UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE
RIBEIRÃO PRETO
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA – ÁREA: ECONOMIA
APLICADA

CLÁUDIA HIROMI OSHIRO

Efeito do pagamento de bônus aos professores sobre a proficiência escolar

ORIENTADOR: PROF. DR. LUIZ GUILHERME DÁCAR DA SILVA
SCORZAFAVE

RIBEIRÃO PRETO

2012

Prof. Dr. João Grandino Rodas
Reitor da Universidade de São Paulo

Prof. Dr. Sigismundo Bialoskorski Neto
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Walter Belluzzo Jr
Chefe do Departamento de Economia

CLÁUDIA HIROMI OSHIRO

Efeito do pagamento de bônus aos professores sobre a proficiência escolar

Dissertação apresentada ao Programa de Pós-Graduação em Economia – Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências. Versão Corrigida. A original encontra-se disponível no Serviço de Pós-Graduação da FEA-RP/USP.

ORIENTADOR: PROF. DR. LUIZ GUILHERME DÁCAR DA SILVA
SCORZAFAVE

RIBEIRÃO PRETO

2012

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

FICHA CATALOGRÁFICA

Oshiro, Cláudia Hiromi

Efeito do pagamento de bônus aos professores sobre a proficiência escolar. Ribeirão Preto, 2011.

67p. : il; 30 cm

Dissertação de Mestrado, apresentada à Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto/USP.
Área de concentração: Economia social

Orientador: Scorzafave, Luiz Guilherme Dácar da Silva

1. Educação 2. Bônus 3. *Accountability*

Folha de aprovação

Cláudia Hiromi Oshiro

Efeito do pagamento de bônus aos professores sobre a proficiência escolar

Dissertação apresentada ao Programa de Pós-Graduação em Economia – Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências. Versão Corrigida. A original encontra-se disponível no Serviço de Pós-Graduação da FEA-RP/USP.

Aprovada em: _____

BANCA EXAMINADORA

Prof. Dr. Luiz Guilherme Dácar da Silva Scorzafave

Instituição: FEARP/USP

Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

*Dedico essa dissertação
a minha família*

AGRADECIMENTOS

Agradeço o apoio e compreensão de toda a minha família, especialmente do Rafael.

Agradeço a todos meus amigos, especialmente minhas queridas Shirley, Juliana, Regina e Maria Cláudia por estarem sempre ao meu lado apesar dos meus muitos momentos de ausência.

Agradeço as minhas amigas de república em Ribeirão Preto, Karen, Daniela e Letícia por me receberem com tanto carinho e ao meu amigo Jorge pela agradável companhia na cidade.

Um agradecimento mais que especial aos meus colegas de mestrado pelo companheirismo e agradável convívio que ajudaram a superar os momentos difíceis durante essa etapa, especialmente aos meus amigos Anderson, Ednilson, João Paulo, Vanderson, Barba e Blanca.

Agradeço a todos os professores da FEARP-USP, mas especialmente aos professores Reynaldo Fernandes e Alexandre Nicolella pelas sugestões no exame de qualificação.

Agradeço a Sandra e Murilo do Departamento de Economia e a Vânia e Érika, da Secretaria de Pós-graduação, que sempre me atenderam com toda atenção e competência.

Agradeço aos colegas da *Tendências* Consultoria que sempre me incentivaram a seguir no mestrado e me deram apoio para finalizar essa dissertação. Um agradecimento especial a Clara, Camila, Minduba, Mariana e Rafael.

Por fim, e mais importante, agradeço ao meu orientador, Luiz Guilherme, que me apoiou do início ao fim da dissertação, sempre atencioso e paciente no esclarecimento das minhas dúvidas e compreensivo nas situações adversas.

RESUMO

OSHIRO, C.H. **Efeito do pagamento de bônus aos professores sobre a proficiência escolar**, 2012. 67f. Dissertação de Mestrado – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, Ribeirão Preto, 2012.

Um dos grandes desafios brasileiros é a melhora do sistema educacional. Embora um avanço considerável tenha ocorrido no que diz respeito ao acesso a escola nas últimas duas décadas, a qualidade do ensino ainda se encontra num nível muito aquém do desejado. Uma das alternativas de política para se alterar esse quadro consiste na responsabilização dos educadores pelo desempenho dos alunos, seja por meio apenas de divulgação pública dos resultados dos estudantes ou adicionando prêmios ou sanções como incentivo ao esforço. A maioria dos trabalhos que avalia programas de responsabilização em diversos países indica efeito positivo sobre o desempenho escolar, mas esse resultado não é um consenso, já que alguns trabalhos não encontraram impacto desse tipo de política. Para o caso brasileiro, ainda existem poucos estudos empíricos que procuram avaliar as políticas de responsabilização. Diante disso, o objetivo deste trabalho é fazer uma avaliação do efeito do programa de bonificação de professores do estado de São Paulo sobre o desempenho escolar dos alunos. Os métodos de *Propensity Score Matching* e Diferenças em Diferenças foram utilizados para medir o impacto do programa sobre as notas médias das escolas. Foram encontrados efeitos positivos e significativos para Matemática (6 pontos na escala SAEB) e Português (3 pontos na mesma escala) para a quarta série do ensino fundamental. Não foram encontrados efeitos para a oitava série. Os resultados se mostraram robustos a especificação de diferentes grupos de controle.

Palavras-chave: Educação, Bônus, *Accountability*

ABSTRACT

OSHIRO, C.H. Effect of teacher bonus payment scheme on school proficiency, 2012.67f. Dissertação de Mestrado – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, Ribeirão Preto, 2012.

One of the biggest brazilian's challenges is the improvement of educational system. Although considerable progress has occurred with regard to access to school in the last two decades, the quality of education is still very low level compared to other countries. One of the alternatives of policy to change this situation is the responsibility of educators for the student performances whether through public disclosure of the results or giving bonuses or penalties as an incentive to effort. Many studies of accountability programs in several countries indicated a positive effect on student grades, but it is not a general agreement, since some studies found no impact of such policy. In the brazilian case, there are few empirical studies measuring accountabilities programs. Thus, the objective of this study is to evaluate the effect of the bonus program for teachers in the state of Sao Paulo over proficiency of students. Methods of Propensity Score Matching and Difference in Differences were used to measure the program's impact on the average scores of schools. We found positive and significant effects on tests of Math (6 points in Saeb scale) and Lecture (3 points in the same scale) for the fourth grade. We didn't find any effects for eighth grade of elementary school. These results are robust to different control groups.

Key-words: Education, Bonus, Accountability

LISTA DE FIGURAS

Figura 1- Evolução da proficiência de Língua Portuguesa e Matemática da quarta série - (1995 a 2009).....	32
Figura 2- Evolução da proficiência de Língua Portuguesa e Matemática da oitava série - (1995 a 2009).....	33
Figura 3- Histograma depois do matching – quarta série - Português.....	37
Figura 4- Histograma depois do matching – quarta série - Matemática.....	38
Figura 5- Histograma depois do matching – oitava série - Português.....	38
Figura 6- Histograma depois do matching – oitava série - Matemática.....	39
Figura 7- Notas de Português e Matemática da Quarta série - Estadual Paulista x Municipal Paulista (2005-2009).....	47
Figura 8- Notas de Português e Matemática da Oitava série - Estadual Paulista x Municipal Paulista (2005-2009).....	47
Figura 9- Notas de Português e Matemática da Quarta série - Estadual Paulista x Restante do Brasil.....	49
Figura 10- Notas de Português e Matemática da Oitava série - Estadual Paulista x Restante do Brasil.....	49
Figura 11- Notas de Português e Matemática da Quarta série - Estadual Paulista x Restante do Sudeste.....	52
Figura 12- Notas de Português e Matemática da Oitava série - Estadual Paulista x Restante do Sudeste.....	52
Figura 13- Notas de Português e Matemática da Quarta série - Estadual Paulista x RJ e ES.....	54
Figura 14- Notas de Português e Matemática da Oitava série - Estadual Paulista x RJ e ES.....	55
Figura 1A: Notas de Português da Quarta série - Estadual Paulista x Estados do Sudeste.....	67
Figura 2A: Notas de Português da Quarta série - Estadual Paulista x Estaduais dos estados do Sudeste.....	67

