que existem nas escolas participantes do programa para a 4ª são também evidenciadas para a 8ª série, em função da própria seleção do programa, que priorizou a adesão de escolas em pior situação. Em relação à amostra 2 (tabela 13 em Anexo) para a 8ª série, isto é, considerando no grupo de tratamento somente as escolas que iniciaram a execução do programa em 2009, os resultados não mudam muito do que o encontrado para a amostra 1. Destaca-se apenas que variáveis relacionadas à infraestrutura passam a ser estatisticamente diferentes para os dois grupos.

4.2 Resultados para a 4ª Série

A estimação do *propensity score* para as amostras 1 e 2 da 4ª série foi realizada através de um modelo *logit*. As variáveis do modelo foram pré-selecionadas a partir da descrição da literatura sobre os fatores que afetam desempenho escolar (capítulo 2) e dos fatores que afetam a probabilidade de participação conforme os critérios estabelecidos pelo MEC para seleção de escolas que poderiam ser atendidas pelo programa.

Contudo, para a escolha do modelo *logit* mais adequado utilizou-se um teste LR, além da verificação dos critérios de informações de *Akaike* – AIC e *Schwarz* – SIC para um conjunto das variáveis (X) inseridas em blocos⁷⁰. O modelo completo mostrou-se mais apropriado comparativamente aos demais modelos com a inserção de apenas um bloco de variáveis ou de alguns blocos. A tabela 14 em anexo mostra o modelo *logit* utilizado na estimação do *propensity score*. Cabe destacar que em um primeiro momento a propriedade de balanceamento não foi satisfeita. Isso porque a variável *prof_superior* (no bloco 1) e *Diretor 30_39* (no bloco 3) não estavam balanceadas. Assim, com a exclusão dessas duas variáveis do modelo foi possível estimar o *propensity score*⁷¹.

Para a amostra 1 da 4ª série, a região de suporte comum foi definida como [0,03375613 a 0,99456773]. Foram eliminadas 1.411 observações que estavam fora da região de suporte comum, ficando a amostra com 498 observações (189 do grupo de tratamento e 309 do grupo de controle). Além disso, foram identificados 5 blocos que garantem que a média do *propensity score* não seja diferente entre o grupo de tratamento e o grupo de controle. No bloco 1 o intervalo do *propensity score* ficou em [0,0337561 a 0,1983259]; no bloco 2 [0,2022216 a 0,3980616]; no bloco 3 [0,4012097 a 0,5949691]; no bloco 4 [0,600171

⁷⁰ Blocos considerados: RM (se escola está na Região Metropolitana), características dos municípios, características socioeconômicas médias dos alunos das escolas e características da escola.

⁷¹ O procedimento de exclusão de variáveis não balanceadas na estimação do *propensity score* é recomendada por Khandker, Koolwal e Samad (2010, p. 183). Realizou-se também o teste LR. Pelo teste LR rejeita-se a hipótese de que os coeficientes de *prof_superior* e *dir30_39* sejam iguais a zero, porém pelo critério da BIC, o modelo sem essas duas variáveis se mostrou melhor.

a 0,7985742] e no bloco 5 [0,8031207 a 0,9945677]. O teste de propriedade de balanceamento para a amostra 1 da 4ª série é apresentado a seguir⁷².

⁷² Os testes de balanceamento para a amostra 2 da 4ª são muito semelhantes a amostra 1, portanto serão omitidos do presente estudo.

Tabela 6: Teste da propriedade de balanceamento do *propensity score* para 4ª série (Amostra1)

	Bloco 1			Bloco 2			Bloco 3			Bloco 4			Bloco 5		
	Diferença de média	Desvio Padrão	Estatística	Diferença de média	Desvio Padrão	Estatística	Diferença de média	Desvio Padrão	Estatística	Diferença de média	Desvio Padrão	Estatística	Diferença de média	Desvio Padrão	Estatística
SEXO	-0.0069	0.0993	-0.3125	0.0064	0.0914	0.2935	-0.0032	0.0805	-0.1625	0.0153	0.0749	0.7406	-0.0217	0.0905	-0.5599
RACA	-0.0033	0.1312	-0.1140	-0.0583	0.1444	-1.7301	0.0228	0.1358	0.6952	0.0395	0.1296	1.1080	0.0200	0.0932	0.5000
MAQ_LAV	0.0342	0.1042	1.4876	-0.0322	0.1002	-1.3693	-0.0132	0.0634	-0.8649	-0.0019	0.0812	-0.0863	0.0071	0.0755	0.2192
CAR	0.0197	0.2032	0.4372	-0.0694	0.2081	-1.4212	0.0334	0.1883	0.7337	0.0226	0.1718	0.4742	0.0158	0.1695	0.2171
INTER_CASA	0.0580	0.1620	1.6277	-0.0634	0.1540	-1.7658	0.0068	0.1125	0.2504	0.0038	0.1153	0.1194	-0.0073	0.1059	-0.1603
DEVER	0.0041	0.0311	0.6015	0.0052	0.0332	0.6568	-0.0027	0.0313	-0.3616	-0.0003	0.0325	-0.0337	-0.0083	0.0322	-0.6012
LER	0.0111	0.0349	1.4464	-0.0053	0.0377	-0.5898	0.0059	0.0390	0.6286	-0.0260	0.0432	-2.2554	0.0220	0.0553	0.9333
NÃO_FALTAR	0.0050	0.0268	0.8482	-0.0093	0.0279	-1.4178	-0.0017	0.0440	-0.1554	0.0031	0.0233	0.4820	0.0142	0.0317	1.0503
MAE_MEDIO	0.0052	0.0769	0.3031	-0.0203	0.0712	-1.2088	0.0276	0.0634	1.8382	-0.0182	0.0556	-1.1935	0.0001	0.0611	0.0043
MAE_SUPER	0.0148	0.0848	0.7874	-0.0103	0.0863	-0.5012	-0.0104	0.0683	-0.6297	0.0035	0.0553	0.2309	0.0178	0.0466	0.8980
MAE_NAORES	0.0083	0.1042	0.3616	0.0033	0.0926	0.1489	-0.0066	0.0832	-0.3264	0.0160	0.0848	0.6837	-0.0555	0.1070	-1.2197
TRAB	-0.0068	0.0733	-0.4189	0.0121	0.0584	0.8790	-0.0036	0.0568	-0.2632	0.0088	0.0539	0.5878	-0.0340	0.0581	-1.3792
REUN	0.0026	0.0524	0.2257	0.0066	0.0620	0.4472	-0.0125	0.0527	-0.9834	-0.0129	0.0621	-0.7487	0.0539	0.0764	1.6750
PRE_ESC	0.0329	0.1683	0.8842	-0.0736	0.1668	-1.8992	-0.0026	0.1675	-0.0642	0.1033	0.1665	2.3346	-0.0772	0.1865	-0.9701
DIR_29	-0.0381	0.0974	-1.7817	0.0345	0.1534	0.9521	-	-	-	0.0294	0.1781	0.5955	-	-	-
DIR_30_a_39	0.0735	0.3602	0.9226	0.1180	0.4486	1.1165	-0.1555	0.3796	-1.7267	-0.1383	0.4648	-1.0817	0.2576	0.4438	1.3694
DIR_40	-0.0353	0.3692	-0.4323	-0.1525	0.4600	-1.4134	0.1555	0.3796	1.7267	0.1089	0.4771	0.8266	-0.2576	0.4438	-1.3694
DIR_POS	-0.1149	0.4482	-1.1608	-0.0239	0.4545	-0.2213	0.1505	0.4934	1.2736	0.0269	0.4911	0.1975	0.0909	0.4361	0.4863
DIR_11_a_15	-0.0021	0.3355	-0.0282	0.0782	0.3394	0.9767	-0.0778	0.2820	-1.1486	-0.0636	0.2981	-0.7726	0.1364	0.2012	1.6069
DIR15	-0.0609	0.2659	-1.0368	0.0305	0.2380	0.5407	0.1026	0.2338	1.8552	-0.0465	0.1781	-0.9471	-0.0606	0.2307	-0.6135
INT_ATI	-0.0449	0.2875	-0.7054	0.0650	0.2780	0.9902	-	-	-	-0.0159	0.4497	-0.1274	0.1212	0.4187	0.6764
ROT_PROF	0.0749	0.2500	1.3578	0.0650	0.2780	0.9902	-0.1166	0.3371	-1.4490	-0.0636	0.2981	-0.7726	-0.0909	0.2783	-0.7638

REF_ESC	-0.0660	0.4482	-0.6659	-0.0027	0.4651	-0.0240	-0.0364	0.4781	-0.3143	-0.0012	0.4104	-0.0107	0.3030	0.4510	1.5926
DIR_SEL_ELE	-0.1372	0.2942	-2.1276	0.0265	0.2953	0.3786	0.1150	0.3371	1.4279	0.0588	0.2477	0.8592	-0.1061	0.2983	-0.8319
PROG_FED	-0.0233	0.4269	-0.2459	-0.0279	0.4356	-0.2694	-0.0885	0.4319	-0.8499	0.1102	0.3380	1.1872	0.1061	0.2983	0.8319
PROG_EST	0.0272	0.4708	0.2609	0.0756	0.4698	0.6795	-0.0529	0.4229	-0.5174	-0.0796	0.4870	-0.5898	0.0000	0.4747	0.0000
PROG_MUN	-0.1488	0.5006	-1.3478	-0.0305	0.4994	-0.2573	0.0612	0.5034	0.5026	0.1310	0.4999	0.9503	0.1970	0.5033	0.9168
MÉDIA_AULA	2.1893	18.4368	0.5365	7.4757	38.2189	0.8272	-11.0405	35.1067	-1.3138	-3.1874	13.9397	-0.8279	2.3258	19.8714	0.2727
MED_MAT_SERIE	-0.0837	4.4675	-0.0846	0.4752	4.6215	0.4335	-1.2550	4.9418	-1.0563	-0.1464	4.6119	-0.1143	0.8012	5.0760	0.3679
TOTAL_MAT	-2.4692	28.5194	-0.3910	3.6804	32.7714	0.4736	1.2076	39.2994	0.1268	-5.6255	39.6670	-0.5117	2.1212	41.1283	0.1201
BIBLIO	-0.0742	0.3130	-1.0726	0.0464	0.2780	0.7052	0.0265	0.3205	0.3411	0.0233	0.1270	0.6617	0.0455	0.2012	0.5270
INTER	0.0351	0.5011	0.3164	0.0133	0.5007	0.1116	-0.1257	0.5018	-1.0420	-0.0367	0.4576	-0.2891	0.2727	0.4965	1.2944
LAB_INFO	-0.1100	0.4953	-1.0048	-0.1194	0.4978	-1.0161	0.0157	0.5003	0.1296	0.1151	0.4576	0.9114	0.2879	0.4438	1.5356
PROF_SUPER	0.1624	0.2045	3.7007	0.0225	0.1864	0.5092	-0.0318	0.1978	-0.6659	-0.0567	0.1544	-1.3417	0.0393	0.2106	0.4346
PROF_POS	0.0373	0.1796	0.9401	0.0169	0.1847	0.3855	-0.0566	0.1737	-1.3618	-0.0167	0.2737	-0.2191	0.0400	0.2821	0.3304
PROF_FUND	20.8280	43.7796	2.1721	-13.6194	48.5066	-1.1927	-17.0736	65.7985	-1.0797	-4.8250	110.4203	-0.1573	38.0606	98.6122	0.9040
POP	147667	505163	1.3253	55958	538945	0.4378	-32890	533396	-0.2545	-150226	596438	-0.9131	-181669	577449	-0.7354
PIB_PER	595.6820	6887.5650	0.3906	-1725.4260	6965.7320	-1.0502	-312.3827	8182.9320	-0.1575	1557.7380	7542.4310	0.7470	803.4780	9611.9350	0.1947
IDEB_mun	-0.0344	0.3233	-0.4802	-0.0251	0.2434	-0.4342	0.0035	0.3009	0.0476	0.1013	0.2145	1.7434	-0.0258	0.1785	-0.3362
IDEB_esc	-0.0749	0.4896	-0.6916	-0.0828	0.5055	-0.6915	0.0677	0.4896	0.5715	0.0847	0.4164	0.7357	0.2864	0.3938	1.7293
Dif_IDEB	-0.0405	0.5179	-0.3536	-0.0577	0.4642	-0.5243	0.0642	0.5212	0.5090	-0.0166	0.4288	-0.1398	0.3121	0.4135	1.7979
TAXA_APRO	-0.3960	7.5614	-0.2365	-2.5493	9.7368	-1.1109	-0.3957	8.3360	-0.1959	2.0548	11.0023	0.6749	5.2197	10.8514	1.1303
MUNIC	-0.1167	0.5012	-1.0542	-0.0252	0.4994	-0.2125	0.0091	0.4962	0.0757	0.1726	0.4999	1.2593	0.1212	0.4510	0.6275

Fonte: elaboração própria a partir dos dados da Prova Brasil 2007, Censo Escolar 2007 e Informações do IBGE e INEP/MEC para o ano de 2007.

Os 5 blocos com as estimativas do *propensity score* foram definidos de forma que a média do *propensity score* não seja estatisticamente diferente entre os grupos de controle e tratamento. Conclui-se a partir dos testes de igualdade de médias que em todos os blocos as médias das variáveis não foram estatisticamente diferentes para os dois grupos⁷³.

Após a estimativa do *propensity score*, realizou-se uma série de estimações do impacto do “Programa Mais Educação” sobre o desempenho escola em língua portuguesa e matemática na Prova Brasil para a 4ª série. Nesse sentido, utilizaram-se os seguintes métodos para a estimação para fins de comparação: (a) mínimos quadrados ordinários, utilizando a nota em nível das escolas em 2009 como variável dependente e variáveis de controle em 2009 (com e sem controles); (b) diferenças em diferenças sem pareamento (com controles); (c) diferenças em diferenças com pareamento⁷⁴ (com controles) e (d) diferenças em diferenças com pareamento e ponderado pelo *propensity score* (com controles).

Destaca-se que as variáveis independentes utilizadas como controle foram acrescentadas por blocos: bloco RM⁷⁵; bloco municipal⁷⁶; bloco socioeconômico⁷⁷ e bloco escola⁷⁸.

⁷³ Embora para algumas variáveis permaneça internamente nos blocos com uma alguma diferença estatística entre os dois grupos, essa diferença não é verificada na totalidade dos blocos. Dessa forma, optou em proceder com as estimações, assim como procedeu Oliveira (2008 e 2010) no seu estudo.

⁷⁴ Considera-se como o pareamento, o procedimento de estimação com as escolas que estão definidas na região de suporte comum e que estavam balanceadas dentro dos blocos. Nesse procedimento, se exclui antes da estimação, as observações que não estavam inseridas no suporte comum. Pelos testes de balanceamento, esse procedimento já seria suficiente para a realização das estimações. Além disso, Khandker, Koolwal e Samad (2010, p. 2000) utilizam esse tipo de pareamento para a realizações de estimações quando as técnicas de PSM e diferenças em diferenças são combinadas.

⁷⁵ Variável do bloco RM: Rm_ou_polo.

⁷⁶ Variáveis do bloco município: Ideb_mun, pib_per, pop e pronasci.

⁷⁷ Variáveis do bloco socioeconômico: Sexo, raca, maq_lav, car, inter_casa, dever, ler, não_faltar, mae_medio, mae_super, mae_naores, trab, reun e pre_esc.

⁷⁸ Variáveis do bloco escola: mat_prova, taxa_ap, municipal, media_aula, total_matriculas, media_mat_turma internet_esc lab_info biblioteca prof_superior prof_pos prof_superior.

4.2.1 Resultados das estimativas do impacto na proficiência em língua portuguesa

A tabela abaixo mostra os resultados das estimativas do impacto do “Programa Mais Educação” sobre o desempenho médio dos alunos de 4ª série em língua portuguesa. A 2ª até a 7ª colunas apresentam as estimativas desse impacto, com coeficiente estimado do efeito tratamento, erro padrão robusto e estatística t, além de informações do número de escolas (observações da amostra). Já as quatro últimas colunas mostram os controles utilizados na estimação.

Tabela 7: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em língua portuguesa para a 4ª série (Amostra 1)

Amostra 1 - 4ª série	Impacto					Controles				
	Coef.	EP Rob.	t	P-valor	R2	Amos tra/tot al	RM	Mun.	Esc.	carac. socioec.
Nota PORT 2009: em nível										
(I) OLS - nível	-11.19***	0.94	-11.80	0.000	4.15%	1909	n	n	n	n
(II) OLS - nível	0.28	0.58	0.48	0.629	82.18%	1909	s	s	s	s
Diferenças em Diferenças										
(III) sem pareamento	2.55***	0.62	4.10	0.000	83.12%	3818	s	s	s	s
(IV) com pareamento	2.13***	0.75	2.83	0.005	83.35%	996	s	s	s	s
(V) com pareamento (ponderado pelo pscore)	2.34**	0.92	2.01	0.011	82.45%	996	s	s	s	s

*estimativas significativas a 10%.

**estimativas significativas a 5%.

*** estimativas significativas a 1%.

A estimativa utilizando o primeiro método de estimação, isto é, MQO e sem controles mostra-se estatisticamente significativa e negativa para a 4ª série (amostra 1), isto é, cerca de -11,19 pontos. Cabe destacar que este resultado é perfeitamente previsível, pois nessa estimação não foi utilizado qualquer tipo de pareamento e qualquer adição de variáveis de controle entre as escolas. Sabe-se pelos testes de médias e de distribuição, apresentados na seção anterior, que as escolas pertencentes ao grupo de controle e tratamento apresentam características pré-tratamento muito diferentes, com vantagem em quase todas as variáveis para o grupo que não participou do programa, principalmente nos fatores relacionados às

características socioeconômicas dos alunos, escolas e do município em que elas estão inseridas, o que repercute no desempenho médio dos alunos de 4ª série das escolas. A estimativa do efeito do programa passa a ser positivo em 0,28 pontos quando é realizada uma tentativa de controle de características através da introdução de co-variáveis. Contudo, com a adição de controles, a estimação é não significativa estatisticamente.

A segunda estimação, mas robusta que o primeira, foi realizado através do método de diferenças em diferenças. Primeiramente foi estimado o modelo de diferenças em diferenças, sem qualquer tipo de pareamento. O estimativa do efeito de tratamento médio sobre os tratados (ATT) para esse método mostrou-se estatisticamente significativo. O impacto do programa é de 2,55 ponto. Ou seja, o “Programa Mais Educação” contribuiu com a elevação em 2,55 pontos nas notas médias de língua portuguesa, a nível de significância de 1%.

Em relação a estimação do modelo de diferenças em diferenças com a realização do pareamento, o ATT estimado é de 2,13 pontos, a 1% de significância. Ressalta-se que esse resultado é próximo ao encontrado na estimação do método de diferenças em diferenças com pareamento e ponderado com o *propensity score*. A estimativa de impacto do programa passa a ser de 2,34⁷⁹ pontos. Cabe ressaltar que não há grandes diferenças entre os coeficientes nas estimativas, o que dá maior confiabilidade aos sinais e significância do impacto estimado. Assim, as evidencias mostram que o “Programa Mais Educação” impacta positivamente o desempenho escolar em língua portuguesa para a 4ª série.

Na amostra 2, ou seja, considerando o grupo de tratamento somente as escolas participantes em 2009, observa-se que os resultados são bastante parecidos com a amostra 1 (tabela 16 em anexo). Cabe destacar que o impacto positivo na proficiência de português do programa, utilizando os métodos de estimação, é inferior que a amostra 1⁸⁰. Isso significa que as escolas que iniciaram o programa em 2008 apresentam um impacto maior do que as que iniciaram somente em 2009.

⁷⁹ Neste caso a 5% de significância.

⁸⁰ Com exceção para a estimação em diferenças em diferenças sem pareamento.

4.2.2 Resultados das estimativas do impacto na proficiência em matemática

A tabela abaixo (tabela 8) apresenta os resultados das estimativas do impacto do “Programa Mais Educação” sobre o desempenho médio dos alunos de 4ª série das escolas na proficiência em matemática.

Tabela 8: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em matemática para a 4ª série (Amostra 1)

Amostra 1 - 4ª série	Impacto					Controles				
	Coef	EP Rob.	T	P-valor	R2	Amostra / total	RM	Mun	Esc	Carac. Socio.
Nota MAT 2009: em nível										
(I) OLS – nível	-14.49***	-1.31	-13.73	0.000	5.16%	1909	n	n	n	n
(V) OLS – nível	-1.54**	0.69	-2.22	0.026	80.66%	1909	s	s	s	s
Diferenças em Diferenças										
(III) sem pareamento	-3.94***	0.71	-5.58	0.000	82.84%	3818	s	s	s	s
(III) com pareamento	-2.38***	0.83	-2.87	0.004	82.38%	996	s	s	s	s
(III) com pareamento (ponderado pelo pscore)	-3.04***	0.99	-3.08	0.002	81.54%	996	s	s	s	s

*estimativas significativas a 10%.

**estimativas significativas a 5%.

*** estimativas significativas a 1%.

As estimativas utilizando o primeiro método de estimação se mostram estatisticamente significativas e negativas para a amostra 1. A estimativa do efeito do programa é de -14,49 pontos, sem controle, e cai para -1,54 pontos quando é realizada a tentativa de controle de características através da introdução de co-variáveis.

Como as escolas pertencentes ao grupo de controle e tratamento apresentam características pré-tratamento muito diferentes para a amostra da 4ª série, com vantagem em quase todas as variáveis para o grupo que não participou do programa, sabe-se que esse resultado para matemática apresenta algum tipo de viés, assim como nos resultados para a língua portuguesa.

Com o método de diferenças em diferenças, as estimativas ainda se mostraram estatisticamente significativas e negativas tanto para as estimativas sem pareamento, com pareamento e com pareamento ponderada pelo *propensity score*. Ressalta-se estes resultados

são próximos, isto é, o ATT estimado no método de diferenças em diferenças sem pareamento é de -3,94 pontos, com pareamento é de -2,39 pontos e com pareamento e ponderado pelo *propensity score* é de -3,04 pontos.

Isso é, para o procedimento considerado mais robustos, o método de diferenças em diferenças, o impacto do programa na proficiência em matemática se mostra negativo. Tanto os sinais e quanto a significância do impacto estimado são mantidos mesmo com a introdução do pareamento e a ponderação pelo *propensity score*. Isto é, há evidências de que o programa apresenta algum impacto negativo sobre a proficiência em matemática para a 4ª série, em termos de importância estatística das estimações realizadas.

Em relação a amostra 2, isto é, considerando o grupo de tratamento somente as escolas participantes em 2009, constata-se que os resultados são bastante parecidos com a amostra 1 (tabela 17 em anexo). Cabe destacar apenas que o impacto negativo, utilizando o método de diferenças em diferenças com pareamento e ponderado pelo *propensity score*, é menor do que a amostra 1, isto é, -2,78 ao invés de -3,04 pontos.

4.3 Resultados para a 8ª Série

As estimações para a 8ª série também foram realizadas a partir dos mesmos procedimentos utilizados para a 4ª série. Isto é, para a estimação do *propensity score* das amostras 1 e 2 da 8ª série utilizou-se um modelo *logit*. Para a escolha do modelo *logit* mais adequado procedeu-se com um teste LR e com a verificação dos critérios de informações de *Akaike* – AIC e *Schwarz* – SIC para um conjunto das variáveis inseridas em blocos⁸¹. O modelo completo mostrou-se mais apropriado em relação aos demais modelos com a inserção de apenas um bloco de variáveis ou de alguns blocos. Além disso, assim como para a 4ª série, as variáveis do modelo foram pré-selecionadas a partir da descrição da literatura sobre os fatores que afetam desempenho escolar e considerando a probabilidade de participação conforme os critérios estabelecidos pelo MEC para seleção de escolas que poderiam ser atendidas pelo programa.

⁸¹ Blocos considerados: RM (se escola está na Região Metropolitana), características dos municípios, características socioeconômicas médias dos alunos das escolas e características da escola

A tabela 15 em anexo apresenta o modelo *logit* utilizado na estimação do *propensity score*. Para a amostra 1 da 8^a série, a região de suporte comum definida foi de [0,01135357 a 0,99068666]. Foram eliminadas 1.132 observações que estavam fora da região de suporte comum, ficando a amostra com 400 observações (137 escolas do grupo de tratamento e 263 escolas do grupo de controle). Além disso, foram identificados 6 blocos que garantem que a média do *propensity score* não seja diferente entre o grupo de tratamento e o grupo de controle. No bloco 1 o intervalo do *propensity score* ficou em [0,0113536 a 0,0997308]; no bloco 2 [0,1017099 a 0,1984549]; no bloco 3 [0,2019927 a 0,3937869]; no bloco 4 [0,4034579 a 0,5998557]; no bloco 5 [0,6004164 a 0,7978771] e no bloco 6 [0,8001507 a 0,9906867]. Na tabela a seguir, o teste de propriedade de balanceamento para a amostra 1 da 8^a série é apresentado.

Tabela 9: Teste da propriedade de balanceamento do *propensity score* para 8ª série (Amostra1)

	Bloco 1			Bloco 2			Bloco 3			Bloco 4			Bloco 5			Bloco 6		
	Diferença de média	Desvio Padrão	Estatística	Diferença de média	Desvio Padrão	Estatística	Diferença de média	Desvio Padrão	Estatística	Diferença de média	Desvio Padrão	Estatística	Diferença de média	Desvio Padrão	Estatística	Diferença de média	Desvio Padrão	Estatística
SEXO	-0.0488	0.0991	-0.8415	0.0129	0.0956	0.3912	0.0098	0.1146	0.3347	-0.0089	0.1027	-0.3038	-0.0104	0.0924	-0.3448	0.0131	0.0983	0.2799
RACA	0.0175	0.1416	0.2110	0.0404	0.1373	0.8550	0.0111	0.1212	0.3556	-0.0189	0.1418	-0.4691	0.0010	0.1253	0.0244	-0.0207	0.1306	-0.3338
MAQ_LAV	0.0180	0.1155	0.2652	-0.0072	0.1115	-0.1858	-0.0138	0.0988	-0.5444	0.0368	0.1260	1.0398	0.0101	0.0846	0.3682	-0.0019	0.0927	-0.0422
CAR	-0.0155	0.2317	-0.1142	0.0200	0.1614	0.3581	-0.0223	0.1894	-0.4603	0.0449	0.2078	0.7649	0.0274	0.1699	0.4972	-0.1001	0.2158	-0.9847
INTER_CASA	0.0483	0.1991	0.4130	-0.0099	0.1774	-0.1611	0.0133	0.1647	0.3144	0.0409	0.1734	0.8356	-0.0228	0.1467	-0.4790	-0.0362	0.1402	-0.5442
DEVER	-0.0312	0.0364	-1.4729	0.0085	0.0484	0.5107	-0.0045	0.0374	-0.4740	0.0065	0.0361	0.6384	0.0008	0.0461	0.0501	-0.0070	0.0483	-0.3041
LER	-0.0074	0.0677	-0.1873	-0.0342	0.0669	-1.5063	0.0066	0.0637	0.4041	0.0026	0.0573	0.1581	0.0254	0.0620	1.2815	-0.0061	0.0607	-0.2121
NÃO_FALTAR	-0.0087	0.0158	-0.9472	0.0067	0.0176	1.1174	-0.0008	0.0141	-0.2098	0.0011	0.0181	0.2105	-0.0054	0.0211	-0.7869	0.0063	0.0199	0.6715
MAE_MEDIO	0.0660	0.1260	0.8945	0.0261	0.1095	0.6921	-0.0043	0.1090	-0.1528	0.0312	0.1179	0.9404	-0.0276	0.0984	-0.8677	-0.0252	0.0901	-0.5909
MAE_SUPER	0.0138	0.0837	0.2804	-0.0148	0.0665	-0.6452	0.0221	0.0621	1.4076	0.0029	0.0497	0.2065	-0.0088	0.0489	-0.5532	-0.0253	0.0515	-1.0458
MAE_NAORES	-0.0269	0.0684	-0.6718	0.0105	0.0726	0.4194	-0.0129	0.0770	-0.6524	0.0138	0.0902	0.5413	0.0033	0.0742	0.1375	-0.0323	0.0646	-1.0635
TRAB	0.0332	0.1013	0.5580	0.0211	0.0712	0.8623	0.0190	0.0800	0.9304	-0.0480	0.0820	-2.1544	-0.0062	0.0746	-0.2542	0.0005	0.0812	0.0120
REUN	0.0690	0.0936	1.2637	0.0032	0.1031	0.0905	-0.0368	0.0816	-1.7982	-0.0139	0.0822	-0.5976	0.0324	0.1136	0.8832	0.0501	0.1064	1.0002
PRE_ESC	-0.0087	0.1380	-0.1076	-0.0327	0.1489	-0.6368	0.0351	0.1360	1.0130	-0.0189	0.2072	-0.3207	0.0602	0.1597	1.1753	0.0038	0.1495	0.0537
DIR_29	0.0085	0.0913	0.1595	0.0536	0.2099	0.7409	-	-	-	-0.0435	0.1400	-1.1058	-	-	-	-	-	-
DIR_30_a_39	0.1111	0.3121	0.6072	0.0250	0.3289	0.2198	-0.1705	0.4200	-1.6129	-0.0109	0.4401	-0.0869	0.2028	0.4584	1.3882	-0.0174	0.4154	-0.0880
DIR_40	-0.1197	0.3224	-0.6332	-0.0786	0.3755	-0.6065	0.1705	0.4200	1.6129	0.0543	0.4507	0.4250	-0.2028	0.4584	-1.3882	0.0174	0.4154	0.0880
DIR_POS	0.1197	0.4137	0.4931	0.0250	0.4888	0.1479	0.0310	0.4674	0.2585	0.0264	0.4761	0.1951	0.0115	0.4204	0.0841	-0.2957	0.4761	-1.3288
DIR_11_a_15	0.1538	0.3586	0.7324	-0.0286	0.2666	-0.3100	-0.0368	0.2387	-0.6024	0.0357	0.1400	0.9047	0.0069	0.2523	0.0841	-0.0217	0.1400	-0.3267
DIR15	0.0342	0.1803	0.3231	0.0536	0.2099	0.7409	0.0281	0.2387	0.4592	-0.0947	0.2715	-1.2465	0.0392	0.2084	0.5793	-0.0870	0.2715	-0.6764
INT_ATI	-0.2051	0.3414	-1.0280	0.0714	0.2404	0.8637	-0.0504	0.3436	-0.5726	0.0559	0.3254	0.6066	-0.0184	0.3665	-0.1544	0.0522	0.4826	0.2274
ROT_PROF	0.1282	0.3321	0.6587	0.0071	0.3103	0.0665	0.0097	0.2877	0.1312	0.0280	0.2376	0.4145	-0.1866	0.4045	-1.4507	0.1348	0.2715	1.0554

REF_ESC	0.0342	0.4602	0.1265	-0.0214	0.4693	-0.1320	-0.0853	0.4910	-0.6788	-0.0730	0.5002	-0.5147	0.2373	0.4903	1.5257	0.0174	0.4154	0.0880	
DIR_SEL_ELE	0.0427	0.2007	0.3629	0.0536	0.2099	0.7409	0.0930	0.2387	1.5453	-0.0947	0.2715	-1.2465	-0.0645	0.2084	-0.9605	-0.0652	0.2376	-0.5789	
PROG_FED	0.0085	0.4703	0.0309	-0.0500	0.4318	-0.3350	0.1056	0.4200	0.9868	-0.1118	0.4284	-0.9260	0.0184	0.3665	0.1544	-0.0913	0.3254	-0.5920	
PROG_EST	0.4530	0.4987	1.5630	0.0821	0.5029	0.4729	0.1415	0.5037	1.1041	0.0450	0.4883	0.3248	-0.2696	0.4954	-1.7275	-0.3261	0.4602	-1.5246	
PROG_MUN	-0.0940	0.4299	-0.3727	0.1571	0.4750	0.9631	-0.1143	0.4784	-0.9372	-0.1925	0.4883	-1.4152	0.2488	0.4954	1.5865	-0.0565	0.5025	-0.2366	
MÉDIA_AULA MED_MAT_SE RIE	1.7692	33.4117	0.0902	-1.0060	13.0689	-0.2226	1.5226	11.6700	0.5092	-1.3296	14.4459	-0.3241	0.2266	13.5237	0.0514	-5.4891	15.1292	-0.7673	
	-3.3806	5.6624	-1.0213	-0.9360	5.2277	-0.5186	2.1946	5.5918	1.5568	0.2445	5.5892	0.1539	-3.7496	6.1855	-1.9408	1.2540	5.5318	0.4777	
TOTAL_MAT	-4.5812	33.1851	-0.2352	-15.7821	31.9012	-1.4533	9.9079	40.5848	0.9575	-0.5078	34.5494	-0.0517	11.7696	36.6620	-0.9969	28.0087	43.2865	1.3867	
BIBLIO	0.2821	0.2354	2.0780	-	-	-	-0.0514	0.2648	-0.7588	0.0435	0.1400	1.1058	-0.0392	0.2084	-0.5793	0.0217	0.1400	0.3267	
INTER	-0.3590	0.4790	-1.2853	-0.1393	0.4693	-0.8628	-0.0620	0.4873	-0.4966	0.2283	0.4826	1.7128	0.0300	0.4903	0.1876	0.2391	0.4154	1.2287	
LAB_INFO	-0.1197	0.4996	-0.4082	-0.0893	0.4980	-0.5193	-0.1550	0.4992	-1.2235	0.0062	0.5002	0.0437	0.2442	0.4682	1.6516	0.1522	0.3475	0.9286	
PROF_SUPER	0.1181	0.1057	1.9331	-0.0207	0.1058	-0.5655	-0.0052	0.1155	-0.1749	0.0162	0.1071	0.5346	-0.0069	0.1262	-0.1688	-0.0269	0.1177	-0.4820	
PROF_POS	0.0898	0.1980	0.7746	0.0210	0.1438	0.4236	-0.0333	0.1973	-0.6605	0.0120	0.1717	0.2463	-0.0108	0.2429	-0.1362	-0.0219	0.3041	-0.1514	
PROF_FUND	4.1197	43.7337	0.1604	-7.0643	38.8112	-0.5272	-12.9777	49.5520	-1.0283	7.5776	55.7531	0.4792	-5.4793	73.0332	-0.2305	34.9652	117.189	8	0.6298
POP	328467	509170	1.1043	-82423	582751	-0.4093	-31552	592225	-0.2076	179236	575351	1.1096	82425	188630	0.4370	-304843	565670	-1.1481	
PIB_PER	4848.84	6307.21		1461.73	5994.87		1379.792	7977.28		2180.39	8577.24		927.984	9045.44		3121.078	8477.68		-0.7787
	60	30	1.3189	00	90	0.7075	0	00	-0.6760	80	00	0.9016	4	70	-0.3153	0	00		
IDEB_mun	0.0214	0.2679	0.1359	0.1064	0.2219	1.4079	-0.0578	0.1910	-1.1927	0.0214	0.1829	0.4128	-0.0438	0.1686	-0.8030	0.0491	0.1629	0.6365	
IDEB_esc	-0.0624	0.5211	-0.2039	-0.0686	0.6066	-0.3270	-0.1868	0.4933	-1.5005	0.2009	0.5773	1.2435	0.1014	0.5537	0.5642	0.1204	0.5124	0.4953	
Dif_IDEB	-0.0838	0.5333	-0.2676	-0.1750	0.6410	-0.7930	-0.1290	0.4835	-1.0476	0.1795	0.5567	1.1496	0.1452	0.5819	0.7711	0.0713	0.5338	0.2810	
TAXA_APRO	5.2120	12.4746	0.7131	-1.9639	10.9254	-0.5207	-5.9763	14.0500	-1.6928	5.9612	14.4174	1.4870	-0.3353	15.0488	-0.0684	6.2478	13.5463	0.9791	
MUNIC	-0.0684	0.4441	-0.2623	0.0107	0.4954	0.0625	-0.1328	0.4873	-1.0704	-0.0575	0.5049	-0.4010	0.1590	0.5045	0.9781	0.1043	0.4602	0.4778	

Fonte: elaboração própria a partir dos dados da Prova Brasil 2007, Censo Escolar 2007 e informações do IBGE e INEP/MEC para o ano de 2007.