LISTA DE TABELAS

Tabela 1- Valores de referência para níveis de proficiência	21
Tabela 2- Descrição das variáveis, médias para Brasil e Estadual São Paulo – 2007	31
Tabela 3- Coeficientes estimados do Probit - quarta série - 2007	35
Tabela 4- Coeficientes estimados do Probit - Oitava série - 2007	36
Tabela 5- Suporte comum	37
Tabela 6- Diferença das médias das variáveis explicativas – Quarta série – Português	40
Tabela 7- Diferença das médias das variáveis explicativas – Quarta série- Matemática	41
Tabela 8- Diferença das médias das variáveis explicativas – Oitava série – Português	42
Tabela 9- Diferença das médias das variáveis explicativas – Oitava série – Matemática	42
Tabela 10- Pseudo R-quadrado	43
Tabela 11- Matching por vizinho mais próximo (quarta e oitava série)	43
Tabela 12- Resultado do Diferenças em Diferenças	44
Tabela 13- Resultado do Diferenças em Diferenças por Município	46
Tabela 14- Diferenças em Diferenças: Estaduais paulistas x Municipais paulistas	48
Tabela 15- Diferenças em Diferenças: Estaduais paulistas x Resto do Brasil	50
Tabela 16- Diferenças em Diferenças: Estaduais paulistas x Resto do Brasil (Estaduais)	50
Tabela 17- Resultado do Diferenças em Diferenças: Estaduais paulistas x Escolas do restante do Sudeste	53
Tabela 18- Resultado do Diferenças em Diferenças: Estaduais paulistas x Escolas estaduais do restante do Sudeste.....	53
Tabela 19- Resultado do Diferenças em Diferenças: Estaduais paulistas x Rio de Janeiro e Espírito Santo.....	55
Tabela 20- Resultado do Diferenças em Diferenças: Estaduais paulistas x Estaduais Rio de Janeiro e Espírito Santo	56
Tabela 21- Diferenças triplas: Estadual SP vs Municipal SP x Estadual Brasil vs Municipal BR	57
Tabela 22- Diferenças triplas: Estadual SP vs Municipal SP x Estadual SE vs Municipal SE	58
Tabela 23- Diferenças triplas: Estadual SP vs Municipal SP x Estadual RJ+ES vs Municipal RJ+ES.....	59
Tabela 1A- Matching por vizinho mais próximo (NN 2) - quarta e oitava série	66
Tabela 2A- Matching por vizinho mais próximo (NN 5) - quarta e oitava série	66

SUMÁRIO

1 INTRODUÇÃO.....	11
2 REVISÃO LITERÁRIA.....	13
3 O PROGRAMA DE BONIFICAÇÃO	20
4 METODOLOGIA.....	24
5 DADOS	29
6 RESULTADOS	34
7 ANÁLISE DE ROBUSTEZ	45
7.1 – Diferenças em diferenças por município	45
7.2 – Diferenças em diferenças: Escolas estaduais x Escolas municipais paulistas ...	47
7.3 – Diferenças em diferenças: Escolas estaduais paulistas x Resto do Brasil	48
7.4 – Diferenças em diferenças: Escolas estaduais paulistas x Sudeste	51
7.4.1 – Diferenças em diferenças: Escolas estaduais paulistas x Sudeste.....	51
7.4.2 – Diferenças em diferenças: Escolas estaduais paulistas x Rio de Janeiro e Espírito Santo	54
7.5 Diferenças triplas (DDD).....	56
7.5.1 – Diferenças triplas (DDD): Estadual SP vs Municipal SP x Estadual Brasil vs Municipal BR.....	57
7.5.2 - Diferenças triplas (DDD): Estadual SP vs Municipal SP x Estadual SE vs Municipal SE.....	58
7.5.3 - Diferenças triplas (DDD): Estadual SP vs Municipal SP x Estadual RJ+ES vs Municipal RJ+ES.....	59
8 CONSIDERAÇÕES FINAIS	60
Referências Bibliográficas.....	62
Apêndice	66

1 INTRODUÇÃO

A melhora do sistema educacional é um dos grandes desafios para o crescimento da economia brasileira. É inegável o avanço ocorrido na educação nas duas últimas décadas, tendo sido a quase universalização do ensino fundamental uma de suas maiores conquistas. No entanto, a formação de capital humano requer educação de qualidade e, nesse sentido, o sistema brasileiro ainda deixa a desejar. Embora os dados do Sistema de Avaliação da Educação Básica (Saeb) apontem crescimento das notas nas últimas avaliações – após período de queda entre 1995 e 2001¹ -, o nível educacional ainda é muito baixo quando comparado a outros países. Segundo o Programa Internacional de Avaliação de Alunos (Pisa) de 2009, o Brasil ocupou a 54ª posição entre 65 países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE).

As políticas públicas na área de educação em todo o mundo voltam suas atenções para a melhora da qualidade em todos os níveis de formação escolar. Diante desse desafio, as reformas educacionais têm como um dos temas centrais programas de responsabilização dos agentes escolares pela melhoria do desempenho dos alunos.

No Brasil, a criação de métodos de avaliação do ensino a partir da década de 1990 tornou possível a elaboração de mecanismos de responsabilização. Um importante passo foi a reformulação do Saeb, em 1995. Com o passar dos anos, houve uma evolução considerável dos sistemas de avaliação nas esferas estaduais e municipais. Hoje, após algumas revisões, muitas metodologias tornaram-se compatíveis, o que permite a comparação de resultados ao longo do tempo e entre pesquisas.

Os progressos levaram a criação de um sistema de transparência do governo federal, intitulado de Plano de Metas Todos pela Educação, além de possibilitar o surgimento do Índice de Desenvolvimento da Educação Básica (Ideb), em 2007. A divulgação pública dos resultados dos exames mais recentes funciona como uma forma de responsabilizar as escolas e seus funcionários pelo desempenho dos alunos e, desta forma, criar incentivos para melhoria. O conhecimento público das notas pode gerar pressão por parte dos pais e da comunidade como um todo sobre as instituições com

¹ De 1995 a 2001 houve queda das notas de Português para a quarta série e de Matemática para quarta e oitava. As notas de Português da oitava série apresentaram queda de 1995 até 1999. Em geral, os resultados apresentaram certa estabilidade até 2005 e depois melhora em 2007 e 2009.

piores resultados. Também é comum oferecer premiações ou impor punições condicionadas ao desempenho em exames (VELOSO, 2009).

Diante desse contexto, o objetivo deste trabalho é investigar se o programa de bonificação de professores e funcionários do estado de São Paulo implementado em 2008, ao responsabilizar os agentes escolares pelo desempenho dos alunos, tem algum efeito sobre as notas nos testes padronizados.

Para este estudo, utilizaremos a metodologia de *Propensity Score Matching* e de Diferenças em Diferenças. Nossa análise de desempenho será feita com base nas notas da Prova Brasil. Esta abordagem consiste, basicamente, em encontrar um grupo de controle mais parecido com o grupo tratamento com base em características observáveis e, então, comparar os resultados que, no nosso caso, são as notas médias das escolas. Por esse motivo, utilizaremos a Prova Brasil e não o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (Saresp), dado que precisamos de resultados para o grupo de controle. Assim, avaliaremos se o programa de bonificação gera incentivos capazes de mudar a conduta do professor e de outros agentes escolares de forma que o aprendizado na escola melhore de uma forma geral levando, portanto, a um aumento da proficiência nos mais diversos testes aplicados.

No Brasil, há uma escassez de estudos empíricos de políticas de responsabilização, como o que é o objeto deste trabalho. Uma das razões é que vários programas adotados não tiveram vida longa a ponto de gerar dados suficientes para estudos empíricos. Diante disso, este trabalho pretende, de alguma forma, preencher essa lacuna, fazendo uso de dados mais recentes que permitem o uso das abordagens de avaliação de políticas públicas citadas.

O restante do trabalho se divide em mais sete seções. A próxima faz uma revisão de importantes artigos sobre políticas de responsabilização em geral. A terceira, explica o funcionamento do programa de bônus paulista. O quarto e quinto capítulos apresentam a base de dados e a metodologia, respectivamente. Os resultados fazem parte da sexta seção. Testes para checar a robustez dos resultados são realizados no capítulo sete e, por fim, algumas considerações finais são feitas no oitavo.

2 REVISÃO LITERÁRIA

As reformas educacionais nas últimas décadas têm como um de seus pilares a responsabilização de agentes escolares pelo aprendizado dos alunos. Uma das reformas mais importantes aconteceu nos Estados Unidos, em 2001, com o *No Child Left Behind Act* (NCLB), que passou a focar a política educacional na questão do desempenho com ênfase em políticas de incentivo.

Programas de *accountability*² foram ganhando espaço na medida em que as velhas práticas de prover insumos e recursos para as escolas se mostraram ineficientes. O aumento dos gastos com educação, por si só, não leva necessariamente a um crescimento da qualidade do ensino. Hanushek (2003) mostrou que recursos escolares estão pouco relacionados com o desempenho dos alunos e com a variação da qualidade dos professores.