Para a 8ª série, 6 blocos com as estimativas do *propensity score* foram definidos de forma que a média do *propensity score* não seja estatisticamente diferente entre os grupos de controle e tratamento. Consta-se a partir dos testes de igualdade de médias que em todos os blocos as médias das variáveis não foram estatisticamente diferentes para os dois grupos.

Com a estimação do *propensity score*, procedeu-se com uma série de estimativas do impacto do programa sobre o desempenho médio dos alunos de 8ª série das escolas em língua portuguesa e matemática na Prova Brasil. Destaca-se que as variáveis independentes dos modelos utilizadas como controle foram acrescentadas por blocos: bloco RM⁸²; bloco municipal⁸³; bloco socioeconômico⁸⁴ e bloco escola⁸⁵, assim como na 4ª série. A seguir apresentam-se os resultados do impacto do “Programa Mais Educação” para língua portuguesa e matemática para a 8ª série.

4.3.1 Resultados das estimativas do impacto na proficiência em língua portuguesa

Na tabela abaixo são apresentados os resultados das estimativas do impacto do “Programa Mais Educação” sobre o desempenho médio dos alunos de 8ª série em língua portuguesa. As colunas 2 a 7 mostram as estimativas do impacto do programa, com coeficiente estimado do efeito tratamento, erro padrão robusto e estatística t, além de informações do número de escolas da amostra. Já as quatro últimas colunas apresentam os controles que foram utilizados para a estimação.

⁸² Variável do bloco RM: Rm_ou_polo.

⁸³ Variáveis do bloco município: Ideb_mun, pib_per, pop e pronasci.

⁸⁴ Variáveis do bloco socioeconômico: Sexo, raca, maq_lav, car, inter_casa, dever, ler, não_faltar, mae_medio, mae_super, mae_naores, trab, reun e pre_esc.

⁸⁵ Variáveis do bloco escola: Port_prova, taxa_ap, municipal, media_aula, total_matriculas, media_mat_turma internet_esc lab_info biblioteca prof_superior prof_pos prof_superior.

Tabela 10: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em língua portuguesa para a 8ª série (Amostra 1)

Amostra 1 - 8ª série	Impacto					Controles				
	Coef	EP Rob.	t	P-valor	R2	Amostra / total	RM	Mun	Esc	Carac. Socio.
Nota PORT 2009: em nível										
(I) OLS – nível	-9.41***	1.32	-7.11	0.000	2.70%	1532	n	n	n	n
(V) OLS – nível	-1.29	0.93	-1.38	0.167	74.66%	1532	s	s	s	s
Diferenças em Diferenças										
(III) sem pareamento	-1.07	1.06	-1.01	0.312	76.83%	3064	s	s	s	s
(III) com pareamento	-0.55	1.24	-0.45	0.655	77.26%	800	s	s	s	s
(III) com pareamento (ponderado pelo pscore)	1.63	1.49	1.10	0.273	81.71%	800	s	s	s	s

*estimativas significativas a 10%.

**estimativas significativas a 5%.

*** estimativas significativas a 1%..

Assim como para a 4ª série, a estimativa utilizando o primeiro método de estimação (MQO em nível e sem controle) mostram-se estatisticamente significativas e negativas para a 8ª série (amostra 1). Conforme os testes de médias e de distribuição, realizados antes do pareamento, sabe-se que as escolas pertencentes aos dois grupos apresentam características pré-tratamento muito desiguais, com vantagem na grande maioria das vezes para o grupo de controle. Daí o resultado negativo também para a 8ª série. Para língua portuguesa, o efeito do programa cai de -9,41 pontos para -1,29 pontos quando é realizada uma tentativa de controle de características através da introdução de variáveis independentes. Entretanto, o efeito do programa passa a ser não significativo.

Contudo, para as demais estimações, isto é, as estimações realizadas com o método de diferenças em diferenças, não foi encontrado qualquer evidência que o programa tenha impactado o desempenho escolar em língua portuguesa medido nos testes de proficiência, tanto para a amostra 1, quanto para a amostra 2.

4.3.2 Resultados das estimativas do impacto na proficiência em matemática

A tabela a seguir (tabela 11) mostra os resultados das estimativas do impacto do “Programa Mais Educação” sobre o desempenho médio dos alunos de 8ª série das escolas na proficiência em matemática.

Tabela 11: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em matemática para a 8ª série (Amostra 1)

Amostra 1 - 8ª série	Impacto					Amostra / total	Controles			
	Coef	EP Rob.	t	P-valor	R2		RM	Mun	Esc	Carac. Socio.
Nota MAT 2009: em nível										
(I) OLS – nível	-12.93***	1.41	-9.18	0.000	3.94%	1532	n	n	n	n
(V) OLS – nível	0.52	0.93	0.56	0.576	78.16%	1532	s	s	s	s
Diferenças em Diferenças										
(III) sem pareamento	-1.07	1.06	-1.01	0.312	76.83%	3064	s	s	s	s
(III) com pareamento	0.42	1.23	0.34	0.732	76.59%	800	s	s	s	s
(III) com pareamento (ponderado pelo pscore)	-2.14	1.40	-1.53	0.127	81.13%	800	s	s	s	s

*estimativas significativas a 10%.

**estimativas significativas a 5%.

*** estimativas significativas a 1%..

Conforme a tabela 11, assim como em todas as outras estimações, os resultados encontrados com a utilização do primeiro método de estimação (MQO e sem controle) se mostram estatisticamente significativos e negativos para o desempenho em matemática da amostra 1 da 8ª série, dado a possibilidade de existência de grande viés devido à heterogeneidade dos grupos. A estimativa do efeito do programa é de -12,93 pontos. Entretanto, para as demais estimações, com o método de diferenças em diferenças, não foi encontrado qualquer evidência de impacto do programa sobre a proficiência em matemática (tanto para a amostra 1, quanto para a amostra 2).

4.4 Análises de robustez

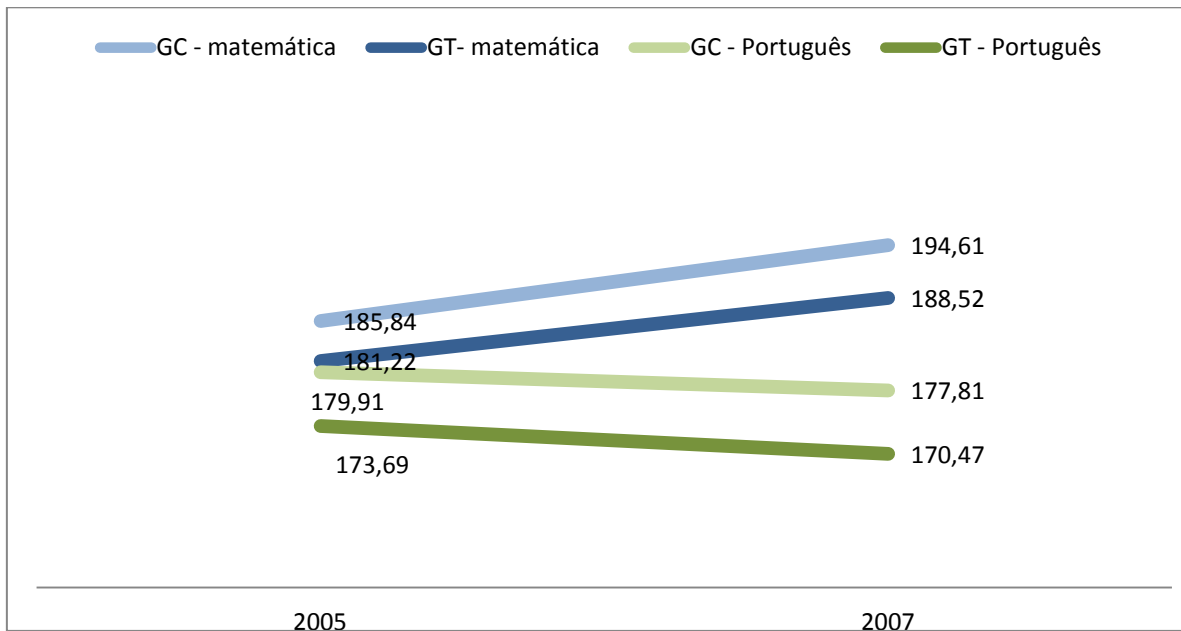
Em relação a análise de robustez, dois procedimentos para testar a validade dos modelos foram realizados. Primeiramente foi testado se a utilização de um grupo de tratamento “placebo” seria capaz de impactar o desempenho escolar. O grupo de tratamento placebo foi constituído de escolas que entraram no programa somente em 2010, ou seja, posterior aos desempenhos escolares de 2009. Foram identificadas 40 escolas do grupo de controle que passaram a participar do programa em 2010. Em relação as estimativas para a 4ª e 8ª série, nenhuma das estimativas em diferenças em diferenças tanto para português como para matemática foram estatisticamente significativas. Assim, como era de se esperar, o impacto do programa não é significativo estatisticamente quando é introduzido um grupo de tratamento “falso” para a avaliação de impacto do “Programa Mais Educação”, o que traz confiabilidade nas estimativas geradas pelo estudo.

No segundo procedimento foi testado se hipótese de trajetórias paralelas anterior ao programa é atendida⁸⁶. Se a hipótese for atendida, então há evidências de não viés, e conseqüentemente as análises que apresentam impacto com significância estatística estão corretamente estimados.

Ao analisar a trajetória dos dois grupos para a 4ª série nas proficiência em matemática e português, no gráfico 11, observa-se que o grupo de tratamento tem menores notas de proficiência do que do grupo de controle, também para 2005. Para português as notas em 2007 são menores nos dois grupos comparativamente a 2005. Na proficiência em matemática, há uma elevação das notas dos dois grupos de 2005 para 2007. Ao observar as linhas, constata-se que a trajetória tanto para português quanto para matemática são quase perfeitamente paralelas, o que sugere que o impacto estimado significativo para as notas na 4ª série não tem viés.

⁸⁶ Considerou-se para composição da amostra somente as escolas dentro do suporte comum.

Gráfico 11: Notas médias das escolas por grupo em matemática e português para a 4ª série, 2005 e 2007

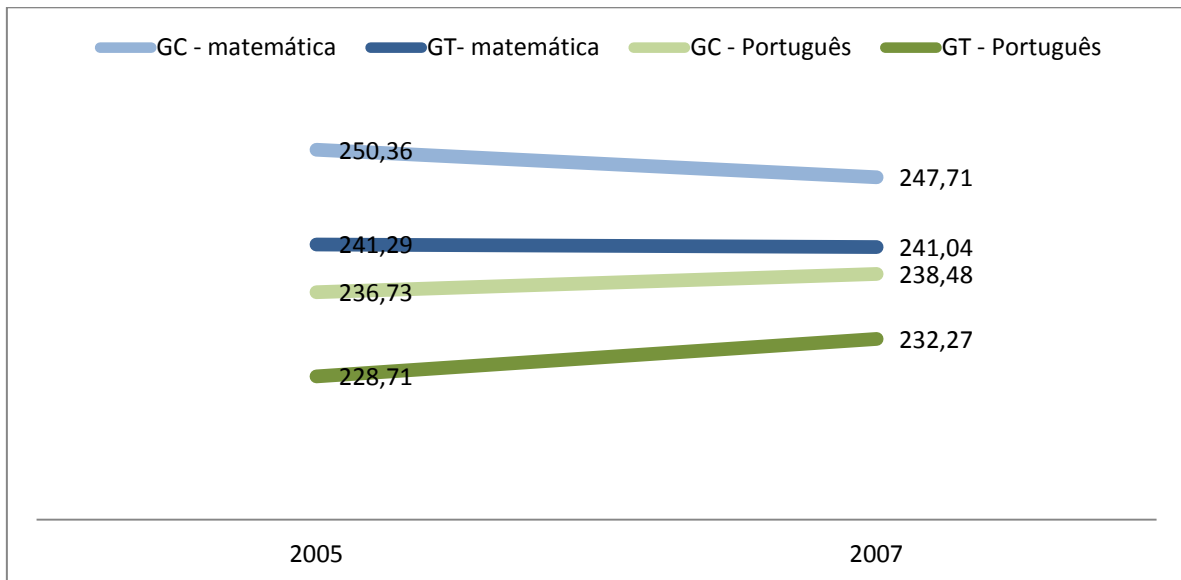


Fonte: Elaboração própria

Nota: GC – Grupo de Controle; GT – Grupo de Tratamento

O gráfico 12 mostra o desempenho para a 8ª série. Assim como na 4ª série, o grupo de tratamento tem menores notas de proficiência em português e matemática do que do grupo de controle, em 2005. Para português, as notas em 2007 são maiores nos os dois grupos comparativamente a 2005. Na proficiência em matemática, há uma redução das notas dos dois grupos de 2005 para 2007. Em relação a análise das linhas, observa-se que a trajetória tanto para português quanto para matemática são quase perfeitamente paralelas, o que sugere que o resultado não significativo para as notas na 8ª série não apresenta viés.

Gráfico 12: Notas médias das escolas por grupo em matemática e português para a 8ª série, 2005 e 2007



Fonte: Elaboração própria

Nota: GC –Grupo de Controle; GT – Grupo de Tratamento

4.5 Discussão dos resultados

O objetivo geral do estudo foi avaliar o impacto do turno integral, expresso pelo “Programa Mais Educação”, no desempenho médio dos alunos das escolas da rede pública no Rio Grande do Sul de 4ª série e 8ª série nos testes padronizados de língua portuguesa e matemática da Prova Brasil. Com o método de diferenças em diferenças e a realização de um pareamento adequado, os resultados mostram que o programa impactou positivamente o desempenho médio das escolas em língua portuguesa da 4ª série, expresso pelo teste de proficiência em português na Prova Brasil. O impacto do “Programa Mais Educação” foi de 2,34 pontos.

Por outro lado, para a 4ª série, o impacto do programa no desempenho em matemática na Prova Brasil não surtiu o mesmo efeito. Após a realização do pareamento e considerando a

ponderação pelo *propensity score*, a estimação de diferenças em diferenças mostra um impacto negativo de -3,04 pontos. Ou seja, considerando um método robusto de estimação, o “Programa Mais Educação” impactou negativamente no desempenho escolar dos alunos em matemática na 4^a série.

Cabe destacar que os resultados encontrados estão coerentes com a literatura. Aquino e Kassouf (2011) avaliaram o impacto do programa “Escola de Tempo Integral” da rede pública estadual de São Paulo sobre a proficiência média das escolas em matemática e português no Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP para a 8^a série. O programa não surtiu efeito sobre o desempenho em matemática, porém os resultados para língua portuguesa resultados foram positivos. Já o estudo de Franco (2009) não encontrou efeito sobre a proficiência em matemática do SAEB para alunos de 4^a série da rede pública de ensino que estudam mais que 5 horas diárias. Ou seja, em outros estudos há evidências de que os programas de educação integral, no formato que estão constituídos, impactem positivamente o desempenho em língua portuguesa e tenham efeito nulo em matemática. Contudo, não foram encontradas razões plausíveis na literatura para o efeito nulo (não significativo estatisticamente) da ampliação da jornada escolar na aprendizagem de matemática.