O debate sobre mecanismos de melhoria de qualidade dos docentes e seus impactos sobre a proficiência do aluno é um dos temas mais importantes relacionados à educação. Dolton e Gutierrez (2011) discutiram as diferenças salariais de professores entre 39 países e suas consequências sobre os resultados dos estudantes. A remuneração dada a esses profissionais reflete, primeiramente, a importância que o país dá à educação. Em países com altos níveis educacionais, os professores estão no topo da distribuição salarial entre diversas categorias profissionais.

Usando dados em painel de 12 anos, foi encontrado que cerca de 10% de aumento nos salários levam a uma melhora de desempenho entre 5% e 10%. Salários mais altos funcionam como uma forma de melhorar a qualidade de ensino no longo prazo, uma vez que incentivam maior esforço dos professores para se manterem no emprego e atraem melhores profissionais, gerando um mercado de trabalho mais competitivo e, desta forma, aumentando a habilidade média dos trabalhadores da área.

No entanto, mudar a estrutura salarial de uma determinada carreira é uma tarefa de prazo mais longo. Dentro do contexto de valorização da profissão de docente, vários países têm adotado políticas de pagamento por mérito. Esse assunto não chega a ser um

² O termo *accountability* tem sido traduzido como transparência e responsabilização, entre outros. Na falta de concordância sobre a melhor tradução, foi mantido o termo em inglês (Fernandes e Gremaud (2009)). Além disso, o sistema de *accountability* é um conceito mais amplo do que apenas a responsabilização. Ele exige uma série de outros parâmetros.

consenso entre os estudiosos de educação. Milgrom e Homstrom (1991), ao analisarem o problema do agente-principal sob modelo multitarefas, citam as discussões sobre os incentivos na área de educação. Os defensores dessa prática se baseiam na teoria microeconômica do problema unidimensional padrão de agente-principal, no qual o incentivo levaria os professores a um maior esforço com impacto positivo sobre a proficiência do aluno. Já o argumento dos que não concordam é que o incentivo levaria o professor a colocar maior esforço em tarefas que podem ser medidas por meio de exames padronizados, deixando de lado funções importantes como, por exemplo, o desenvolvimento da criatividade e de diversas habilidades não cognitivas que talvez não sejam medidas nas provas.

Um dos programas norte-americanos de bonificação de professores mais antigos é o da Carolina do Norte. Desde 1996, o estado remunera os professores de acordo com o desempenho da escola. Vigdor (2008) investigou o programa utilizando dados de 2002/03 e 2005/06 e observou melhora nos exames de Matemática. No entanto, o autor faz ressalvas quanto à estrutura de incentivos adotada – baseada nos resultados gerais da escola –, que pode levar ao problema do *free-rider*.

O mesmo programa foi avaliado no estudo de Lauen (2011) utilizando a metodologia de Regressão Descontínua (Sharp) e para um período diferente. O ponto de descontinuidade é encontrado em 2007, quando houve mudança nas fórmulas de cálculo da premiação do programa. Desta vez, foram encontrados efeitos positivos e significativos apenas para o exame de Leitura do quinto ano. Para Matemática os ganhos foram semelhantes ao de Português, mas não estatisticamente significantes. O ganho de leitura no ponto de descontinuidade foi de 0,1 desvio padrão. O autor chama a atenção para a importância de estudos futuros que avaliem o impacto num prazo mais longo.

Figlio e Kenny (2007) combinaram dados do *National Education Longitudinal Survey*, de 2000, com uma pesquisa própria e fizeram uma análise *cross-section* para verificar a relação entre incentivos dos professores e desempenho dos alunos nos Estados Unidos. Eles mostraram que escolas que oferecem pagamento por mérito apresentaram notas maiores, sobretudo aquelas que têm um mínimo de participação dos pais. Entretanto, não descartam a possibilidade de que escolas mais eficientes – e que, portanto, tem pontuação maior – sejam mais propensas a adotar essas políticas. Neste caso, o aumento das notas seria um resultado espúrio. Com a metodologia empregada no trabalho, porém, não foi possível investigar tal possibilidade.

Outro estudo importante nessa área foi realizado por Lavy (2009), em Israel. O programa experimental recompensava individualmente os professores - com bônus em dinheiro - pela melhora do desempenho dos estudantes em exames de matrícula no ensino médio. Explorando a natureza dos dados, o autor utilizou a metodologia de Regressão Descontínua e de Diferenças em Diferenças e mostrou que professores alteraram métodos de ensino, o que levou ao aumento das notas. O autor qualifica incentivos aos docentes como um “método promissor de melhora da qualidade escolar”. A ressalva com relação à esse trabalho é que o experimento, que teve duração de apenas um ano, não permitiu estudar efeitos dos incentivos em outras coortes e identificar efeitos de longo prazo.

Também em caráter temporário, foi estruturado um programa de premiação de professores no Quênia condicionado ao desempenho dos estudantes. Ao fazer a análise de tal projeto, Glewwe et al. (2003) encontraram maiores ganhos de notas em escolas participantes durante o período do programa. Todavia, não verificaram mudanças em atitudes dos professores, seja em relação à maior presença em sala de aula ou nos métodos pedagógicos. A única mudança aconteceu na alocação de tempo para treinamento para a prova, o que deve explicar o aumento das notas. Os autores atribuem à natureza temporária da política os resultados não muito animadores.

Muralidharan e Sundararaman (2009) analisaram os resultados de um experimento aleatório no estado rural indiano de Andhra Pradesh que bonificava professores com melhoria da pontuação média num exame de avaliação independente em escolas primárias. Os resultados encontrados após dois anos de programa foram de aumento de 0,28 e 0,16 desvio padrão em testes de Matemática e de Linguagem, respectivamente. Matérias para as quais não havia incentivos também tiveram melhor desempenho. Os estudiosos chamam a atenção para a natureza do programa, que provavelmente teria maior chance de sucesso em países com características mais próximas à Índia e, talvez, em outros ambientes de baixos níveis de aprendizagem.

Atkinson et. al. (2009) analisaram efeitos de um programa de pagamento de professores na Inglaterra baseado em desempenho sobre pontuação nos testes e sobre o valor adicionado usando dados longitudinais e método de Diferenças em Diferenças. O programa inglês remunera o professor de acordo com o seu desempenho em cinco áreas. Uma delas é a melhora das notas dos alunos nos exames. A intenção deste projeto foi criar mecanismos de incentivo de esforço dos professores, além de atrair e manter bons

profissionais para a área de educação. Os resultados mostraram melhora de cerca de 40% nas notas por aluno.

Glazerman et al (2009) avaliaram os efeitos do *Teacher Advancement Program* (TAP) de Chicago, implementado em 2007, que promove uma remuneração extra e novas responsabilidades ao docente por meio da designação de Professor Mentor. A bonificação fica atrelada ao desempenho do aluno em sala de aula. O modelo também inclui reuniões semanais de professores e observações regulares em sala de aula por uma equipe que auxilia os professores no cumprimento de suas metas. Além da melhora das notas, a combinação de premiação e treinamento tem a intenção de atrair e reter professores talentosos. Os estudiosos não conseguiram encontrar nenhuma melhora com relação às notas ou à satisfação dos professores.

Springer et al. (2010) analisaram os resultados do *Project on Incentives in Teaching* (POINT), um projeto realizado no Sistema Escolar Metropolitano de Nashville, entre 2006/07 e 2008/09, do qual participavam voluntariamente professores de Matemática do *middle school* (sexta a oitava séries) que eram remunerados de acordo com ganhos de notas dos alunos em testes padronizados. De modo geral, não foram encontrados resultados muito favoráveis. Os alunos dos professores designados aleatoriamente para o grupo de tratamento não tiveram desempenho melhor do que aqueles cujos professores estavam no grupo controle. Evidências positivas foram encontradas apenas para a quinta série, sobretudo no segundo ano do experimento.

Fryer (2011) analisou o impacto dos incentivos financeiros a professores sobre o desempenho dos alunos na cidade de Nova Iorque. Para o estudo foram utilizados dados de cerca de 200 escolas aleatorizadas no período de 2006/2007 a 2009/2010. Além de buscar os efeitos sobre a proficiência, o estudo verificou se houve alguma alteração sobre o comportamento do professor por meio de análise de dois indicadores: falta e repetência. Não foram encontradas evidências de mudança no comportamento dos professores ou alunos, tampouco indícios de melhoras de notas. Pelo contrário, os efeitos estimados foram negativos para ensino básico e *middle school*.

Com relação à literatura brasileira, os estudos sobre responsabilização são mais pautados pelas discussões sobre implementação e os problemas associados a realidade do País que têm impossibilitado a continuidade de alguns programas estaduais. Apesar dos grandes avanços em termos de mobilização social e em debates sobre a educação, o que dá força para maior responsabilização dos agentes escolares, poucos são os

programas de *accountability* forte adotados que tiveram duração mais longa passíveis de avaliações empíricas.

Atualmente, dois estados apresentam, desde 2008, programas de bonificação de professores por desempenho de maior visibilidade - São Paulo e Pernambuco -, o que já tem permitido a realização de alguns estudos.

Menezes-Filho e Tavares (2011) avaliaram as implicações da volatilidade nos resultados dos exames escolares ao longo do tempo para o resultado do pagamento de bônus de professores de São Paulo. Eles concluíram que cerca de 35% da variância das notas é devido a amostragem e a outros fatores aleatórios e 43% deste é devido a fatores que afetam as escolas e alunos no dia da prova. O estudo verificou que cerca de 11% das escolas que atingiram as suas metas não o teriam feito caso se controlassem os erros de medida e sugere que os *policy makers* devam utilizar todas informações disponíveis para traçar metas para modelos educacionais de remuneração por desempenho.

Afora isso, ao menos o que é de nosso conhecimento, outros estudos publicados sobre responsabilização no Brasil não apresentam resultados empíricos, mas trazem debates de suma importância, principalmente, para os governos que consideram implementar algum tipo de programa de recompensa financeira aos agentes escolares.

Um primeiro empecilho para a implementação de sistemas mais eficazes apontado por Andrade (2008), é a forte tradição de que melhoras na educação decorrem de aumento de gastos, o que, conforme já mostrou Hanushek (2003), não acontece necessariamente.

Outro ponto importante destacado por Andrade (2008) é a falta de apoio por parte dos professores e a força de seus sindicatos, que são contrários aos salários atrelados ao desempenho dos alunos. Brooke (2006) explica que nos Estados Unidos e na Inglaterra, o sucesso da *accountability* esteve ligado à força do governo, que fez mudanças nas questões trabalhistas e na estrutura de incentivos dos professores.

Andrade (2008) e Brooke (2006) também indicam a falta de continuidade das políticas no médio e longo prazo – como no momento das trocas de governos, por exemplo – como um grave problema brasileiro.

Sobre as experiências no País, Brooke (2006) analisou três programas, dos quais dois tinham remuneração baseada em notas. O Programa Nova Escola, criado em 2000, no Rio de Janeiro, depois de várias modificações passou a ter, em 2004, gratificação de professores e pessoal de apoio segundo o desempenho, fluxo escolar e gestão. O complexo sistema de avaliação é alvo de várias críticas, principalmente por parte dos

professores, cujo sindicato, inclusive, moveu três ações contra o governo relativas ao programa. Já o Prêmio Educacional Escola do Novo Milênio – Educação Básica de Qualidade no Ceará, de 2001, recompensava os membros das equipes das 100 melhores escolas em termos de notas. Havia também recompensa para os melhores alunos. O autor levanta dois problemas no desenho do programa: (i) o uso do desempenho como critério único para todas as escolas sem considerar grupos de referências com base nas características socioeconômicas dos alunos e (ii) a rotatividade de pessoal e a impossibilidade da escola de decidir sobre a contratação e dispensa de professores e de gerar incentivos.

Sobre autonomia de escolas, estados, municípios e federação, Andrade (2008) diz que para o sucesso de um sistema de *accountability*, é necessário que gestores das escolas tenham autonomia, o que não ocorre no caso brasileiro. Fernandes e Gremaud (2009) ressaltam que a responsabilidade e a autonomia pela educação básica na grande maioria de estados e municípios restringem a ação do governo federal.

Ao considerar o sistema educacional brasileiro e as suas peculiaridades, Fernandes e Gremaud (2009) alertam para efeitos indesejáveis caso as estruturas de incentivo não sejam adequadas. O exemplo citado é que escolas mal avaliadas por terem alunos com piores condições socioeconômicas podem não ter incentivo para melhorar, já que o desempenho não reflete o esforço. A imposição de sanções, como a menor destinação de recursos as escolas com baixo desempenho, também pode ser prejudicial, uma vez que o aluno, nesse caso, seria punido uma segunda vez. Apesar de todas as críticas ao sistema educacional brasileiro e das difíceis barreiras para implementação de uma política de *accountability* apropriada, os autores reconhecem que importantes decisões têm sido tomadas para melhorar os sistemas de avaliação.

Brooke (2006) cita como importante passo a criação de parâmetros curriculares e mecanismo para homogeneizar a educação, como o Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e Valorização do Magistério (FUNDEF), o SAEB e a Avaliação Nacional do Rendimento no Ensino Escolar (ANRESC).

Enquanto no Brasil boa parte da literatura estuda ainda as questões relacionadas à implementação dos programas, os estudos internacionais, principalmente nos Estados Unidos, têm apresentado resultados sobre as diversas políticas adotadas. Não há um consenso sobre o sucesso de políticas que envolvem premiações a professores. Embora grande parte da literatura internacional mostre indícios de impactos positivos sobre as

notas, existem casos notórios, como o de Nova Iorque (FRYER, 2011), que chegaram até a resultados que mostravam impactos negativos.

Esta dissertação pretende contribuir para o estudo na área por meio de um trabalho empírico a respeito do efeito de premiação de docentes sobre o desempenho do aluno.

3 O PROGRAMA DE BONIFICAÇÃO

Para a análise do impacto da bonificação de professores sobre proficiência dos alunos no estado de São Paulo, é importante detalhar os principais aspectos do programa que chegou a pagar em 2010 cerca de R\$ 650 milhões em bônus a aproximadamente 210 mil funcionários da rede estadual.

As regras do sistema de remuneração foram lançadas pelo Programa de Qualidade da Escola (PQE) em maio de 2008 e o projeto de lei referente ao programa foi aprovado apenas em dezembro do mesmo ano³. Em março de 2009, o governo estadual remunerou cerca de 196 mil servidores baseado em resultados de 2008.

A Secretaria de Educação do Estado de São Paulo estipulou metas para cada série da escola e remunera todos professores do ciclo avaliado e os demais funcionários. O desenho do programa pode ser interessante no sentido de que toda a equipe escolar pode cooperar para um melhor ambiente escolar. Segundo Lavy (2009), em educação, incentivos para grupos são melhores que incentivos individuais pela própria natureza do processo educacional, que envolve trabalho em equipe, esforços de colegas professores e multitarefas. Embora não seja possível descartar a existência de *free-riders*, as regras do programa tentam eliminar o problema do absenteísmo. Funcionários com mais de 1/3 de faltas no ano não têm direito ao bônus.

As metas são estabelecidas com base no Índice de Desenvolvimento da Educação do Estado de São Paulo (Idesp) para cada etapa escolar (*s*), que por sua vez, é composto por dois outros índices:

$$IDESP_s = ID_s \times IF_s \quad (1)$$

No qual *ID* é o Índice de Desempenho e *IF* é o Índice de Fluxo e serão detalhados mais a frente. A ideia por trás dessa composição é que uma escola pode ser considerada boa quando a maior parte dos alunos desenvolve as competências e habilidades requeridas para a sua série num período de tempo ideal - o ano letivo (SECRETARIA DA EDUCAÇÃO DO ESTADO DE SÃO PAULO, 2009a).

O ID é baseado nas notas do Saresp, exame de avaliação externa paulista, que é aplicado para a segunda, quarta, sexta e oitava séries (ou terceiro, quinto, sétimo e nono

³ Projeto de Lei Complementar no. 1078

ano) do ensino fundamental, além da terceira série do ensino médio nas escolas administradas pelo governo estadual. O exame avalia a proficiência dos alunos de acordo com as exigências do Sistema Nacional de Avaliação da Educação Básica, que define o aprendizado desejado para o aluno no final de cada ciclo. São aplicadas provas de Língua Portuguesa com Redação, Matemática e Ciências Humanas (Geografia e História). Para o pagamento de bônus, considera-se apenas a proficiência em Matemática e Português.