Além disso, não há estudos que mostrem que a educação integral impacte negativamente o desempenho em matemática. No caso específico do “Programa Mais Educação” para as estimações mais robustas, o programa apresentou um impacto negativo para a proficiência para a 4^a série na Prova Brasil (resoluções de problemas). Uma das razões para isso pode ser o fato da pouca ênfase dada ao programa ao desenvolvimento cognitivo dos alunos no ensino da matemática para a resolução de problemas (ênfase dada na Prova Brasil), o que pode ter prejudicado o desempenho médio dos alunos das escolas nos testes de proficiência. Sabe-se que o *kit* de matemática do ensino fundamental compreende jogos para o desenvolvimento do raciocínio matemático, porém não há nos *kit* metodologia específica e nem orientação para construção de problemas matemáticos para os alunos resolverem (esse compreende a capacidade do aluno em montar estratégias, raciocinar logicamente e verificar se sua estratégia foi válida)⁸⁷.

⁸⁷ Segundo Rodrigues e Magalhães (2011, p.1), a “Resolução de Problemas é uma metodologia de ensino de Matemática muito eficaz, pois propicia uma mobilização de saberes no sentido de buscar a solução”.

Além disso, destacam-se também que o MEC só estabelece entre atividades obrigatórias, apenas uma atividade específica do macrocampo “acompanhamento pedagógico”, sendo que há diversas atividades incluídas nesse macrocampo, além das atividades com matemática. Incluem-se nesse macrocampo as atividades de letramento, línguas estrangeiras, ciências, história e geografia. Ou seja, não necessariamente, as atividades extra-classes se concentraram no desenvolvimento cognitivo da matemática (ênfase na resolução de problemas). Ressalta-se também a necessidade de interação dessa matéria com outras atividades desenvolvidas do programa, como por exemplo, no macrocampo da inclusão digital com atividades que contemplem o uso de *software* educacionais específicos para o ensino da matemática (desafios e etc).

Uma outra hipótese que se apresenta em relação ao impacto negativo do programa sobre o desempenho médio escolar em matemática na 4ª série no Rio Grande do Sul pode estar relacionada ao fato de que o programa tenha retido os alunos, que sem o programa, estariam fora da escola no ano de 2009. Ou seja, ao impedir que os alunos com piores desempenhos abandonassem a escola, e considerando que estes tenham as piores notas em matemática, é possível que o desempenho médio em matemática dos alunos participantes do programa tenha diminuído ao longo do tempo. Se por um lado o programa pode não ter contribuído com a melhoria no desempenho em matemática, por outro, ele pode ter tido êxito na redução da taxa de evasão escolar.

Em relação aos resultados da 8ª série. Há duas razões plausíveis para o efeito nulo do programa na proficiência média dos alunos das escolas tanto de matemática como de português. Primeiramente, é possível que o programa não tenha funcionado adequadamente para os alunos das séries finais do ensino fundamental das escolas do Rio Grande do Sul. Uma outra hipótese é que esses alunos não tenham sido contemplados com o programa, e sendo assim, as secretarias municipal e estadual, tenham priorizado em um primeiro momento a implementação do programa nas séries iniciais do ensino fundamental. Cabe ressaltar que nesse último caso, a recomendação do MEC às unidades executoras do programa (MEC(2009b), era para selecionar turmas de participantes dos programas que contemplassem alunos das séries finais da 2ª fase do ensino fundamental (8º e/ou 9º anos), além dos alunos das séries finais da 1ª fase do ensino fundamental (4º e/ou 5º anos). Por outro lado, caso estes não tenham participado do programa, a análise realizada na seção anterior contribui como teste de robustez à presente pesquisa. Isso porque o teste de robustez tem como objetivo

verificar se a mesma metodologia seria capaz de estimar algum impacto a uma população que não foi exposta diretamente ao tratamento (BIONDI; VASCONCELLOS; MENEZES-FILHO, 2009 e VASCONCELLOS; BIONDI; MENEZES-FILHO, 2009), mas que de alguma forma esteja relacionada a um grupo que recebeu o tratamento. Esperar-se-ia assim que esse impacto fosse nulo para a 8ª série, caso contrário alguma variável não observada estaria gerando viés aos resultados das estimativas para a 4ª série.

Em relação a primeira possibilidade para o resultado da 8ª série (e também a resultado negativo da 4ª série para matemática) , assim como destaca a literatura sobre avaliação de políticas públicas (CANO, 2006), é necessário a introdução de outros tipos de avaliação complementares a avaliação de impacto para verificar as razões do resultado nulo ou negativo de um programa, tais como as avaliações de processo, conforme visto na revisão da literatura (capítulo 2). Assim, buscar-se-ia saber se determinado resultado deficiente pode ser atribuído a dificuldades de implementação e/ou não eficácia na intervenção. Sabe-se apenas de algumas as dificuldades já levantadas pela literatura para a implementação do “Programa Mais Educação” em outras localidades do Brasil.

O relatório de avaliação do TCU (2010), por exemplo, aponta dificuldades na execução do “Programa Mais Educação” no município de João Pessoa/PB. Em visitas técnicas foi constatado falhas na distribuição da merenda escolar e acompanhamento técnico da execução do programa por parte de nutricionistas. Além disso, houve também atraso de repasses pelo MEC, o que inviabilizou o início do programa em uma escola auditada. Já Parente e Azevedo (2011) realizaram o monitoramento do programa nas escolas do município de Itabaiana/SE em 2010. As principais dificuldades levantadas para a implementação do programa foram: espaço para desenvolvimento das oficinas, espaço para banho e higienização, troca constante de monitores no processo, número reduzido e sobrecarga de trabalho para o pessoal de apoio, evasão no programa e rivalidade entre professor e monitores.

Contudo, como esses problemas podem ser localizados e específicos a uma realidade, é necessária uma avaliação do andamento do programa, particular e generalizada, no Rio Grande do Sul. Uma alternativa interessante seria separar as escolas em dois grupos de tratamento (um com escolas que apresentaram bom andamento do programa e outro com as escolas que apresentaram os maiores problemas), para verificar se há impacto diferenciado para esses dois grupos. Ou seja, a avaliação de processo, combinada com a avaliação de impacto, ajudaria na melhor compreensão das dificuldades de implementação do programa.

Assim como outras avaliações realizadas (BIONDI; VASCONCELLOS; MENEZES-FILHO, 2009 e VASCONCELLOS; BIONDI; MENEZES-FILHO, 2009) foi possível detectar um maior impacto conforme o número de anos de participações no programa. Isso porque as escolas que iniciaram o programa em 2009 (amostra 2) obtiveram resultados inferiores do que o conjunto de todas as escolas participantes (sejam as que iniciam em 2008 ou 2009, isto é, amostra 1) para a 4ª série para português. Acredita-se que esse resultado se deve ao fato das escolas que iniciaram em 2008 tiveram mais tempo para se readequar e se adaptar ao programa do que as que iniciaram em 2009. Ou seja, as escolas iniciaram o programa em 2008 teriam adquirido maior aprendizado na implementação do programa do que as que aderiram ao programa em 2009, o que sugere que o tempo de experiência no programa é relevante para os resultados.

Cabe destacar algumas limitações desta dissertação. Primeiramente, o fato da unidade de análise ser a escola e não o aluno constitui-se como um grande limitador do estudo. Mesmo que a SEDUC tenha declarado a adesão em massa pelos alunos no Rio Grande do Sul, seria necessário verificar em cada escola, os alunos efetivos no programa. Além disso, detalhes como o tempo de exposição ao programa, presença nas aulas, bem como as atividades individuais por alunos também deveriam ser analisadas. Entretanto para que isso aconteça, é necessário que cada escola tenha um sistema de monitoramento constituído para o acompanhamento dos alunos no programa.

Outras análises de impacto com a utilização de outra metodologias de pareamento também poderiam ter sido realizadas. Cabe aqui mencionar que levou-se em consideração para o pareamento no presente estudo, as escolas que estavam contidas na região de suporte comum e que estavam balanceadas. Contudo, a aplicação do pareamento por *Kernel*, *Caliper* e *Vizinho mais Próximos* também deveriam ser realizados a fins de comparação.

Adicionalmente, o tamanho da amostra para o Rio Grande do Sul é ainda pequeno e, portanto, seria importante aumentar o número de observações analisadas. Uma solução para isso, seria trabalhar com a unidade de análise os alunos e não as escolas, observando por exemplo as notas do Sistema de Avaliação do Rendimento Escolar do Rio Grande do Sul – SAERS, e acompanhando os alunos que realizaram a prova em 2007 (na 2ªsérie/3º ano) e depois em 2010 (na 5ªsérie/ 6ºano).

Além disso, para consolidação dos resultados é necessário a ampliação das amostras para outros estados e para o Brasil. Isto é, novas pesquisas devem ser realizadas para a investigação do comportamento do “Programa Mais Educação” em outras regiões. Sabe-se,

por exemplo, que a implementação do programa contou com um número maior de escolas nas regiões norte e nordeste do país. Portanto, o efeito do programa pode ser diferenciado e/ou intensificado, dependendo da realidade que está sendo investigada.

Destaca-se também que dentro das finalidades do programa, apenas um aspecto foi investigado, isto é, a melhoria do rendimento e do aproveitamento escolar. Entretanto, para uma avaliação completa de impacto do “Programa Mais Educação” é necessário que outros objetivos, que são passíveis de mensuração, sejam investigados, como por exemplo, a redução da evasão, da reprovação, da distorção idade/série e o combate ao trabalho infantil. Ou seja, é necessário que outras finalidades do programa sejam analisadas para a melhor compreensão dos impactos do programa sobre a formação das crianças, bem como, as possíveis interações que existam entre elas e a repercussão dessas interações no impacto de cada finalidade.

Por fim, duas recomendações são apresentadas. Ressalta-se, primeiramente, que as contribuições aqui apresentadas não têm a intenção de esgotar o assunto. Pelo contrário, a avaliação de impacto de efetividade objetiva é apenas umas das avaliações apresentadas pela literatura. Nesse sentido, para uma avaliação adequada do programa, dentro da ideia de ciclo de políticas públicas, torna-se fundamental a realização das demais avaliações descritas na literatura, como complemento e aprofundamento das análises apresentadas no presente estudo.

Ademais, além dos estudos acadêmicos, é fundamental que dentro da estrutura do “Fórum Mais Educação”, instituído pela portaria que cria o programa, e dos “Comitês Metropolitanos, Regionais ou Estaduais” do Mais Educação, se estabeleça uma instância de avaliação/monitoramento do programa nas escolas, para que possa produzir registros sobre a implantação, a execução e os resultados. A integração de informações e experiências de atores e políticas envolvidas no “Programa Mais Educação” são determinantes para a efetividade e sucesso do programa.

5 CONCLUSÕES

Realizou-se a avaliação do impacto do “Programa Mais Educação”, implementado em 2008, sobre o desempenho médio dos alunos de 4^a e 8^a séries das escolas na proficiência em língua portuguesa e matemática da Prova Brasil na rede pública do Estado do Rio Grande do Sul. Um dos objetivos do programa é a implementação de ações pedagógicas e socioeconômicas para melhoria do rendimento e o aproveitamento escolar.

O “Programa Mais Educação” teve início em 2008 em todo o país. Participaram nesse ano 1.409 escolas públicas estaduais e municipais de ensino fundamental localizadas em 54 municípios de 25 estados, além do Distrito Federal. Em 2009, o programa foi ampliado, atingindo 5 mil escolas em todos os estados e no Distrito Federal. No Rio Grande do Sul, 274 escolas participaram do programa no ano de 2009. Diante da proposta de melhoria da qualidade da educação através da educação em tempo integral e sua abrangência territorial, avaliou-se o impacto da participação das escolas do Rio Grande do Sul no “Programa Mais Educação” em 2008 e 2009 nas notas médias de língua portuguesa e matemática na Prova Brasil 2009 (Inep/MEC) para a 4^a série e 8^a série do ensino fundamental.

Utilizou-se para fins de comparação diferentes métodos de estimação: mínimos quadrados ordinários, utilizando a nota em nível das escolas em 2009 como variável dependente e diferenças em diferenças (este com e sem pareamento e com e sem ponderação pelo *propensity score*). Constatou-se a necessidade de um bom pareamento entre as escolas participantes e não participantes, dada as diferenças acentuadas existentes nas características observáveis entre os dois grupos.

Demonstrou-se também que o “Programa Mais Educação” tem um efeito positivo e significativo de 2,34 pontos nas notas médias de português das escolas participantes do programa na Prova Brasil 2009 na 4^a série do ensino fundamental. Esse impacto é maior para as escolas que iniciaram o programa em 2008, dado o tempo maior para o planejamento e adequação das escolas em relação as exigências e desafios impostos para implementação do programa.

Por outro lado, para a proficiência em matemática (resolução de problemas) da 4^a série, o efeito do programa mostrou-se negativo, ou seja, um impacto de -3,04 pontos. Uma das

razões para isso pode ser o fato da pouca ênfase dada ao programa ao desenvolvimento cognitivo dos alunos no ensino da matemática na resolução de problemas (ênfase dada na Prova Brasil). Além disso, não foi encontrado efeito algum sobre o desempenho escolar dos alunos da 8ª série tanto para português quanto para matemática, o que supõe que ou o programa não tenha funcionado adequadamente para os alunos das séries finais do ensino fundamental das escolas do Rio Grande do Sul ou alunos dessas séries não foram contemplados com o programa.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABADIE, Alberto. Emiparametric Difference-in-Differences Estimators. **Review of Economic Studies**, v. 72, n .1, p, 1-19, 2005.

ALBERNAZ, Ângela; FERREIRA, Francisco; FRANCO, Creso. Qualidade e equidade no ensino fundamental brasileiro. **Pesquisa e Planejamento Econômico**. PPE/ IPEA, v.32, n.3, dez de 2002.

AMARAL, Luiz Felipe; MENEZES-FILHO, Naercio. A Relação entre Gastos Educacionais e Desempenho Escolar. In: Encontro Nacional de Economia – ANPEC 2008, XXXVI, 2008, Salvador. **Anais...** Salvador: ANPEC, 2008.

ANGRIST, Joshua; BETTINGER, Eric; KREMER, Michael. Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia. **American Economic Review**, 2005.

_____. HAHN, Jinyong. When to Control for Covariates? Panel Asymptotics for Estimates of Treatment Effects. **Review of Economics and Statistics**. Vol. 86, n. 1, p.58–72, 2004.

_____. IMBENS, Guido; RUBIN, David. Identification of Causal Effects Using Instrumental Variables. **Journal of the American Statistical Association**, 91, p. 444-472, 1996.

_____. PISCHKE, J-S. **Mostly Harmless Econometrics: an Empiricist`s Companion**. New Jersey: Princeton University Press, 2009.

AQUINO, Juliana Maria de; KASSOU, Ana Lúcia. A ampliação da jornada escolar melhora o desempenho acadêmico dos estudantes? Uma avaliação do programa escola de tempo integral da rede pública do estado de São Paulo. In: Encontro Nacional de Economia – ANPEC 2011, XXXIX, 2011, Foz do Iguaçu. **Anais...** Foz do Iguaçu: ANPEC, 2011.

ARAÚJO, Aloísio et al. A educação infantil e sua importância na redução da violência, In: VELOSO, Fernando, et al (Org.). **Educação Básica no Brasil: construindo o país do futuro**, Rio de Janeiro: Elsevier, 2009.

ARAÚJO, Fernando; LOUREIRO, Maria Rita. Por uma Metodologia Pluridimensional de Avaliação da LRF. **Revista de Administração Pública**. Rio de Janeiro, 39 (6), p.123-152, nov./dez, 2005.

ARRETCHE, Marta T. Tendências no Estudo sobre Avaliação, In: **Avaliação de Políticas Sociais: uma questão em debate**, Elizabeth Melo Rico (org.), 5º edição, São Paulo: Cortez editora - Instituto de Estudos Especiais, 2007.

ASHENFELTER, Orley. Estimating the Effect of Training Programs on Earnings. **Review of Economics and Statistics**, 60, p, 47-57, 1978.

_____. CARD, David. Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. **Review of Economics and Statistics**, 67, 648-660, 1985.

ATHEY, Susan; IMBENS, Guido. Identification and Inference in Nonlinear Difference-In-Differences Models. **Econometrica**, vol. 74, n. 2, 2006.

AUSTIN, Peter. A Critical Appraisal of Propensity-Score Matching in the Medical Literature between 1996 and 2003. **Statistics in Medicine**. v. 27, n. 12, p.2037–49, 2008a.

_____. Discussion of ‘A Critical Appraisal of Propensity-Score Matching in the Medical Literature between 1996 and 2003: Rejoinder. **Statistics in Medicine**. v. 27, n. 12, p. 2066–69, 2008b.

BANERJEE, Abhijit et al. Remedying Education: Evidence from Two Randomized Experiments in India. **Quarterly Journal of Economics**, MIT Press, vol. 122, n. 3, p. 1235-1264, 2007.