O ID representa a quantidade relativa de alunos em cada nível de proficiência, divididas em quatro níveis: (i) Abaixo do básico, (ii) Básico, (iii) Adequado e (iv) Avançado. Os valores de referência do Saresp para a definição dos níveis de desempenho encontram-se na tabela a seguir:

Tabela 1- Valores de referência para níveis de proficiência

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
Abaixo do básico	<150	<175	<200	<225
Básico	entre 150 e 200	entre 175 e 225	entre 200 e 275	entre 225 e 300
Adequado	entre 200 e 250	entre 225 e 275	entre 275 e 325	entre 300 e 350
Avançado	>250	>275	>325	>350

Fonte: Secretaria de Educação do Estado de São Paulo

Assim, o ID para cada disciplina (j) e para cada etapa de escolarização (s) é dado por:

$$ID_{js} = \left(1 - \frac{def_{js}}{3}\right) \times 10 \quad (2)$$

Na fórmula, def é um indicador que mede a defasagem em relação às expectativas de aprendizado.

$$def_{js} = \frac{(3 \times AB) + (2 \times B) + (1 \times Ad) + (0 \times Av)}{100} \quad (3)$$

Em que AB , B , Ad e Av representam respectivamente a proporção de alunos nos níveis abaixo do básico, básico, adequado e avançado em relação ao total de alunos avaliados.

Quando a escola apresenta grande defasagem, ou seja, nível de proficiência baixo, o índice de desempenho é baixo. O ID varia de zero a dez e a média aritmética

entre os índices de Língua Portuguesa (*LP*) e Matemática (*Mat*) representa o ID para cada etapa de escolarização.

$$ID_s = \frac{ID_{LP} + ID_{Mat}}{2} \quad (4)$$

O indicador de Fluxo (IF) é dado pela taxa média de aprovação em cada etapa de escolarização e é dado por:

$$IF_s = \frac{A_i + A_{i+1} + \dots + A_n}{n} = \frac{\sum_{i=1}^n A_i}{n} \quad (5)$$

Em que A_i é a taxa de aprovação na série i e n é o número de séries⁴ da etapa de escolarização considerada.

$$A_i = \frac{\text{número de alunos aprovados}}{\text{número total de alunos matriculados}} \quad (6)$$

A partir dos dados do Idesp, calcula-se a parcela que a escola cumpre da meta. O Índice de Cumprimento (IC) em determinado ano (t) para cada etapa escolar (s) é dado pela seguinte fórmula:

$$IC_{s,t} = \left(\frac{IDESP_{s,t} - IDESP_{s,t-1}}{META_{s,t} - IDESP_{s,t-1}} \right) \quad (7)$$

O IC é limitado no intervalo de 0% a 120%. Para valores negativos, considera-se o piso (zero). Para valores superiores a 120%, considera-se o cumprimento de 120%.

Um adicional por qualidade passou a ser considerado a partir de 2009. Esse índice reflete a posição da escola em relação à média das outras escolas (Idesp global) e à sua meta de longo prazo (2030). O Índice de Qualidade (IQ) do ano t para a série s é calculado pela seguinte fórmula:

$$IQ_{s,t} = \left(\frac{IDESP_{s,t} - IDESP(global)_{s,t}}{META_{s,2030} - IDESP(global)_{s,t}} \right) \quad (8)$$

⁴ Atualmente, com o ensino fundamental de nove anos, n é igual a cinco para as séries iniciais do EF e quatro para as séries finais.

Para efeito de pagamento, considera-se a soma dos indicadores ($IQ+IC$), que está limitado entre 0 e 120%.

As metas de longo prazo do Idesp para cada série são iguais para todas as escolas e foram estabelecidas levando-se em consideração níveis de aprendizado comparáveis aos de países da Organização para a Cooperação e o Desenvolvimento Econômico (OCDE). As metas para 2030 para quarta e oitava séries do fundamental são, respectivamente, 7,0 e 6,0.

Em 2009, o Idesp médio para a quarta série da rede estadual foi de 3,86. Já para a oitava série foi de 2,84. As escolas têm, portanto, um grande desafio até 2030, pois, na média, estão muito abaixo do nível desejado.

Visando atingir esse nível de pontuação, foram estabelecidas metas anuais (que entram no cálculo do IC e IQ) para cada escola de acordo com a distância da meta de longo prazo seguindo uma função de crescimento logística (FERRAZ, 2009). Dessa forma, escolas que partem de uma pontuação baixa têm que apresentar uma melhora porcentual maior relativamente àquelas com maiores notas. Embora a variação seja maior, é razoável acreditar que em escolas piores, menores esforços possam gerar maiores mudanças de notas. Portanto, embora escolas melhores tenham que apresentar uma menor variação de notas, como já se encontram em um patamar elevado, os esforços necessários podem ser maiores e o cumprimento da meta pode ser mais difícil. Nesse sentido, torna-se interessante verificar se o desenho do programa é capaz de gerar incentivos tanto para escolas mais fracas quanto para escolas mais fortes, uma vez que elas devem ser muito diferentes em vários aspectos.

Com relação ao pagamento do bônus, o valor varia de acordo com a parcela da meta cumprida e é feito a professores do ciclo que cumpriu a meta e para funcionários com nível de absenteísmo abaixo de 1/3. Em escolas que cumprem 100% da meta, por exemplo, os funcionários recebem o equivalente a 2,4 salários. Em escolas que superaram a meta prevista, o bônus pode chegar a 2,9 salários. Escolas que não cumprem as metas, mas que, nos dois anos anteriores foram classificadas entre as 10% melhores, também recebem uma remuneração. Esse seria um incentivo para escolas boas, cujas metas são difíceis de cumprir, pois as notas já estão num patamar elevado. Funcionários das diretorias regionais recebem bônus pela média da região em que atuam.

4 METODOLOGIA

Neste trabalho, pretende-se avaliar o efeito da política de *accountability* forte adotada no estado de São Paulo a partir de 2008 sobre o desempenho dos alunos. Para averiguar o impacto desse programa, deveríamos comparar o desempenho em duas situações, com e sem a presença da política. No entanto, é impossível observar essas duas situações simultaneamente.

Segundo Abadie et al. (2001), se a exposição à política for aleatória, pode-se comparar dois grupos: um que recebeu o incentivo (grupo de tratamento) e outro que não (grupo controle). Depois, bastaria calcular a diferença do desempenho nas duas situações para avaliar o impacto da *accountability*. Entretanto, em nosso estudo, a escolha não é aleatória, uma vez que apenas escolas da rede estadual de São Paulo receberam essa política de incentivos na forma de bônus em 2008.

Para lidar com esse problema, a metodologia adotada foi o *matching*, ou pareamento. A abordagem consiste em separar os indivíduos em dois grupos (tratamento e controle), considerando características observáveis. Desse modo, a aleatoriedade da escolha estaria garantida, condicional em características observáveis. Em nosso trabalho, o grupo de tratamento seria formado por escolas estaduais paulistas, enquanto o de controle abrange escolas de estados que não contam com esse programa e escolas das redes municipais do estado de São Paulo.

Após o pareamento, podemos eliminar os efeitos de características não observáveis que afetam o desempenho e que são invariantes no tempo. Fazemos isso, utilizando um estimador de Diferenças em Diferenças e comparando os resultados nas duas situações.

Para formalizar o procedimento que será realizado, considere o status tratamento de um indivíduo i por meio de uma variável binária D que assume o valor 1 quando participa do programa e 0, caso contrário. Em adição, suponha que o valor da variável de interesse para o indivíduo i seja Y_{i1} , caso ele seja submetido ao tratamento (1) e Y_{i0} , caso não seja exposto ao tratamento (0). O efeito do tratamento para o indivíduo i (τ_i) é dado por $\tau_i = Y_{i1} - Y_{i0}$.

O efeito médio do tratamento sobre os tratados (*average treatment effect on treated* – ATT) é dado por:

$$\tau_{ATT} = E(Y_{i1} - Y_{i0} | D_i = 1) = E(Y_{i1} | D_i = 1) - E(Y_{i0} | D_i = 1) \quad (9)$$

No entanto, o contrafactual de um indivíduo sob tratamento, $E(Y_{i0} | D_i = 1)$, não é observável.

O efeito médio do tratamento sobre os não tratados (*average treatment effect on the control* – ATC) é dado por:

$$\tau_{ATC} = E(Y_{i1} - Y_{i0} | D_i = 0) = E(Y_{i1} | D_i = 0) - E(Y_{i0} | D_i = 0) \quad (10)$$

Neste caso, o contrafactual do indivíduo que não recebe tratamento é $E(Y_{i1} | D_i = 0)$, que não é observável.

Os resultados observáveis são, portanto, apenas $E(Y_{i1} | D_i = 1)$ e $E(Y_{i0} | D_i = 0)$. Se pudéssemos considerar que $E(Y_{i0} | D_i = 1) = E(Y_{i0} | D_i = 0)$, poderíamos substituir $E(Y_{i0} | D_i = 0)$ na equação (9) e, então, teríamos o ATT. Mas essa é uma suposição muito forte, pois indivíduos tratados e não tratados são diferentes. Uma estimação feita desta forma muito provavelmente estaria viesada.

É possível evitar esse viés se for assumida a Hipótese da Independência Condicional (*Conditional Independence Assumption* - CIA), que garante que a escolha do grupo de tratamento é independente dos resultados. Além disso, é necessário garantir que, dado um conjunto de características observáveis X , indivíduos com as mesmas características têm a mesma probabilidade de participar do grupo tratamento ou controle. Dessa forma:

$$Y_{i1}, Y_{i0} \perp D_i | X_i \quad (11)$$

$$Y_{i1}, Y_{i0} \perp D_i | X_i \Rightarrow (Y_{i0} | X_i, D_i = 1) = (Y_{i0} | X_i, D_i = 0) = (Y_{i0} | X_i) \quad (12)$$

em que $\perp$ denota independência.

A CIA afirma que os resultados potenciais independem da participação do programa dadas as características observáveis X .

Podemos reescrever a equação (9) da seguinte forma:

$$\tau_{ATT} = E(Y_{i1} - Y_{i0} | D_i = 1, X) = E(Y_{i1} | D_i = 1, X) - E(Y_{i0} | D_i = 0, X) \quad (13)$$

Assim, se vale a CIA, teremos uma estimativa não viesada do ATT. Utilizaremos a metodologia de *matching* (ou pareamento), que consiste em encontrar um grupo de controle o mais parecido possível ao grupo de tratamento de modo que os resultados desses possam ser comparados. Assim, condicionando nas variáveis X , é possível comparar os dois grupos como se fosse um experimento aleatório.

Quanto maior o número de características observáveis, maior a semelhança entre os indivíduos, tornando-os comparáveis. Mas, se o número de variáveis observadas for muito grande, menor é a probabilidade de encontrar alguém parecido. Ao considerar todas as covariadas relevantes, o pareamento pode apresentar o problema da multidimensionalidade. Para n características observáveis, o número de possíveis valores para o vetor X será 2^n .

Um modo de resolver o problema da multidimensionalidade é o *propensity score* (escore de propensão), proposto por Rosenbaun e Rubin (1983). O método consiste em estimar uma probabilidade, $P(X)$, do indivíduo receber tratamento, dadas suas covariadas. Dessa forma, o pareamento é feito com base em apenas um escalar.

Esse método atribuirá uma probabilidade entre zero e um de a escola receber tratamento. Seja $P(X)$ a probabilidade de um indivíduo i receber tratamento, definida como:

$$P(X) = Pr(D_i = 1|X) \quad (14)$$

Segundo esta metodologia, X pode ser substituído por $P(X)$:

$$\begin{aligned} Y_{i1}, Y_{i0} &\perp D_i | X_i \\ Y_{i1}, Y_{i0} &\perp D_i | P(X_i) \end{aligned}$$

Mesmo controlando pelas variáveis observadas, ainda podem existir características não observadas que afetam o resultado de interesse. Para este trabalho, a base de dados nos permite utilizar o método de Diferenças em Diferenças (DD) para eliminar o efeito de características individuais não observáveis que são fixas no tempo.

O DD consiste na comparação da diferença dos resultados (Y) do grupo de tratamento e controle ao longo do tempo (diferença no tempo). Considere Y^A , resultado antes da intervenção Y^D o resultado depois da intervenção; DD determina o efeito da intervenção sobre o resultado de interesse Y . Dito de outra forma:

$$DD = E(Y_1^D - Y_1^A | D = 1) - E(Y_0^D - Y_0^A | D = 1) \quad (15)$$

No entanto a última parcela do lado direito da equação $E(Y_0^D - Y_0^A | D = 1)$ não é observável. Assim, para que seja possível calcular o estimador de DD, é necessário

admitir a Hipótese de Tendência Comum que é dada por: $E(Y_0^D - Y_0^A|D = 1) = E(Y_0^D - Y_0^A|D = 0)$.

Se a hipótese é válida, o impacto médio do programa é dado por:

$$DD = E(Y_1^D - Y_1^A|D = 1) - E(Y_0^D - Y_0^A|D = 0) \quad (16)$$

Uma vez que já foi determinado o grupo de comparação por meio do *matching*, o DD pode ser escrito como a diferença do ATT antes e depois da intervenção.

$$DD = [E(Y_1^D|D = 1) - E(Y_0^D|D = 0)] - [E(Y_1^A|D = 1) - E(Y_0^A|D = 0)] \quad (17)$$

ou

$$\beta_3 = \tau_{ATT}^D - \tau_{ATT}^A \quad (18)$$

O estimador DD também pode ser calculado pelo método de Mínimos Quadrados Ordinários a partir do seguinte modelo:

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 t + \beta_3 D_i t + \varepsilon_{it} \quad (19)$$

Em que:

$D_i = 1$, se recebeu tratamento e 0, caso contrário

$t = 1$, depois do tratamento e 0, antes.

No contexto do presente trabalho, reescrevemos:

$$Nota\ média_{it} = \beta_0 + \beta_1 EE_SP_i + \beta_2 Ano + \beta_3 EE_SP_i * Ano + \varepsilon_{it} \quad (20)$$

Em que:

$EE_SP_i = 1$, se é uma escola estadual paulista (ou seja, recebeu tratamento) e 0, caso contrário

$Ano = 1$, se é referente a 2009 (depois do tratamento) e 0 se é referente a 2007(antes).

No nosso caso, o β_3 estimado será o impacto do programa sobre o desempenho do aluno.

Para testar a robustez dos resultados dessa estratégia de identificação, o método de diferenças em diferenças será testado com base em outros grupos de controle, a saber: (i) escolas municipais paulistas; (ii) escolas dos demais estados brasileiros; (iii) escolas dos demais estados do Sudeste (MG, ES e RJ) e (iii) escolas do ES e RJ. Esses resultados são apresentados nas seções 7.2 a 7.4. A ideia aqui é verificar o quão

robustos são os resultados a alterações no grupo de controle, isto é, se a comparação com escolas semelhantes de regiões distintas não altera os resultados.

O método de DD pressupõe que as nota média das escolas estaduais paulistas tenha a mesma trajetória da nota do grupo de controle. Todavia, ainda pode haver fatores (não relacionados ao programa de bonificação) que afetem as notas dos grupos de forma desigual como, por exemplo, as diferenças inerentes entre as administrações estaduais e municipais. Para eliminar essas possíveis diferenças, podemos combinar mais de um grupo de controle por meio das diferenças triplas (DDD).

Os modelos estimados serão do tipo:

$$Nota\ média_{it} = \beta_0 + \beta_1 SP_i + \beta_2 EE_i + \beta_3 Ano + \beta_4 SP_i * Ano + \beta_5 EE_i * Ano + \beta_6 EE_i * SP_i + \beta_7 EE_i * SP_i * Ano + \varepsilon_{it} \quad (21)$$

Em que:

$SP_i = 1$, se a escola pertence ao estado de São Paulo e 0, caso contrário

$EE_i = 1$, se a escola é estadual e 0, caso contrário

$Ano = 1$, depois do tratamento e 0, antes.

Neste caso, o β_7 é o nosso coeficiente de interesse e pode ser escrito como:

$$\hat{\beta}_7 = [(Nota_{E,SP,2009} - Nota_{E,SP,2007}) - (Nota_{M,SP,2009} - Nota_{M,SP,2007})] - [(Nota_{E,F,2009} - Nota_{E,F,2007}) - (Nota_{M,F,2009} - Nota_{M,F,2007})] \quad (22)$$

Em que E =Estadual , M =Municipal, SP = Estado de São Paulo e F =Fora do Estado de São Paulo.

5 DADOS

Para avaliarmos o efeito do programa de bonificação sobre a proficiência dos alunos utilizaremos os dados da Prova Brasil. Embora a bonificação esteja atrelada às notas do Saresp, que é o exame aplicado pelo governo do estado de São Paulo, esses dados limitam nossa avaliação e a aplicação de metodologias de avaliação de políticas públicas. Como o exame é realizado apenas em São Paulo, não existem outros estados que possam servir como contrafactual.

Felizmente, um dos principais avanços na área de educação foi o surgimento dos exames de avaliação escolar. Muitos estados possuem sistemas próprios de avaliação e adotam padrões muito parecidos. Alguns exemplos são o Simave (MG), o Saepe (PE) e o Saers (RS), entre outros.

Em nível federal é aplicada a Prova Brasil a cada dois anos. A Prova Brasil ou Avaliação Nacional do Rendimento Escolar (Anresc) tem caráter censitário e juntamente com a Avaliação Nacional da Educação Básica (Aneb) que é amostral, compõe o Saeb. São realizados testes de Matemática (com foco na resolução de problemas) e de Língua Portuguesa (com foco em leitura) para a quarta e oitava série do fundamental e terceiro ano do ensino médio. Este exame tem características muito semelhantes ao Saresp: usam como base a TRI (Teoria de resposta ao item) e avaliam os anos finais de cada ciclo (quarta e oitava séries do fundamental e terceiro ano do ensino médio) em pelo menos duas disciplinas básicas (Matemática e Português).