BARROS, Ricardo Paes de. et al. Determinantes do desempenho educacional no Brasil. **Pesquisa e Planejamento Econômico**. V. 31, n. 1, p. 1-42, abr, 2001.

BECKER, Sacha; ICHINO, Andrea, Estimation of Average Treatment Effects Based on Propensity Scores. **The Stata Journal**, 2(4): p, 358-377, 2002.

BERNARDONI, Doralice; SOUZA, Marta Cristina de; PEIXE, Blênio César.

Fortalecimento da função avaliação de políticas públicas: estudo de caso do processo de avaliação da política de emprego e renda da secretaria de estado do trabalho, emprego e promoção social do Paraná. Escola de Governo do Paraná, 1998. Disponível em:

<http://www.repositorio.seap.pr.gov.br/arquivos/File/gestao_de_politicas_publicas_no_paran_a_coletanea_de_estudos/cap_6_trabalho_emprego_e_promocao_social/capitulo_6_1.pdf>

Acesso em: 17/02/2011.

BERTRAND, Marianne; DUFLO, Esther; MULLAINATHAN, Sendhil. How Much Should We Trust Differences- in-Differences Estimates? **Quarterly Journal of Economics**, Vol 119(1), 249-275, 2004.

_____. MULLAINATHAN, Sendhil. Are Emily and Brandon more Employable than Latoya and Tyrone? Evidence on Racial Discrimination in the Labor Market from a Large Randomized Experiment. **American Economic Review**, v. 94, n. 4, Set, 2004.

BIONDI, Roberta Loboda; FELÍCIO, Fabiana de. **Atributos escolares e o desempenho dos estudantes**: uma análise em painel dos dados do SAEB. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Brasília-DF, 2007

_____. VASCONCELLOS, Lígia; MENEZES-FILHO, Naercio. **Avaliando o impacto da Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP) no desempenho de matemática nas avaliações educacionais**. Seminário Acadêmico FGVSP, São Paulo: 4 nov, 2009.

BRASIL. Lei nº 9.394, de 30 de dezembro de 1996, Estabelece as diretrizes e bases da educação nacional, **Diário Oficial da União**: seção 1, Brasília, DF, ano CXXXI, n. 248, p. 1-9, 23 dez. 1996.

BRASIL. Portaria nº 17 de 24 abril de 2007. Institui o Programa Mais Educação, que visa fomentar a educação integral de crianças, adolescentes e jovens, por meio do apoio a atividades sócio-educativas no contra-turno escolar. **Diário Oficial da União**: seção 1, Brasília, DF, ano 2007, p, 69-70, 24 abr, 2007.

CALIENDO, Marco; KOPEINING, Sabine. **Some practical guidance for the implementation of propensity score matching**. Bonn, Germany: Institute for the Study of Labor (IZA), IZA Discussion Papers, n. 1588, 2005.

CAMERON, Colin; TRIVEDI, Pravin. **Microeconometrics**. Methods and Application, Cambrigde, 1995.

CANO, Ignácio. **Introdução à Avaliação de Programas Sociais**. Editora Fundação Getúlio Vargas. 3º edição, Rio de Janeiro, 2006.

CARD, David; SULLIVAN Daniel. Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment. **Econometrica**, vol, 56, no, 3, 497-530, 1988.

CASTRO, Cláudio. Entrevista. **Zero Hora**, Porto Alegre, 4 mar, 2010, Disponível em <<http://alternativasintepe.blogspot.com/2010/02/especialista-diz-que-ensino-integral.html>>. Acesso em: 17/02/2011.

CAVALIERE, Ana Maria. Educação integral: uma nova identidade para a escola brasileira? **Educação & Sociedade**. Vol. 23, nº 81. P. 47-270. Campinas: dez, 2002.

_____. Tempo de escola e qualidade na educação pública. **Educação & Sociedade**, vol, 28, nº 100 – Especial. p. 1015-1035, Campinas: out, 2007.

CEDEPLAR/ UFMG e FUNDAÇÃO ITAÚ, **Avaliação do Impacto do Programa Escola Integrada de Belo Horizonte**. In: Seminário Itaú Internacional de Avaliação Econômica de Projetos Sociais, V, 2008, São Paulo: Fundação Itaú, 2008.

COELHO, Lígia Martha. **Escola pública de horário integral: um tempo (fundamental) para o ensino fundamental**, In: ABRAMOVICZ, Anete; MOLL, Jaqueline. Para além do fracasso escolar, Campinas, Papirus, p. 191-208, 1997.

COOK, Thomas. Waiting for Life to Arrive: A History of the Regression-Discontinuity Design in Psychology, Statistics, and Economics. **Journal of Econometrics**, Vol. 142(2): p, 636-654, 2008.

CORSEUIL, Carlos e MOURA, Rodrigo. **O SIMPLES Federal e a geração de empregos na Indústria**, Disponível em: <<http://www.ie.ufjf.br/datacenterie/pdfs/seminarios/pesquisa/texto190509.pdf>>2009, Acesso em: 15/05/2010.

COSTA, Leandro O.; ARRAES, Ronaldo A.; VERA-HERNANDEZ, Marcos. **Identificação Parcial do Efeito das Escolas Privadas Brasileiras**. Fórum BNB de Desenvolvimento. XVII. Fortaleza: Banco do Nordeste do Brasil, 2011. Disponível em: <http://www.bnb.gov.br/content/aplicacao/eventos/forumbnb2011/docs/2011_identificacao_parcial.pdf>. Acesso em: 19 fev 2011.

COTTA, Teresa. Metodologias de avaliação de programas e projetos sociais: análise de resultado e impacto. **Revista do Serviço Público**, ano 49, nº 2: 103-123, Abr-jun, ENAP, 1998.

CURI, Andréa Zaitune; MENEZES-FILHO, Naercio Aquino. Determinantes dos gastos com Educação no Brasil. **Pesquisa e Planejamento Econômico**. v. 40, n. 1, abr. 2010.

CURRIE, Janet; Early Childhood Education Programs. **Journal of Economic Perspectives**. V. 15, nº 2, p. 213–238, 2001.

_____. THOMAS, Duncan. Does Head Start Make a Difference? **American Economic Review** v. 85:3, p. 341–64, 1995.

_____. Does Head Start Help Hispanic Children? **Journal of Public Economics**. V. 74:2, p. 235–62, 1999.

D'ATRI, Fabiana. **Municipalização do Ensino Fundamental da Rede Pública**. Os Impactos Sobre o Desempenho Escolar. Dissertação de Mestrado. Escola de Economia de São Paulo. Fundação Getúlio Vargas . São Paulo, 2007.

DE JESUS, Girlene Ribeiro; LAROS Jacob Arie. Eficácia escolar: Regressão Multinível com Dados de Avaliação em Larga Escala. **Avaliação Psicológica**. 3(2), p. 93-106, 2004.

DONALD, Stephen; LANG, Kevin. Inference with Difference in Differences and Other Panel Data. **Review of Economics and Statistics**. Vol. 89, n.2, p. 221-233, 2007.

DUENHAS, Rogério; FRANÇA, Marco Túlio e GONÇALVES, Flávio. Análise de políticas públicas de abertura de escolas de ensino fundamental na redução de diferentes tipos de violência com utilização de mínimos quadrados ponderados por propensity score matching. In: Encontro de Economia da Região Sul – ANPEC SUL 2010, XIII, 2010, Porto Alegre/RS. **Anais...** Porto Alegre, 11 a 13 ago 2010.

DUFLO, Esther. Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. **American Economic Review** v. 91, p. 795-813 , 2001.

DYE, Thomas. Mapeamento dos modelos de análise de políticas públicas, In: HEIDEMANN, Francisco; SALM, José Francisco. **Políticas Públicas e Desenvolvimento**. 1º Edição, Brasília: Editora Universidade de Brasília, 2009.

FELÍCIO, Fabiana de; VASCONCELLOS, Lígia. O efeito da educação infantil sobre o desempenho escolar medido em exames padronizados. In: Encontro Nacional de Economia – ANPEC, XXXV, 2007, Recife/PB. **Anais...** Recife, de 4 a 7 de dezembro de 2007.

FERNANDES, Reynaldo; MENEZES FILHO, Naércio; ZYLBERSTAJN, Hélio. **PLANFLOR: Política ou ilusória?** Informações FIPE, set. de 2000.

FERRÃO, Maria Eugénia; BELTRÃO, Kaizô Iwakami; SANTOS, Denis Paulo. Modelo de Regressão Multinível: Aplicação ao Estudo do Impacto da Política de Não-Repetência no Desempenho Escolar dos Alunos da 4ª Série, In: **Pesquisa e Planejamento Econômico**. Vol.32, n.3, Rio de Janeiro: IPEA, 2002.

_____. O Impacto de políticas de não-repetência sobre o aprendizado dos alunos da 4ª série. **Pesquisa e Planejamento Econômico**. v.32, n.3. Rio de Janeiro: IPEA, dez 2002.

FERRO, Andrea. **Avaliação do impacto dos Programas de Bolsa Escola no Trabalho Infantil no Brasil**. Dissertação (Mestrado em Economia) - Escola Superior de Agricultura. Universidade de São Paulo, São Paulo, 2003.

_____. **Jovens e adultos no mercado do trabalho: impacto dos programas condicional de renda e uma análise agregada usando series temporais**, Tese (Doutorado em Economia) - Escola Superior de Agricultura. Universidade de São Paulo, São Paulo, 2007.

FIGUEIREDO, Cláudia Maria G.; HADDAD, Mônica A.; FREGUGLIA, Ricardo. **Quality of Education and Public Resources Allocation in Brazil**. TD. 007/2011. Programa de Pós-Graduação em Economia Aplicada - FE/UFJF. Juiz de Fora: UFJF, 2011. Disponível em: <<http://www.ufjf.br/poseconomia/files/2011/08/TD-007-2011-CI%C3%A1udia-Gomes-et al.pdf>> Acesso em: 19 Fev 2012.

FIGUEIREDO, Marcos; FIGUEIREDO, Angelina. Avaliação política e Avaliação de políticas: um quadro de referência teórica. **Análise e Conjuntura**, n° 3 set/dez, p. 107-127. Belo Horizonte, 1986.

FISHER, Ronald. **The Design of Experiments**. 1° edição, Oliver and Boyd, London, 1925.

FRAKER, Thomas; MAYNARD, Rebecca, The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs. **Journal of Human Resources**, Vol, 22, No, 2, p 194–227, 1987

FRANÇA, Marco Túlio Aniceto. **Gestão pública de ensino, política e desenvolvimento**. 2011, Tese (Doutorado em Desenvolvimento Econômico) - Programa Pós Graduação em Desenvolvimento Econômico. Universidade Federal do Paraná, Curitiba, 2011.

FRANÇA. Marco Túlio; DUENHAS, Rogério. A promoção da escola como espaço de convivência comunitária e qualidade de educação: uma análise de diferenças em diferenças para as escolas municipais de Curitiba entre os anos de 2005-2007. In: Encontro de Economia da Região Sul – ANPEC SUL 2011, XIV, 2011, Florianópolis. **Anais...** Florianópolis, 09 a 10 de jun de 2011.

FRANCO, Ana Maria Paiva. Os determinantes do aprendizado com dados de um painel de escolas do SAEB. **Informações FIPE**. Março de 2009.

FRANCO, Creso; MANDARINO, Mônica; ORTIGÃO, Maria Isabel. O projeto pedagógico e os resultados escolares. **Pesquisa e Planejamento Econômico**. v.32, n.3. Rio de Janeiro: IPEA, dez 2002.

FRÖLICH, Markus. Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators. **Review of Economics and Statistics**, 86(1): 77–90, 2004a.

_____. A Note on the Role of the Propensity Score for Estimating Average Treatment Effects. **Econometric Reviews**, 23(2): 167–74, 2004b.

FUERST, J.S ; FUERST, Dorothy. Chicago Experience with an Early Childhood Program: The Special Case of the Child Parent Center Program. **Urban Education**. 28, p. 69–96, 1993.

GARCES, Eliana; THOMAS, Duncan; CURRIE, Janet. Longer-Term Effects of Head Start. **NBER Working Paper n . 8054**, Cambridge MA, Dez de 2000.

GERMANI, Bernadete. **Educação de tempo integral: passado e presente na rede municipal de ensino de Curitiba**, 2006. Dissertação (Mestrado em Educação) - Programa de Pós-Graduação em Educação. Pontifícia Universidade Católica do Paraná, Curitiba, 2006.

GOMES, Maria do Carmo. **Escola de tempo integral: redimensionar o tempo ou a educação?**, 2009. Dissertação (Mestrado em Educação) - Programa de Pós-Graduação em Educação, Pontifícia Universidade Católica de Campinas, Campinas, 2009.

GOMES, Sônia Maria; MAGALHÃES, André. **Qual a relação entre a merenda escolar, carência e o distúrbio nutricionais e a defasagem idade-série do aluno de 1ª a 8ª séries de escolas públicas brasileiras?** Fórum BNB de Desenvolvimento. XVII. Fortaleza: Banco do Nordeste do Brasil, 2011. Disponível em:

<http://www.bnb.gov.br/content/aplicacao/eventos/forumbnb2011/docs/2011_qual_relacao.pdf>. Acesso em: 19 Fev. 2012.

GONÇALVES, Maria Elizete; RIOS-NETO, Eduardo; CÉSAR, Cibele. **Aplicação do modelo hierárquico logístico longitudinal à análise da trajetória escolar (4ª a 8ª série) no ensino fundamental.** Fórum BNB de Desenvolvimento. XVII. Fortaleza: Banco do Nordeste do Brasil, 2011. Disponível em:

<http://www.bnb.gov.br/content/aplicacao/eventos/forumbnb2011/docs/2011_aplicacao_escolar.pdf>. Acesso em: 07/12/2011.

HAHN, Jinyong. On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. **Econometrica**. V. 66, n.2, p. 315–31, 1998.

_____. TODD, Petra; VANDERKLAUW, Wilbert. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. **Econometrica**, v. 69 n. 1, p. 201-209, 2000.

HANSEN, B. The Essential Role of Balance Tests in Propensity-Matched Observational Studies: Comments on ‘A Critical Appraisal of Propensity- Score Matching in the Medical Literature between 1996 and 2003’ by Peter Austin. **Statistics in Medicine**, 27(12): 2050–54, 2008.

HECKMAN, James; ICHIMURA, Hidehiko; TODD, Petra. Matching as an Econometric Evaluation Estimator. **Review of Economic Studies**, v. 65 n. , p. 261–94, 1998.

_____. LALONDE, Robert; SMITH, Jeffrey. The Economics and Econometrics of Active Labor Market Programs, In: **Handbook of Labor Economics**, vol, 3, ed, ASHENFELTER, Orley e CARD, David. p. 1865–2097. Amsterdam: North-Holland, 1999.

_____. ROBB, Richard. Alternative Methods for Evaluating the Impact of Interventions. HECKMAN e SINGER (eds.). **Longitudinal Analysis of Labor Market Data**, Cambridge: Cambridge University Press, 1985.

HEIDEMANN, Francisco. Do sonho do progresso às políticas de desenvolvimento, In _____. SALM, José Francisco. **Políticas Públicas e Desenvolvimento**, Brasília: Editora Universidade de Brasília, 1º Edição, 2009.

HILL, Jennifer. Discussion of Research Using Propensity-Score Matching: Comments on ‘A Critical Appraisal of Propensity-Score Matching in the Medical Literature between 1996 and 2003’ by Peter Austin. **Statistics in Medicine**, 27(12): 2055–61, 2008.

HIRANO, Keisuke; IMBENS, Guido; RIDDER, Geert. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. **Econometrica**, 71(4) p. 1161-1189, Jul, 2003.

IACUS, Stefano; KING, Gary; PORRO, Giuseppe. Multivariate Matching Methods that are Monotonic Imbalance Bounding. **Journal of the American Statistical Association** v. 106, p. 345–361, 2011.

IMBENS, Guido. Nonparametric Estimation of Average Treatment Effects Under Exogeneity: a Review. **The Review of Economics and Statistics**. Vol. 86, n.1, p. 4–29, 2004.

_____. ANGRIST, Joshua. Identification and Estimation of Local Average Treatment Effects. **Econometrica**, Vol, 61, N. 2, 467-476, 1994.

_____. WOOLDRIDGE, Jeffrey. **Recent Developments in the Econometrics of Program Evaluation**, NBER WP 14251, 2008. Disponível em <<http://www.atypon-link.com/AEAP/doi/abs/10.1257/jel.47.1.5>>. Acesso em: 08/02/2011.

_____. **Recent Developments in the Econometrics of Program Evaluation**. *Journal of Economic Literature* 2009, 47:1, 5–86. Disponível em <<http://www.atypon-link.com/AEAP/doi/abs/10.1257/jel.47.1.5>>. Acesso em: 8 fev 2011.

IPEA. **BRASIL: o estado de uma nação** – Estado, crescimento e desenvolvimento: a eficiência do setor público no Brasil. AMORIM, Ricardo L.; CAMPOS, André G.; GARCIA, Ronaldo C. (editores). Brasília: IPEA, 2007.

KHANDKER, Shahidur; KOOLWAL, Gayatri; SAMAD, Hussain. **Handbook on impact evaluation: quantitative methods and practices**. The International Bank for Reconstruction and Development. Washington, D, C: The World Bank, 2010.

KING, Gary et al. **Comparative Effectiveness of Matching Methods for Causal Inference**. Working Paper, 2011. Disponível em: <<http://j.mp/jCpWmk>>. Acesso em: 19/11/2011.

KLEIN, Ruben. Utilização da Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB). **Ensaio: Avaliação e Políticas Públicas em Educação**. Rio de Janeiro, 11, 40, p, 283-96, 2003.

LALONDE, Robert. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. **American Economic Review**. 76, 604-620, 1986.

LAROS, Jacob Arie; MARCIANO, João Luiz. Índices educacionais associados à proficiência em língua portuguesa: um estudo multinível. **Avaliação Psicológica**. 7(3), p, 371-389, 2008.

LEE, David. **The Electoral Advantage of Incumbency and the Voter's Valuation of Political Experience: A Regression Discontinuity Analysis of Close Elections**. Department of Economy, University of California, 2001.

LEE, Valerie et al. Are Head Start Effects Sustained? A Longitudinal Follow-up Comparison of Disadvantaged Children Attending Head Start, No Preschool, and Other Preschool Programs. **Child Development**. 61, p. 495–507, 1990.

LEE, Wang-Sheng. **Propensity Score Matching and Variations on the Balancing Test**. The University of Melbourne, 2006. Disponível em: <https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=esam06&paper_id=217> Acesso em: 08/02/2011.

LOUREIRO, André; SILVA, Victor; OLIVEIRA, Jimmy; COSTA, Leandro. Uma avaliação dos resultados dos programas de combate ao analfabetismo no Brasil. In: Encontro Nacional

de Economia da ANPEC, XXXVI, 2008, 09-12 de dez 2008, Salvador. **Anais...** Salvador: ANPEC, 2008.

MACHADO, Ana Flávia; MORO, Sueli; MARTINS, Ludiemy; RIOS, Juan. Qualidade do Ensino em Matemática: Determinantes do Desempenho de Alunos em Escolas Públicas Estaduais Mineiras. **Revista Economia**. Belo Horizonte: Jan-Abr, 2008.

MANSKI, Charles. Nonparametric Bounds on Treatment Effects. **American Economic Review Papers and Proceedings**, 80, p. 319-323, 1990.

_____. **Partial Identification of Probability Distributions**, New York: Springer-Verlag, 2003.

MENEZES, Tatiane Almeida de; SOARES, Sammara Cavalcanti. **Os determinantes do desempenho escolar: uma análise para o Estado de Pernambuco**. Fórum BNB de Desenvolvimento, XVI. Fortaleza: Banco do Nordeste do Brasil, 2010. Disponível em < <http://www.bnb.gov.br/content/aplicacao/eventos/forumbnb2010/docs/determinantes.pdf>>> Acesso em 10/06/2011.

MENEZES-FILHO, Naercio. **Os Determinantes do Desempenho Escolar do Brasil**. Instituto Futuro Brasil, Ibmec-SP, FEA-USP. 2007. Disponível em << http://veja.abril.com.br/gustavo_ioschpe/arquivos_270908/Menezes-Filho%202007%20-%20Os%20Determinantes%20do%20Desempenho%20Escolar%20no%20Brasil.pdf>>. Acesso em 16/06/2011.

MIGUEL, Edward; KREMER, Michael. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. **Econometrica**, v. 72, n. 1, p. 159-217, 2004.

MINISTÉRIO DA EDUCAÇÃO. **Educação integral**: texto referência para o debate nacional - Série Mais Educação, Brasília. MEC, SECAD, 2009a.

_____. **Programa Mais Educação passo a passo**. Brasília: MEC, SECAD, 2009b. Disponível em < portal.mec.gov.br/dmdocuments/passoapasso_maiseducacao.pdf > Acesso em: 08/02/2011.

_____. **Educação integral/educação integrada e(m) tempo integral**: concepções e práticas na educação brasileira Mapeamento das experiências de jornada escolar ampliada no Brasil. Brasília: MEC, SECAD, 2009c. Disponível em < http://portal.mec.gov.br/index.php?option=com_docman&task=doc_download&gid=2426&Itemid=> Acesso em: 08/02/ 2011.

_____. **Manual de educação integral para obtenção de apoio financeiro através do programa dinheiro direto na escola – PDDE, no exercício de 2009**. Brasília: MEC, SECAD, 2009d.

_____. **Microdados da Prova Brasil 2007**. Brasília: INEP/MEC, 2009e.

_____. **Manual de educação integral para obtenção de apoio financeiro através do programa dinheiro direto na escola – PDDE, no exercício de 2010**. Brasília: MEC, SECAD, 2010.

_____. **Microdados da Prova Brasil 2009**. Brasília: INEP/MEC, 2011.

MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO. **Relatório de Avaliação do Plano Plurianual**. Relatório de Avaliação. Anexo 1. Brasília: MPOG, 2009a. Disponível em:

<http://www.planejamento.gov.br/secretarias/upload/Arquivos/spi/plano_plurianual/avaliacao_PPA/relatorio_2009/09_anexo1_parte1.pdf>. Acesso em 22/11/2011.

_____. **Relatório de Avaliação do Plano Plurianual**. Relatório de Avaliação. Anexo 2. Brasília: MPOG, 2009b. Disponível em:

<http://www.planejamento.gov.br/secretarias/upload/Arquivos/spi/plano_plurianual/avaliacao_PPA/relatorio_2009/09_anexo2.pdf>. Acesso em 22/11/2011.

_____. **Relatório de Avaliação do Plano Plurianual (PPA) 2008 - 2011 - Ano base 2009**: Ministério da Educação. Brasília: MPOG, 2010a.

_____. **Indicadores de Programas** - Guia Metodológico, Brasília: MPOG, SPI, 2010b.

_____. **Relatório de Avaliação do Plano Plurianual**. Relatório de avaliação. Volume I, Parte 1. Brasília: MPOG, 2011a. Disponível em:

<http://www.planejamento.gov.br/secretarias/upload/Arquivos/spi/plano_plurianual/avaliacao_PPA/relatorio_2011/11_volumeI_parte1.pdf>. Acesso em 22/11/2011.

_____. **Relatório de Avaliação do Plano Plurianual**. Relatório de avaliação. Vol. II. Caderno Final. Brasília: MPOG, 2011b. Disponível em:

<http://www.planejamento.gov.br/secretarias/upload/Arquivos/spi/plano_plurianual/avaliacao_PPA/relatorio_2011/11_volumeII_caderno_final.pdf> Acesso em 22/11/2011.

MIRANDA, Carlos; COSTA, Cristina (org). Ações de Combate à Pobreza Rural: metodologia para avaliação de impactos. **Série Desenvolvimento Rural Sustentável**. 1º Ed, Vol. 6,. Brasília: Instituto Interamericano de Cooperação para a Agricultura, 2007.

MOTA, Silvia. **Educação de tempo integral**: da concepção à prática. 2008. Dissertação (Mestrado em Educação) - Programa de Pós-Graduação em Educação, Pontifícia Universidade Católica de Santos, Santos, 2008.

NATIS, Lilian. Modelos hierárquicos lineares. **Estudos em Avaliação Educacional**. n°23, jun-jul de 2001.

NEYMAN, Jerzy. On the Application of Probability Theory to Agricultural Experiments. **Essay on Principles**. Section 9, Translated in Statistical Science, (with discussion), vol 5, No 4, 465–480, 1923.

OLIVEIRA, Jaqueline Maria de. **Custo-efetividade de políticas de redução do tamanho da classe e ampliação da jornada escolar**: uma aplicação de estimadores de matching. 2008. Dissertação (Mestrado em Economia) - Departamento de Economia, Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2008.

_____. **Custo-efetividade de políticas de redução do tamanho da classe e ampliação da jornada escolar**: uma aplicação de estimadores de matching. Rio de Janeiro: BNDES, 2010.

PACHECO, Suzana. Proposta Pedagógica, Elementos para o debate necessário, In: MINISTÉRIO DA EDUCAÇÃO – MEC. Educação integral. **Boletim n° 13**, Brasília, 1998.

PARENTE, Cláudia da Mota; AZEVEDO Érica do Nascimento. Monitoramento do Programa Mais Educação: Educação Integral em Construção. In: Simpósio Brasileiro de Política e Administração da Educação, 25., 2011, São Paulo. **Anais eletrônicos...** São Paulo: Associação Nacional de Política e Administração da Educação – ANPAE, 2011. Disponível em:

<<http://www.anpae.org.br/simposio2011/cdrom2011/PDFs/trabalhosCompleto/comunicacaoRelatos/0105.pdf>> Acesso em: 04/01/2012.

PAZELLO, Elaine Toldo. **Avaliação Econômica de Projetos Sociais**. 2ª edição, Fundação Itaú Social, 2007.

PIRES, Gisele. **Escola pública integrada: as impressões dos professores e especialistas das escolas de Blumenau – SC**. 2007. Dissertação (Mestrado em Educação) - Programa de Mestrado Acadêmico em Educação, Universidade do Vale do Itajaí, Itajaí, 2007.

PORTER, Jack. **Estimation in the Regression Discontinuity Model**. Department of Economics, University of Wisconsin at Madison, 2003.

RESENDE, Anne. **Avaliando resultados de um programa de transferência de renda: o impacto do Bolsa-Escola sobre os gastos das famílias brasileiras**. 2006. Dissertação (Mestrado em Economia) - Centro de Desenvolvimento e Planejamento Regional da Faculdade de Ciências Econômicas, UFMG, Belo Horizonte, 2006.

ROCHA, Vanderson Amadeu da; BELLUZO, Walter. Avaliação do programa de descentralização de gastos públicos no sistema municipal de ensino fundamental de São Paulo. In: Encontro Nacional de Economia – ANPEC 2010, XXXVIII, 2010, Salvador. **Anais...** Salvador: ANPEC, 2010.

RODRIGUES, Adriano; MAGALHÃES, Shirlei Cristina. A resolução de problemas nas aulas de matemática: diagnosticando a prática pedagógica. **Revista eletrônica da Fundação Educacional de Oliveira**. Ano 1, n. 1, Jan-Jul, 2011. Disponível em:

<<http://www.feol.com.br/sites/Revista%20eletronica/artigos/RESOLU%C7%C3O%20DE%20PROBLEMAS%20NAS%20AULAS%20DE%20MATEM%C1TICA%20%5B%20Adriano%20Rodrigues%20e%20Shirlei%20Cristina%20Magalh%C3%93es%205D.pdf>> Acesso em: 04/01/2012.

ROSENBAUM, Paul; RUBIN, Donald. The Central Role of the Propensity Score in Observational Studies for Causal Effects. **Biometrika**, 70, p, 41-55, 1983.

_____. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. **American Statistician**, 39, p, 33-38, 1995.

RUBIN, Donald. Matching to Remove Bias in Observational Studies. **Biometrics**, 29, p. 159-183, 1973a.

_____. The Use of Matched Sampling and Regression Adjustments to Remove Bias in Observational Studies. **Biometrics**, 29, p. 185-203, 1973b.

_____. Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies. **Journal of Educational Psychology**, 66, p. 688-701, 1974.

_____. Assignment to Treatment Group on the Basis of a Covariate. **Journal of Educational Statistics**. 2(1), p. 1-26, 1977.

_____. Bayesian inference for causal effects: The Role of Randomization, **Annals of Statistics**, 6:34–58, 1978.

_____. Estimating Causal Effects from Large Data Sets Using Propensity Scores. **Annals of Internal Medicine**, 127(5 Part 2) p, 757–63, 1997.

_____. THOMAS, Neal. Matching Using Estimated Propensity Scores: Relating Theory to Practice. **Biometrics**. V. 52, n. 1, p. 249–64, 1996.

SCHNEIDER, Anne. Pesquisa avaliativa e melhoria da decisão política: evolução histórica e guia prático, In: HEIDEMANN, Francisco; SALM, José Francisco. **Políticas Públicas e Desenvolvimento**. 1º Edição, Brasília: Editora Universidade de Brasília, 2009.

SECCHI, Leonardo. **Políticas Públicas: conceitos, esquemas de análise, casos práticos**. Editora Cengage Learning. São Paulo, 2010.

SHADISH, Willian; CAMPBELL, Donald; COOK, Thomas. **Experimental and Quasi-experimental Designs for Generalized Causal Inference**. Houghton and Mifflin, Boston, 2002.

SILVA, Samara Paladino Roriz e. **Análise dos efeitos de programas educacionais: o caso projeto jovem de futuro do Instituto Unibanco**. 2010, Dissertação (Mestrado em Economia) - Escola de Economia de São Paulo da Fundação Getúlio Vargas, São Paulo, 2010.

SMITH, Jeffrey; TODD, Petra. Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods. **American Economic Review** v. 91, n.2, p.112–18, 2001.

_____. Does Matching Overcome Lalonde’s Critique of Nonexperimental Estimators? **Journal of Econometrics**. v. 125(1–2), p. 305–53, 2005.

SOARES, José; MAROTTA, Luana. Desigualdade no sistema de ensino fundamental brasileiro. In: VELOSO, Fernando et al (Org.). **Educação Básica no Brasil: construindo o país do futuro**, Rio de Janeiro: Elsevier, 2009.

SOARES, Tufi Machado. Modelo de três níveis hierárquicos para a proficiência dos alunos de 4ª série avaliados no teste de língua portuguesa do SIMAVE/PROEB-2002. **Revista Brasileira de Educação**, nº, 29 p,73-88, Agosto de 2005.

_____. Influência do professor em sala de aula sobre a proficiência alcançada pelos alunos avaliados no SIMAVE -2002. **Estudos em Avaliação Educacional**. n. 28, jul-dez, 2003.

_____. MENDONÇA, Márcia Cristina. Construção de um modelo de regressão hierárquico para os dados do SIMAVE-2000. **Pesquisa Operacional**. vº 23, nº 3, p, 421-441, Set a dez de 2003.

STOCK, James; WATSON, Mark. **Econometria**. 1º edição. Editora Addison Wesley Bra, 1994.

STUART, Elizabeth. Developing Practical Recommendations for the Use of Propensity Scores: Discussion of ‘A Critical Appraisal of Propensity Score Matching in the Medical Literature between 1996 and 2003 by Peter Austin’. **Statistics in Medicine**. V. 27, n. 12, p. 2062–65, 2008.

TAVARES, Héilton Ribeiro. **Teoria da Resposta ao Item para Dados Longitudinais**. 2001. Tese (Doutorado em Estatística) - Universidade de São Paulo, São Paulo, 2001.

TEMPLE, Judy; REYNOLDS, Arthur; MIEDEL, Wendy. Can Early Intervention Prevent High School Dropout? **Urban Affairs**. Março, 35: v.1, p. 31–56, 2000.

TODOS PELA EDUCAÇÃO. **De olho nas metas**. 2010. MOVIMENTO TODOS PELA EDUCAÇÃO. 2010. Disponível em: <http://www.todospelaeducacao.org.br//arquivos/biblioteca/de_olho_nas_metas_2010_final_su_mario.pdf> Acesso em: 11 fev 2011.

TRIBUNAL DE CONTAS DA UNIÃO. **Relatório de Avaliação de Programa Ação Apoio à Formação Continuada de Professores do Ensino Fundamental**. Relator Ministro Valmir Campelo. TCU, Brasília: Secretaria de Fiscalização e Avaliação de Programas de Governo, 2006.

_____. **Relatório de auditoria operacional: Programa Universidade para Todos (ProUni) e Fundo de Financiamento ao Estudante do Ensino Superior (FIES)**. Relator Ministro José Jorge. TCU, 2009.

_____. **Relatório de Acompanhamento do Programa Dinheiro Direto na Escola**. Município de João Pessoa. Primeira Câmara, TC 012.092/2010-0. TCU, 2010.

UNICEF – FUNDO DAS NAÇÕES UNIDAS PARA A INFÂNCIA. Caminhos do direito de aprender. **Boas práticas de 26 municípios que melhoraram a qualidade da educação**, Brasília, 1º ed, Vol. 66, 2010. Disponível em: <http://www.unicef.org/brazil/pt/resources_18427.htm>. Acesso em: 19 fev 2011.

VAITSMAN, Jeni; RODRIGUES, Roberto W. S; PAES-SOUSA, Rômulo. **O Sistema de Avaliação e Monitoramento das Políticas e Programas Sociais: a experiência do Ministério do Desenvolvimento Social e Combate à Fome do Brasil**. Organização das Nações Unidas para a Educação, a Ciência e a Cultura, 2006.

VANDERKLAUW, Wilbert. A Regression–discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrollment. **International Economic Review**, v. 43, n. 4. p, 1249-87, 2002.

VASCONCELLOS, Lígia; BIONDI, Roberta; MENEZES-FILHO, Naercio, **Avaliando o impacto do programa Escrevendo o Futuro no desempenho das escolas públicas na Prova Brasil e na renda futura dos estudantes**. In: Encontro Nacional de Economia – ANPEC 2009, XXXVII, 2009, Salvador. Anais... Salvador: ANPEC, 2009.