Segundo a Secretaria da Educação do Estado de São Paulo (2009b), em 2007 foram introduzidas mudanças importantes no Saresp num esforço para tornar o exame mais adequado tecnicamente às características de um sistema de avaliação em larga escala que permitisse acompanhar a evolução da qualidade do sistema estadual de ensino ao longo dos anos. Além de o Saresp ter adequado as habilidades avaliadas aos padrões do Saeb, a escala de proficiência passou a ser igual ao da prova nacional. Assim, a partir de 2007, os resultados obtidos pelos alunos paulistas nos dois exames ao longo dos anos tornaram-se passíveis de comparação.

Dessa forma, a semelhança das avaliações nos leva a acreditar que o efeito de um programa que pretende melhorar o ensino deve ter impacto positivo nas notas dos dois exames.

Além da Prova Brasil nos proporcionar comparações com escolas de outros estados, ela evita alguns tipos de *gaming*⁵. Como este exame não está diretamente atrelado ao pagamento de bônus, não há incentivos para seleção de alunos, ajuda na prova ou outros tipos de fraudes que podem ocorrer no dia da aplicação. Mesmo que não seja possível evitar todos os efeitos indesejados, o uso do exame nacional ajuda a minimizar alguns deles.

Para realização de nossas estimações, utilizaremos as notas das escolas da Prova Brasil de 2007 e 2009 de Língua Portuguesa e Matemática para a quarta série (ou quinto ano) e oitava série (ou nono ano).

Optou-se por não usar os dados de 2005, pois nas escolas estaduais paulistas - que são justamente as de interesse deste estudo-, o exame foi realizado com uma amostra de alunos. Embora todas as escolas tenham participado, naquelas em que havia mais de uma turma por série, nem todas fizeram o exame, o que poderia viesar nossos resultados caso a amostra não tenha sido aleatória.

Além das provas, os alunos, professores e diretores respondem a um questionário socioeconômico que fornecem as covariadas necessárias para a estimação do *Propensity Score*. Foram testadas variáveis conhecidas da literatura que podem estar relacionadas com a proficiência dos alunos e mantidas na estimação do escore de propensão apenas aquelas que se mostraram significativas e com maiores coeficientes.

A tabela 2 mostra a descrição dos dados da Prova Brasil de 2007. As diferenças entre as covariadas da quarta e oitava série das escolas estaduais paulistas não são muito grandes. Já em comparação com os dados do País como um todo, as escolas estaduais de São Paulo mostraram indicadores melhores. A média da proporção de alunos que tem computador com internet em casa é maior entre alunos das escolas estaduais paulistas do que da média do Brasil tanto para a quarta quanto para a oitava série. A proporção de mães com escolaridade acima do ensino médio também é maior para alunos das escolas estaduais paulistas e, além disso, os pais nesses grupos tendem a comparecer mais às reuniões nas escolas.

Com relação às características de professores, a diferença mais importante é que a proporção com salário acima de R\$1300 na quarta série está mais concentrada nas escolas estaduais paulistas (30%), já na média do Brasil este número cai para 20%. Com relação aos diretores também existe diferença salarial considerável. 58% dos diretores

⁵ comportamentos estratégicos intencionais, voltados apenas à aumentar a nota média das prova, sem necessariamente melhorar o aprendizado dos alunos

em escolas do primeiro ciclo do fundamental recebem mais de R\$1.500 no País como um todo, esse número sobe para 96% nas escolas estaduais paulistas. Para a oitava série, a parcela dos diretores em escolas estaduais paulistas que ganham acima daquele valor é de 97%, enquanto para o conjunto de todas as unidades de federação é de 70%.

Tabela 2- Descrição das variáveis, médias para Brasil e Estadual São Paulo – 2007

	Quarta série		Oitava série	
	Brasil	São Paulo (Estadual)	Brasil	São Paulo (Estadual)
Proporção de alunos com a característica dentro da escola				
Pardos	0.415	0.422	0.450	0.407
Branco	0.316	0.380	0.351	0.436
Negro	0.107	0.091	0.107	0.096
Computador e Internet	0.166	0.292	0.201	0.311
Moram com 1 pessoa	0.022	0.026	0.035	0.032
Mãe completou a 8a, mas não o Ensino Médio	0.126	0.141	0.157	0.184
Mãe superior incompleto	0.087	0.119	0.177	0.205
Mãe superior completo	0.086	0.110	0.060	0.057
Pais sempre vão à reunião	0.549	0.654	0.546	0.665
Possuem de 1 a 20 livros em casa	0.520	0.523	0.593	0.602
Possuem de 21 a 100 livros em casa	0.157	0.157	0.192	0.183
Possuem mais de 100 livros em casa	0.091	0.091	0.063	0.055
Trabalham fora	0.135	0.129	0.206	0.182
Fizeram o maternal	0.334	0.339	0.420	0.422
Fizeram pré-escola	0.340	0.400	0.370	0.429
Nunca reprovaram	0.603	0.813	0.620	0.786
Moram com os pais	0.537	0.618	0.598	0.617
Moram com pai ou a mãe	0.334	0.324	0.297	0.322
Proporção de professores com a característica dentro da escola				
Experiência de até 2 anos	0.328	0.552	0.336	0.456
Experiência de 3 a 9 anos	0.388	0.225	0.402	0.395
Salário até R\$1300	0.303	0.081	0.249	0.074
Salário mais de R\$ 1300	0.194	0.301	0.312	0.289
Diretores (Dummy=1 se possui a característica e 0, caso contrário)				
Salário mais de R\$ 1500	0.588	0.959	0.705	0.971
Experiência de até 4 anos	0.536	0.251	0.510	0.250
Experiência de 5 a 15 anos	0.385	0.558	0.401	0.542
Conselho de escola se reúne pelo menos uma vez/ano	0.832	0.987	0.891	0.988
Conselho de classe se reúne pelo menos uma vez/ano	0.808	0.986	0.881	0.987

Fonte: Elaboração própria com base nos dados do Inep

Vale ressaltar ainda que o número de escolas no grupo de tratamento utilizado no *matching* é menor do que o número total de escolas oferecido na base de dados original devido à perda de informações das covariadas nos questionários da Prova Brasil de alunos, professores ou diretores. Para a quarta série, das 1987 escolas estaduais paulistas iniciais, restaram 1.753 para Matemática e 1.754 para Português. Já para oitava

série, a perda foi menor e passou de 3.490 originais para 3.330 para Matemática e 3.339 para Português.

Mesmo assim, a média de notas não apresentou diferença significativa com relação à amostra original. Houve maiores perdas nos questionários da quarta série e, mesmo que isso possa implicar em algum viés de seleção (caso escolas cujos alunos respondam a questionários sejam melhores, por exemplo), o *Propensity Score* ajudará a resolver esse problema.

Os gráficos abaixo mostram a evolução da proficiência para alunos da quarta e oitava série na rede estadual paulista. Os valores de 1995 a 2005 referem-se à base de dados do Saeb e os valores de 2007 e 2009 foram calculados com base nos microdados da Prova Brasil. Apesar de os dados do Saeb serem amostrais, os dados são representativos para as redes estaduais.

Figura 1- Evolução da proficiência de Língua Portuguesa e Matemática da quarta série - (1995 a 2009)

Fonte: Inep

Figura 2- Evolução da proficiência de Língua Portuguesa e Matemática da oitava série - (1995 a 2009)

Fonte: Inep

Para a quarta série, aumentos consideráveis nas notas ocorreram entre 2007 e 2009 (antes e depois da bonificação). Considerando o mesmo período, a evolução foi bem menos favorável para a oitava série, para o qual houve melhora apenas na nota de Língua Portuguesa, enquanto a nota de Matemática ficou relativamente estável.

6 RESULTADOS

Neste capítulo, serão apresentados os resultados da estimação do *Propensity Score Matching* e do método de Diferenças em Diferenças.

Estimamos o escore de propensão, ou seja, a probabilidade de ser tratado, por meio do modelo Probit. No contexto deste trabalho, será estimada a probabilidade de ser uma escola estadual paulista.

A vasta disponibilidade de dados da Prova Brasil 2007 nos permite utilizar um número muito grande de variáveis explicativas. Quanto mais covariadas relevantes conseguirmos utilizar, mais o grupo de controle tenderá a ser parecido com o grupo de tratamento, porém, a dificuldade de encontrar um controle aumenta. Num primeiro exercício, ao utilizar um conjunto muito grande de variáveis explicativas, poucos indivíduos foram pareados, levando a uma pequena região de suporte comum. Desta forma, optou-se por um modelo com um número menor de covariadas.