VELOSO, Fernando. 15 Anos de Avanços na Educação no Brasil: Onde Estamos? In: _____. et al (Org.). **Educação Básica no Brasil: construindo o país do futuro**. Rio de Janeiro: Editora Elsevier, 2009.

7 ANEXOS

Quadro 7: Revisão da literatura – Modelos hierárquicos

REFERÊNCIA	BASE DE DADOS	TIPO DE ESTIMAÇÃO	ADMIST./REGIÃO	SÉRIE	VARIÁVEL DE RESULTADO	VARIÁVEIS DE CAUSA/ IMPACTO
Natis (2001)	SARESP ¹ 1997	MH de 2 níveis: aluno e escola	Pública estadual / SP	1ª a 4ª série	Proficiência em português	<u>Positivo</u> : diretor atuante; <u>Negativo</u> : defasagem idade-série
Ferrão, Beltrão e Dos Santos (2002 e 2003)	SAEB ² 1999	MH de 2 níveis: aluno e escola	Públicas e privadas / Sudeste	4ª e 8ª séries do EF e 1ª série do EM	Proficiência em matemática, ciências, língua portuguesa, história e geografia	Política de não-repetência (não-significativa); <u>Negativo</u> : raça; defasagem idade-série
Soares e Mendonça (2003)	SIMAVE/PROEB ³ 2000	MH de 2 níveis: turma e escola	Pública estadual / MG	4ª série	Proficiência em matemática e português	<u>Positivo</u> : escore socioeconômico e anos de escolaridade da mãe; frequência do aluno a cultos religiosos; <u>Negativo</u> : defasagem escolar; anos de abandono da escola; raça
Albernaz, Ferreira e Franco (2003)	SAEB 1999	MH de 2 níveis: aluno e escola	Públicas e privadas / Brasil	8ª série	Proficiência em matemática	<u>Positivo</u> : nível socioeconômico da família; nível socioeconômico médio da clientela da escola; escolaridade do professor; qualidade da infraestrutura física; <u>Negativo</u> : raça (negra)
Soares (2003)	SIMAVE/PROEB 2000	MH de 2 níveis: aluno e turma	Pública estadual / MG	4ª série	Proficiência em português	<u>Positivo</u> : escore socioeconômico do aluno e turma; características dos professores; ambiente da sala de aula <u>Negativo</u> : defasagem; sexo (masculino); raça (negra)

Franco, Mandarino e Ortigão (2003)	SAEB 1999	MH de 2 níveis: aluno e escola	Pública e privada / Brasil	8ª série	a)Proficiência em matemática b)Equidade intraescolar socioeconômica	Projetos pedagógicos (não significativo) para a; <u>Negativo</u> : projetos pedagógicos para b
De Jesus e Laros (2004)	SAEB 2001	MH de 3 níveis: aluno, sala de aula e escola	Pública e privada / Brasil	8ª série	Proficiência em português	<u>Positivo</u> : nível socioeconômico médio da escola (NSE médio) e nível de escolaridade médio dos pais; frequência que o aluno faz o dever de casa; <u>Negativo</u> : atraso escolar; aluno que trabalha
Soares (2005)	SIMAVE/PROEB 2002	MH de 3 níveis: aluno, turma e escola	Pública estadual / MG	4ª série	Proficiência em português	<u>Positivo</u> : escore socioeconômico; pré-escola; dedicação do professor; <u>Negativo</u> : defasagem escolar do aluno; raça negra; sexo (masculino)
Machado et al. (2008)	SIMAVE/PROEB; Censo Escolar; SICA e IMRS 2003 ⁴	MH de 3 níveis: aluno, escola e município	Pública estadual /MG	4ª série do EF e 1ª série EM	Proficiência em matemática	<u>Positivo</u> : não reprovação, escolaridade da mãe; mais de 20 livros em casa; <u>Positivo</u> : fazer o dever de casa para EF; <u>Positivo</u> : não fazer tarefas domésticas para EM
Laros e Marciano (2008)	SAEB 2001	MH de 2 níveis: aluno e escola	Pública e privada /Brasil	3ª série EM	Proficiência em português	<u>Positivo</u> : nível socioeconômico da escola; recursos culturais na família do aluno; quantidade de meninas na turma; <u>Negativo</u> : atraso escolar; cobrança e incentivo dos pais do aluno e uso do computador no dever de casa
França (2011)	SAEB 2003 e IBGE ⁵	MH de 3 níveis: estudantes, escolas e regiões	Públicas / Brasil	4ª série	Proficiência em matemática	<u>Positivo</u> : capital econômico, capital humano; rede privada; infraestrutura escolar ; % corpo docente com ES; <u>Negativo</u> : repetência; desigualdade de renda, características dos estudantes (sexo, etnia, nº de pessoas

Gonçalves, Rios-Neto e César (2011)	Avaliação de desempenho: fatores associados – CEDEPLAR /INEP 1999 e 2003; Censo Escolar 1999	MH logístico longitudinal, 3 níveis: trajetória escolar do aluno, variáveis fixas do aluno e escola	Públicas/ estados do NE, NO e CO	4ª e 8ª séries (mesmo aluno)	Repetência	que habitam a residência) <u>Positivo</u> : recursos físicos e humanos das escolas; proficiência média da escola; desempenho dos colegas; proficiência anterior do aluno e nível socioeconômico dos alunos; <u>Negativo</u> : se o aluno trabalha; contexto socioeconômico regional (Nordeste)
-------------------------------------	--	---	----------------------------------	------------------------------	------------	--

Notas: 1, SARESP - Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo; 2, SAEB - Sistema de Avaliação da Educação Básica; 3, SIMAVE/PROEB - Sistema Mineiro de Avaliação da Educação Pública/ Programa de Avaliação da Rede Pública de Educação Básica de Minas Gerais; 4, SICA- Sistema de informação custo aluno de MG; IMRS- Indicadores de responsabilidade social de MG; 5, IBGE - Instituto Brasileiro de Geografia e Estatística;

Quadro 8: Revisão da literatura – Indicadores de qualidade

REFERÊNCIA	BASE DE DADOS	TIPO DE ESTIMAÇÃO	ADMIST./REGIÃO	SÉRIE	VARIÁVEL DE RESULTADO	VARIÁVEIS DE CAUSA/ IMPACTO
Barros et al. (2001)	PNAD ¹ 1996 e PPV ² 1996-1997	MQO	Indivíduos entre 11 e 25 anos/ áreas urbanas do Nordeste e Sudeste	EF I, EF2 e EM	Desempenho educacional (escolaridade)	<u>Positivo</u> : escolaridade dos pais, e em particular da mãe; escolaridade dos professores (para EF2); <u>Negativo</u> : escolaridade dos professores (para EM no NE); salário esperado da população
Felício e Vasconcellos (2007)	SAEB ³ 2003 e Prova Brasil 2005	Pareamento e Painel de Efeitos Fixos	Pública/ Brasil	4ª série	Proficiência em matemática	<u>Positivo</u> : educação infantil
Biondi e Felício (2007)	SAEB 1999 e 2003	MQO e Painel de efeitos fixos	Rede pública/Brasil	4ª série	Proficiência em matemática	<u>Positivo</u> : ausência de rotatividade dos professores ao longo do ano; a experiência média dos professores superior a dois anos em sala de aula e a existência na escola de conexão com a internet; <u>Negativo</u> : proporção de professores que participam de treinamento; escolha do diretor por eleição (ao invés de seleção e eleição); existência de laboratório de informática

Menezes-Filho (2007)	SAEB 2003	Decomposição da Variância e MQO	Pública / 10 estados e DF	4ª e 8ª séries do EF e 1ª série EM	Proficiência em matemática e português	<u>Positivo</u> : número de horas-aula; educação da mãe; número de livros; presença de computador em casa; trabalho fora de casa; idade de entrada no sistema escolar
Amaral e Menezes-Filho (2008)	Prova Brasil 2005, FINBRA ⁴ , IPEA ⁵ , EDUDATABRASIL/INEP	Regressão quantílica e MQO	Rede pública/ Brasil	4ª e 8ª séries	Proficiência em matemática e português	Gastos por aluno no ensino fundamental (sem relevância prática significativa)
Aquino (2008)	SAEB 2005	MQO	Públicas e privadas /Brasil	4ª série	Proficiência em matemática e português	<u>Positivo</u> : presença dos pais nas refeições familiares e reuniões; livros extraclasse nos domicílios; escolaridade da mãe; <u>Negativo</u> : sexo (masculino); reprovação; trabalho infantil
Luz (2008)	Fatores associados ao desempenho escolar - Cedeplar/INEP 1999 a 2003; Censo Escolar 2001	PS e efeito médio do tratamento no tratado (ATT)	Pública estadual/ RM do PA, RO, PE, SE, GO e MS	7ª série e ano seguinte (2002 e 2003)	Proficiência em matemática	<u>Negativo</u> : repetência
Oliveira (2008 e 2010)	SAEB 2005	<i>Generalized</i> PSM;	Rede pública urbana/ Brasil	4ª série	Proficiência em matemática	<u>Positivo</u> : redução no tamanho de classe; ampliação da jornada escolar
Franco (2009)	SAEB 1997, 1999, 2001, 2003 e 2005	Painel de efeitos fixos	Públicas (estaduais e municipais) e privada /	4º série	Proficiência em matemática	<u>Positivo</u> : sexo masculino; morar com o pai e a mãe;

			Brasil			nível de educação da mãe; turno maior que 5 horas (só nas escolas privadas); <u>Negativo</u> : atraso escolar; raça (negra); trabalha; número de matrículas na 4ª série
França e Gonçalves (2010)	SAEB 2003	PSM, Regressão ponderada pelo PS	Públicas (municipal e estadual) e privadas / Brasil	4ª série	Proficiência em matemática	<u>Positivo</u> : aluno de escola privada; índices de capital material e humano; fragmentação na câmara; <u>Negativo</u> : nº de pessoas que moram na casa, uma reprovação, duas reprovações, etnia, morar no norte, nordeste
Menezes e Soares (2010)	SAEB 2007	MQO	Pública estadual/ PB	4ª série	Proficiência em matemática e português	<u>Positivo</u> : número de horas-aula; escolaridade dos professores; tempo que o docente leciona numa mesma escola; aluno tem um acervo mínimo de livros em casa; faz o dever de casa; escolaridade da mãe; <u>Negativo</u> : defasagem idade-série; trabalhar fora de casa; formação heterogênea das turmas; alta rotatividade dos professores

Curi e Menezes-Filho (2010)	PDANs 2001 a 2006 e POF ⁶ 2002-2003 e IBGE	Probit; Modelo Censurado de Heckman e Tobit	Públicas e privadas /Brasil	EI, EF I, EF2 e EM	Probabilidade de o aluno estar mat. na rede privada	<u>Positivo</u> : educação da mãe, a renda familiar, a oferta de escolas públicas, o custo da educação no estado e a região de moradia
Gomes e Magalhães (2011)	ASBRAN ⁷ , IBGE ⁸ e IPEA	Distribuição Binomial Negativa	Pública/ Brasil	1ª a 8ª série	defasagem idade-série do aluno	<u>Negativo</u> : carência nutricional
Figueiredo, Haddad, e Freguglia (2011)	Prova Brasil 2005/2007, Censo Escolar 2004/2006, FINBRA 2004/2006, FIRJAN ⁹ 2005/2007, RAIS ¹⁰ 2004/2006, Ministério do Desenvolvimento 2004/2006, Censo Demográfico/ IBGE 2000	Painel Simples	Rede pública/ Brasil	4ª e 8ª séries	Proficiência em matemática e português	<u>Positivo</u> : gastos públicos em educação (coeficientes muito pequenos em magnitude)
Costa, Arraes e Vera-Hernandez (2011)	SAEB 2005	PSM e MQO	Públicas e privadas / Brasil	4ª série	Proficiência em matemática e português	<u>Positivo</u> : escolas privadas

Notas: 1, PNAD - Pesquisa Nacional por Amostragem de Domicílio; 2, PPV/MS - Pesquisa sobre Padrões de Vida; 3, SAEB - Sistema de Avaliação da Educação Básica; 4, FINBRA - Finanças Municipais do Brasil- STN; 5, IPEA- Instituto de Pesquisa Econômica Aplicada; 6, POF - Pesquisa de Orçamentos Familiares –IBGE; 7, ASBRAN - pesquisa da Associação Brasileira de Nutrição; 8, IBGE - Instituto Brasileiro de Geografia e Estatística; 9, FIRJAN - Federação das Indústrias do Rio de Janeiro; 10, RAIS-Relação Anual de Informações Sociais.

Tabela 12: Estatísticas descritivas e testes de médias da amostra 2 para a 4ª série

	Escolas Controles = 1720		Escolas tratadas = 140		Teste de igualdade de médias	Teste Kolmogorov- Smirnov
	média	desvio-padrão	média	desvio-padrão		
Prova Brasil						
Nota_4a_MAT_2007	199.61	16.49	187.96	13.43	***	***
Nota_4a_PORT_2007	181.73	15.30	170.05	12.61	***	***
Nota_4a_MAT_2009	210.42	19.07	196.24	13.33	***	***
Nota_4a_PORT_2009	189.14	16.47	178.21	12.34	***	***
SEXO	0.49	0.11	0.50	0.09	n	n
RACA	0.59	0.16	0.51	0.12	***	***
MAQ_LAV	0.86	0.13	0.86	0.08	n	***
CAR	0.70	0.27	0.58	0.19	***	***
INTER_CASA	0.24	0.16	0.25	0.12	n	**
DEVER	0.98	0.03	0.97	0.03	n	***
LER	0.97	0.04	0.96	0.04	**	***
NÃO_FALTAR	0.98	0.03	0.97	0.03	***	***
MAE_MEDIO	0.12	0.08	0.11	0.07	n	**
MAE_SUPER	0.11	0.09	0.10	0.07	n	**
MAE_NAORES	0.33	0.12	0.33	0.09	n	*
TRAB	0.12	0.09	0.11	0.06	n	*
REUN	0.92	0.06	0.89	0.06	***	***
PRE_ESC	0.74	0.16	0.53	0.19	***	***
DIR_29	0.03	0.16	0.01	0.12	n	n
DIR_30_a_39	0.21	0.40	0.25	0.43	n	n
DIR_40	0.77	0.42	0.74	0.44	n	n
DIR_POS	0.69	0.46	0.64	0.48	n	n
DIR_11_a_15	0.09	0.29	0.09	0.29	n	n
DIR15	0.05	0.22	0.04	0.19	n	n
PRO_SEL	0.00	0.05	0.00	0.00	n	n
INT_ATI	0.10	0.29	0.16	0.37	**	n
ROT_PROF	0.07	0.25	0.09	0.28	n	n
REF_ESC	0.75	0.43	0.76	0.43	n	n
DIR_SEL_ELE	0.08	0.27	0.09	0.28	n	n
PROG_FED	0.73	0.44	0.88	0.33	***	***
PROG_EST	0.41	0.49	0.31	0.47	**	n
PROG_MUN	0.36	0.48	0.51	0.50	***	***
Censo Escolar 2007						
MÉDIA_AULA	244.17	22.56	245.90	12.73	n	**
MED_MAT_SERIE	24	5	28	5	***	***
TOTAL_MAT	50	26	87	43	***	***
BIBLIO	0.96	0.20	0.92	0.27	**	n
INTER	0.66	0.47	0.56	0.50	**	n
LAB_INFO	0.55	0.50	0.65	0.48	**	n

PROF_SUPER	0.83	0.18	0.77	0.21	***	***
PROF_POS	0.28	0.22	0.25	0.24	n	***
PROF_FUND	78	48	112	96	***	***
Informações do IBGE – 2007						
POP	177.733	336.197	641.885	576.287	***	***
PIB_PER	16,223.97	11,558.05	19,181.58	8,683.12	***	***
RM_OU_POLO	0.57	0.50	1.00	0.00	***	***
Informações do MEC e INEP 2007						
IDEB_mun	4.54	0.46	4.19	0.24	***	***
IDEB_esc	4.58	0.74	3.86	0.57	***	***
Dif_IDEB	0.03	0.61	-0.33	0.47	***	***
PRONASCI	0.29	0.45	1.00	0.00	***	***
TAXA_APRO	89.71	9.57	83.93	9.68	***	***
MUNIC	0.46	0.50	0.67	0.47	***	***

Nota: (***) Nível de significância de 1%; (**) Nível de significância de 5%; (*) Nível de significância de 10%

Tabela 13: Estatísticas descritivas e testes de médias da amostra 2 para a 8ª série

	Escolas Controles = 1720		Escolas tratadas = 93		Teste de igualdade de médias	Teste Kolmogorov- Smirnov
	média	desvio- padrão	média	desvio- padrão		
Prova Brasil						
Nota_4a_MAT_2007	252.24	17.71	241.80	16.48	***	***
Nota_4a_PORT_2007	239.57	15.73	232.36	16.02	***	***
Nota_4a_MAT_2009	258.56	18.47	246.32	15.78	***	***
Nota_4a_PORT_2009	250.58	16.26	241.50	14.87	***	***
SEXO	0.46	0.11	0.46	0.09	n	n
RACA	0.67	0.15	0.56	0.13	***	***
MAQ_LAV	0.84	0.14	0.84	0.08	n	n
CAR	0.75	0.26	0.59	0.19	***	***
INTER_CASA	0.30	0.19	0.31	0.15	n	n
DEVER	0.96	0.04	0.95	0.05	**	*
LER	0.90	0.07	0.89	0.06	n	n
NÃO_FALTAR	0.99	0.02	0.99	0.02	n	n
MAE_MEDIO	0.21	0.12	0.20	0.10	n	n
MAE_SUPER	0.08	0.08	0.05	0.05	***	***
MAE_NAORES	0.14	0.08	0.14	0.07	n	n
TRAB	0.21	0.12	0.15	0.07	***	***
REUN	0.89	0.09	0.81	0.11	***	***
PRE_ESC	0.80	0.16	0.65	0.17	***	***
DIR_29	0.02	0.15	0.01	0.10	n	n
DIR_30_a_39	0.19	0.39	0.25	0.43	n	n
DIR_40	0.79	0.41	0.74	0.44	n	n
DIR_POS	0.71	0.45	0.66	0.48	n	n
DIR_11_a_15	0.10	0.29	0.05	0.23	n	*
DIR15	0.04	0.20	0.04	0.20	n	n
PRO_SEL	0.01	0.07	0.00	0.00	n	***
INT_ATI	0.11	0.31	0.17	0.38	*	n
ROT_PROF	0.09	0.28	0.10	0.30	n	n
REF_ESC	0.73	0.44	0.72	0.45	n	n
DIR_SEL_ELE	0.07	0.26	0.05	0.23	n	n
PROG_FED	0.71	0.45	0.89	0.31	***	***
PROG_EST	0.48	0.50	0.37	0.48	**	**
PROG_MUN	0.29	0.45	0.47	0.50	***	***
Censo Escolar 2007						
MÉDIA_AULA	252.45	18.63	254.59	14.24	***	***
MED_MAT_SERIE	25	6	28	5	***	***
TOTAL_MAT	51	30	72	41	***	***
BIBLIO	0.98	0.13	0.96	0.20	*	n
INTER	0.74	0.44	0.66	0.48	*	*
LAB_INFO	0.63	0.48	0.73	0.45	**	**
PROF_SUPER	0.87	0.15	0.86	0.12	n	n
PROF_POS	0.29	0.21	0.29	0.26	n	n
PROF_FUND	91	44	148	97	***	***
Informações do IBGE – 2007						
POP	169.048	333.763	729.615	598.286	***	***
PIB_PER	16.013,92	10.359,39	19.871,26	8.581,74	***	***
RM_OU_POLO	0.54	0.50	1.00	0.00	***	***
Informações do MEC e INEP 2007						

IDEB_mun	3.84	0.48	3.38	0.20	***	***
IDEB_esc	3.81	0.72	3.24	0.60	***	***
Dif_IDEB	-0.03	0.56	-0.14	0.54	*	*
PRONASCI	0.27	0.45	1.00	0.00	***	***
TAXA_APRO	84.29	12.20	78.91	14.19	***	***
MUNIC	0.37	0.48	0.62	0.49	***	***

Nota: (***) Nível de significância de 1%; (**) Nível de significância de 5%; (*) Nível de significância de 10%

Tabela 14: Estimação do propensity score – amostra 1 - 4ª série

Estimation of the propensity score

note: pronasci != 1 predicts failure perfectly

pronasci dropped and 1227 obs not used

note: sel_alunos != 0 predicts failure perfectly

sel_alunos dropped and 3 obs not used

note: IDEB_municipio_em_2007 dropped because of collinearity

note: RM_ou_polo dropped because of collinearity

Iteration 0: log likelihood = -401.55292

Iteration 1: log likelihood = -259.75755

Iteration 2: log likelihood = -239.92854

Iteration 3: log likelihood = -237.08873

Iteration 4: log likelihood = -236.99165

Iteration 5: log likelihood = -236.9915

Logistic regression

Number of obs = 679

LR chi2(42) = 329.12

Prob > chi2 = 0.0000

Pseudo R2 = 0.4098

Log likelihood = -236.9915

Tratamento	Coef.	Std. Err.	z	P> z	[95% Conf.Interval]	
Nota_4a_PORT_2007	-0.06429	0.026325	-2.44	0.015	-0.11589	-0.0127
Nota_4a_MAT_2007	0.119824	0.026749	4.48	0	0.067398	0.172251
IDEB_esc	-4.24006	0.773048	-5.48	0	-5.75521	-2.72492
TAXA_APRO	0.004314	0.018402	0.23	0.815	-0.03175	0.040382
Dif_IDEB	0.876968	0.624345	1.4	0.16	-0.34673	2.100661
PIB_PER	4.12E-05	2.26E-05	1.82	0.069	-3.13E-06	8.56E-05
POP	1.69E-07	3.75E-07	0.45	0.652	-5.66E-07	9.04E-07
MUNIC	2.750181	0.590345	4.66	0	1.593126	3.907235
SEXO	-1.43555	1.479201	-0.97	0.332	-4.33473	1.463631
RACA	1.591333	1.040031	1.53	0.126	-0.44709	3.629757
MAQ_LAV	3.59024	1.695909	2.12	0.034	0.266319	6.914161
CAR	-0.6827	0.860469	-0.79	0.428	-2.36919	1.003791
INTER_CASA	0.811423	1.524934	0.53	0.595	-2.17739	3.800238
DEVER	6.866438	4.474066	1.53	0.125	-1.90257	15.63545
LER	0.861981	3.60085	0.24	0.811	-6.19555	7.919517
NÃO_FALTAR	-5.89581	4.452135	-1.32	0.185	-14.6218	2.830214
MAE_MEDIO	-1.08963	2.14182	-0.51	0.611	-5.28752	3.108263
MAE_SUPER	-0.24422	2.2022	-0.11	0.912	-4.56045	4.072015
MAE_NAORES	0.683407	1.358765	0.5	0.615	-1.97972	3.346537
TRAB	-0.69296	2.146281	-0.32	0.747	-4.8996	3.513673
PRE_ESC	-2.24273	0.767714	-2.92	0.003	-3.74742	-0.73803
REUN	-3.27297	2.14032	-1.53	0.126	-7.46792	0.921978
MÉDIA_AULA	0.003067	0.004701	0.65	0.514	-0.00615	0.01228
TOTAL_MAT	-0.00072	0.004623	-0.16	0.876	-0.00978	0.008342

MED_MAT_SERIE		0.023306	0.030881	0.75	0.45	-0.03722	0.083832
INTER		0.293183	0.28071	1.04	0.296	-0.257	0.843364
BIBLIO		0.124014	0.456649	0.27	0.786	-0.771	1.01903
LAB_INFO		0.852199	0.282711	3.01	0.003	0.298096	1.406302
DIR_29		-1.18806	1.056481	-1.12	0.261	-3.25872	0.882609
DIR_40		0.036275	0.302623	0.12	0.905	-0.55686	0.629405
DIR_POS		0.006402	0.275992	0.02	0.981	-0.53453	0.547336
DIR_11_a_15		-0.23372	0.406063	-0.58	0.565	-1.02959	0.562149
DIR15		-0.01677	0.543981	-0.03	0.975	-1.08296	1.049412
DIR_SEL_ELE		0.25185	0.419349	0.6	0.548	-0.57006	1.073758
PROF_POS		-1.85514	0.76474	-2.43	0.015	-3.35401	-0.35628
PROF_FUND		0.003469	0.002531	1.37	0.17	-0.00149	0.008429
INT_ATI		0.147774	0.380194	0.39	0.698	-0.59739	0.892941
ROT_PROF		0.663874	0.434742	1.53	0.127	-0.1882	1.515953
REF_ESC		-0.35137	0.312792	-1.12	0.261	-0.96443	0.261692
PROG_FED		-0.29058	0.412945	-0.7	0.482	-1.09993	0.518779
PROG_EST		0.816073	0.412264	1.98	0.048	0.00805	1.624095
PROG_MUN		-0.01645	0.454222	-0.04	0.971	-0.90671	0.873805
<u>_cons</u>		0.056418	6.514138	0.01	0.993	-12.7111	12.82389

Tabela 15: Estimação do propensity score – amostra 1 - 8ª série

Estimation of the propensity score

note: pronasci != 1 predicts failure perfectly

pronasci dropped and 1013 obs not used

note: sel_alunos != 0 predicts failure perfectly

sel_alunos dropped and 3 obs not used

note: IDEB_municipio_em_2007 dropped because of collinearity

note: RM_ou_polo dropped because of collinearity

note: idade3039 dropped because of collinearity

Iteration 0: log likelihood = -298.62748

Iteration 1: log likelihood = -190.56156

Iteration 2: log likelihood = -171.6208

Iteration 3: log likelihood = -167.32665

Iteration 4: log likelihood = -166.95284

Iteration 5: log likelihood = -166.9489

Iteration 6: log likelihood = -166.9489

Logistic regression

Number of obs = 516

LR chi2(43) = 263.36

Prob > chi2 = 0.0000

Pseudo R2 = 0.4409

Log likelihood = -166.9489

Tratamento	Coef.	Std.Err.	z	P> z	[95% Conf.Interval]	
Nota_4a_PORT_2007	-0.01065	0.019054	-0.56	0.576	-0.04799	0.026699
Nota_4a_MAT_2007	0.020151	0.019121	1.05	0.292	-0.01733	0.057628
IDEB_esc	-6.06071	0.943157	-6.43	0	-7.90926	-4.21215
TAXA_APRO	0.001811	0.016579	0.11	0.913	-0.03068	0.034306
Dif_IDEB	4.96472	0.87007	5.71	0	3.259414	6.670027
PIB_PER	3.68E-05	2.35E-05	1.57	0.117	-9.23E-06	8.29E-05
POP	7.19E-07	4.32E-07	1.66	0.096	-1.28E-07	1.56E-06
MUNIC	3.536511	0.752375	4.7	0	2.061884	5.011138
SEXO	-0.1996	1.463712	-0.14	0.892	-3.06842	2.669228
RACA	0.690508	1.396825	0.49	0.621	-2.04722	3.428234
MAQ_LAV	3.552769	1.914024	1.86	0.063	-0.19865	7.304188
CAR	-0.65237	1.09053	-0.6	0.55	-2.78977	1.485031
INTER_CASA	-2.00179	1.527562	-1.31	0.19	-4.99576	0.992173
DEVER	-3.03015	4.070224	-0.74	0.457	-11.0076	4.947342
LER	-3.87635	2.518617	-1.54	0.124	-8.81275	1.060049
NÃO_FALTAR	-5.00363	8.851693	-0.57	0.572	-22.3526	12.34537
MAE_MEDIO	0.606868	1.912948	0.32	0.751	-3.14244	4.356177
MAE_SUPER	-0.46125	3.202088	-0.14	0.885	-6.73723	5.814724
MAE_NAORES	-2.9901	1.938869	-1.54	0.123	-6.79021	0.810012
TRAB	3.932602	2.036277	1.93	0.053	-0.05843	7.923631

PRE_ESC		-2.58424	1.064409	-2.43	0.015	-4.67044	-0.49803
REUN		-1.72224	1.645269	-1.05	0.295	-4.9469	1.502432
MÉDIA_AULA		-0.00446	0.008749	-0.51	0.61	-0.02161	0.012683
TOTAL_MAT		-0.02306	0.007404	-3.11	0.002	-0.03757	-0.00855
MED_MAT_SERIE		0.010177	0.030773	0.33	0.741	-0.05014	0.070491
INTER		0.427326	0.353274	1.21	0.226	-0.26508	1.11973
BIBLIO		-0.16693	0.755871	-0.22	0.825	-1.64841	1.314551
LAB_INFO		0.658192	0.337242	1.95	0.051	-0.00279	1.319174
DIR_29		-1.56895	1.355284	-1.16	0.247	-4.22525	1.087361
DIR_40		0.119388	0.368269	0.32	0.746	-0.60241	0.841182
DIR_POS		-0.39738	0.324837	-1.22	0.221	-1.03405	0.239284
DIR_11_a_15		-1.43899	0.612404	-2.35	0.019	-2.63928	-0.2387
DIR15		0.406486	0.670902	0.61	0.545	-0.90846	1.721431
DIR_SEL_ELE		-0.15661	0.669717	-0.23	0.815	-1.46923	1.156009
PROF_POS		-1.66632	0.955364	-1.74	0.081	-3.5388	0.206156
PROF_SUPER		-2.06361	1.54575	-1.34	0.182	-5.09323	0.966001
PROF_FUND		0.012925	0.004225	3.06	0.002	0.004644	0.021207
INT_ATI		0.143788	0.449722	0.32	0.749	-0.73765	1.025227
ROT_PROF		0.897213	0.502836	1.78	0.074	-0.08833	1.882752
REF_ESC		-0.37101	0.340535	-1.09	0.276	-1.03845	0.296425
PROG_FED		0.32217	0.469733	0.69	0.493	-0.59849	1.242829
PROG_EST		0.9195	0.483828	1.9	0.057	-0.02879	1.867786
PROG_MUN		-0.99433	0.533098	-1.87	0.062	-2.03918	0.050523
_cons		29.39861	9.576054	3.07	0.002	10.62989	48.16733

Tabela 16: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em língua portuguesa para a 4ª série (Amostra 2)

Amostra 2 - 4ª série	Impacto					Controles				
	Coef.	EP Rob.	t	P-valor	R2	Amostra/total	RM	Mun.	Esc.	carac. socioecon.
Nota PORT 2009: em nível										
(I) OLS - nível	-10.94***	1.11	-9.83	0.000	3.08%	1860	n	n	n	n
(II) OLS - nível	0.28	0.60	0.46	0.644	82.24%	1860	s	s	s	s
Diferenças em Diferenças										
(III) sem pareamento	2.58***	0.68	3.81	0.000	83.10%	3720	s	s	s	s
(IV) com pareamento	1.99**	0.81	2.45	0.014	84.41%	866	s	s	s	s
(V) com pareamento (ponderado pelo pscore)	2.09*	1.07	1.96	0.050	83.78%	866	s	s	s	s

*estimativas significativas a 10%.

**estimativas significativas a 5%.

*** estimativas significativas a 1%.

Tabela 17: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em matemática para a 4ª série (Amostra 2)

Amostra 2 - 4ª série	Impacto					Controles				
	Coef	EP Rob.	T	P-valor	R2	Amostra / total	RM	Mun	Esc	Carac. Socio.
Nota MAT 2009: em nível										
(I) OLS – nível	-14.17***	1.21	-11.68	0.000	3.85%	1860	n	n	n	n
(V) OLS – nível	-1.30*	0.71	-1.83	0.067	80.65%	1860	s	s	s	s
Diferenças em Diferenças										
(III) sem pareamento	-3.65***	0.76	-4.80	0.000	82.82%	3720	s	s	s	s
(III) com pareamento	-2.20**	0.89	-2.48	0.013	83.48%	866	s	s	s	s
(III) com pareamento (ponderado pelo pscore)	-2.78**	1.07	-2.60	0.010	83.26%	866	s	s	s	s

*estimativas significativas a 10%.

**estimativas significativas a 5%.

*** estimativas significativas a 1%.

Tabela 18: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em língua portuguesa para a 8ª série (Amostra 2)

Amostra 2 - 8ª série	Impacto					Controles				
	Coef	EP Rob.	t	P-valor	R2	Amostra / total	RM	Mun	Esc	Carac. Socio.
Nota PORT 2009: em nível										
(I) OLS – nível	-9.08***	1.60	-5.69	0.000	1.82%	1488	n	n	n	n
(V) OLS – nível	-1.43	1.04	-1.38	0.167	74.69%	1488	s	s	s	s
Diferenças em Diferenças										
(III) sem pareamento	-0.74	1.22	-0.60	0.548	76.82%	2976	s	s	s	s
(III) com pareamento	-0.33	1.40	-0.23	0.816	76.88%	646	s	s	s	s
(III) com pareamento (ponderado pelo pscore)	0.28	1.58	0.18	0.859	85.90%	646	s	s	s	s

*estimativas significativas a 10%.

**estimativas significativas a 5%.

*** estimativas significativas a 1%.

Tabela 19: Resultados das estimativas de impacto do “Programa Mais Educação” no desempenho em matemática para a 8ª série (Amostra 2)

Amostra 2 - 8ª série	Impacto					Controles				
	Coef	EP Rob.	t	P-valor	R2	Amostra / total	RM	Mun	Esc	Carac. Socio.
Nota MAT 2009: em nível										
(I) OLS – nível	-12.25***	1.70	-7.19	0.000	2.55%	1488	n	n	n	n
(V) OLS – nível	0.64	1.04	0.61	0.540	78.03%	1488	s	s	s	s
Diferenças em Diferenças										
(III) sem pareamento	-0.07	1.22	-0.06	0.953	77.85%	2976	s	s	s	s
(III) com pareamento	0.35	1.40	0.25	0.805	76.57%	646	s	s	s	s
(III) com pareamento (ponderado pelo pscore)	-1.02	1.49	-0.68	0.495	83.25%	646	s	s	s	s

*estimativas significativas a 10%.

**estimativas significativas a 5%.

*** estimativas significativas a 1%.