As especificações que apresentaram uma boa qualidade do *matching*, com covariadas balanceadas e maior região de suporte comum foram as seguintes:

Tabela 3- Coeficientes estimados do Probit - quarta série - 2007

Variável dependente: Escola estadual paulista	Português	Matemática
Variáveis de alunos		
Pardo	0,730*** (0,133)	0,705*** (0,133)
Tem computador com internet	2,285*** (0,160)	2,240*** (0,160)
Escolaridade da mãe (8 a 10 anos)	1,665*** (0,311)	1,654*** (0,312)
Possui de 1 a 20 livros em casa	-2,489*** (0,214)	-2,513*** (0,214)
Possui de 21 a 100 livros em casa	-4,032*** (0,294)	-3,999*** (0,294)
Possui mais de 100 livros em casa	-3,492*** (0,365)	-3,458*** (0,365)
Trabalha fora	4,322*** (0,266)	4,258*** (0,265)
Fez maternal	-1,412*** (0,116)	-1,422*** (0,115)
Nunca foi reprovado	6,333*** (0,198)	6,341*** (0,198)
Mora com os pais	1,079*** (0,384)	1,080*** (0,386)
Mora com o pai ou a mãe	1,578*** (0,431)	1,607*** (0,433)
Variáveis de professores		
Experiência de até 2 anos	0,774*** (0,053)	0,742*** (0,053)
Experiência de 3 a 9 anos	-0,334*** (0,062)	-0,370*** (0,061)
Salário até R\$1300	-1,083*** (0,076)	-1,115*** (0,077)
Variáveis de diretores		
Salário mais de R\$1500	0,816*** (0,065)	0,827*** (0,065)
Experiência de até 4 anos	-0,682*** (0,053)	-0,682*** (0,053)
Experiência de 5 a 15 anos	-0,206*** (0,049)	-0,201*** (0,049)
Conselho de escola se reúne pelo menos uma vez/ano	1,034*** (0,101)	1,029*** (0,101)
Conselho de classe se reúne pelo menos uma vez/ano	0,563*** (0,116)	0,547*** (0,115)
Constante	-8,125*** (0,415)	-8,065*** (0,417)
No. de observações	32.083	32.064
Pseudo R-quadrado	0,469	0,468

Erro padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

Tabela 4- Coeficientes estimados do Probit - Oitava série - 2007

Variável dependente: Escola estadual paulista	Português	Matemática
Variáveis de alunos		
Tem computador com internet	0,915*** (0,096)	0,897*** (0,093)
Escolaridade da mãe (8 a 10 anos)	1,637*** (0,229)	1,673*** (0,226)
Possui de 1 a 20 livros em casa	-2,110*** (0,230)	-2,139*** (0,227)
Possui de 21 a 100 livros em casa	-3,478*** (0,244)	-3,530*** (0,239)
Possui mais de 100 livros em casa	-3,404*** (0,398)	-3,651*** (0,391)
Trabalha fora	0,445*** (0,163)	
Fez maternal	-0,616*** (0,083)	-0,599*** (0,081)
Nunca foi reprovado	6,967*** (0,147)	6,717*** (0,143)
Mora com os pais	3,067*** (0,329)	2,745*** (0,323)
Mora com o pai ou a mãe	6,372*** (0,361)	5,862*** (0,353)
Variáveis de professores		
Salário até R\$1300	-0,999*** (0,053)	-0,751*** (0,050)
Variáveis de diretores		
Salário mais de R\$1500	1,006*** (0,057)	1,019*** (0,057)
Constante	-8,765*** (0,327)	-8,147*** (0,311)
No. de observações	23.204	23.201
Pseudo R-quadrado	0,421	0,408
Erro padrão entre parênteses		
*** p<0.01, ** p<0.05, * p<0.1		
Fonte: Elaboração própria		

Após a estimação do escore de propensão foi feito o *matching* utilizando o método do primeiro vizinho mais próximo. Outros métodos foram testados como robustez (Tabelas 1A e 2A), e os resultados foram semelhantes. Como a amostra é relativamente grande, não há muita diferença nos resultados dos diversos métodos já que, assintoticamente, eles devem convergir. Para o prosseguimento de nossa análise optamos fazer o DD para a amostra pareada pelo método de *nearest neighbor* para o primeiro vizinho, sem reposição e com caliper de 0,002.

A tabela 5 mostra que o número de escolas tratadas e controles dentro do suporte comum são relativamente grandes. A boa aderência do *matching* também pode ser notada ao observar a distribuição do escore de propensão (figuras 3 a 6).

Tabela 5- Suporte comum

		Suporte Comum		
		Fora do suporte	Dentro do suporte	Total
Quarta série	Português	Controles	0	30.329
		Tratados	260	1.754
	Matemática	Controles	0	30.311
		Tratados	266	1.487
Oitava série	Português	Controles	0	19.865
		Tratados	764	3.339
	Matemática	Controles	0	19.871
		Tratados	720	3.330

Fonte: Elaboração própria

Figura 3- Histograma depois do matching – quarta série - Português

Fonte: Elaboração própria

Figura 4- Histograma depois do matching – quarta série - Matemática
Fonte: Elaboração própria

Figura 5- Histograma depois do matching – oitava série - Português
Fonte: Elaboração própria

Figura 6- Histograma depois do matching – oitava série - Matemática

Fonte: Elaboração própria

Outro procedimento importante nesse tipo de metodologia é a checagem das condições de balanceamento. As tabelas a seguir (tabelas 6 a 9) mostram a média das variáveis no tratamento e no controle. A hipótese nula deste teste é que a diferença das médias é zero, ou seja, as médias dos dois grupos são iguais. É esperado que, antes do pareamento, as médias sejam diferentes. Depois de pareados, percebe-se que na maioria das covariadas não foi possível rejeitar a nula e, portanto, tem-se um *matching* com um bom balanceamento de covariadas.

Tabela 6- Diferença das médias das variáveis explicativas – Quarta série – Português

Português - quarta série	Antes do Matching			Depois do Matching		
	Tratamento	Controle	p>t	Tratamento	Controle	p>t
Variáveis de alunos						
Pardo	0,422	0,415	0,076	0,421	0,42	0,821
Tem computador com internet	0,291	0,158	0,000	0,287	0,291	0,53
Escolaridade da mãe (8 a 10 anos)	0,141	0,125	0,000	0,141	0,142	0,95
Possui de 1 a 20 livros em casa	0,523	0,520	0,259	0,523	0,521	0,289
Possui de 21 a 100 livros em casa	0,156	0,157	0,558	0,157	0,158	0,849
Possui mais de 100 livros em casa	0,091	0,091	0,572	0,092	0,092	0,702
Trabalha fora	0,129	0,135	0,003	0,127	0,127	0,971
Fez maternal	0,339	0,334	0,231	0,341	0,342	0,916
Nunca foi reprovado	0,813	0,590	0,000	0,805	0,805	0,999
Mora com os pais	0,618	0,532	0,000	0,618	0,617	0,849
Mora com o pai ou a mãe	0,324	0,335	0,000	0,323	0,322	0,749
Variáveis de professores						
Experiência de até 2 anos	0,552	0,315	0,000	0,527	0,524	0,837
Experiência de 3 a 9 anos	0,226	0,398	0,000	0,241	0,252	0,295
Salário até R\$1300	0,082	0,317	0,000	0,089	0,081	0,2
Variáveis de diretores						
Salário mais de R\$1500	0,959	0,566	0,000	0,954	0,952	0,608
Experiência de até 4 anos	0,251	0,552	0,000	0,277	0,28	0,838
Experiência de 5 a 15 anos	0,558	0,375	0,000	0,557	0,542	0,398
Conselho de escola se reúne pelo menos uma vez/ano	0,987	0,824	0,000	0,985	0,977	0,141
Conselho de classe se reúne pelo menos uma vez/ano	0,986	0,797	0,000	0,985	0,986	0,878

Fonte: Elaboração própria

Tabela 7- Diferença das médias das variáveis explicativas – Quarta série- Matemática

Matemática - quarta série	Antes do Matching			Depois do Matching		
	Tratamento	Controle	p>t	Tratamento	Controle	p>t
Variáveis de alunos						
Pardo	0,422	0,415	0,076	0,421	0,419	0,62
Tem computador com internet	0,291	0,158	0,000	0,286	0,290	0,45
Escolaridade da mãe (8 a 10 anos)	0,141	0,125	0,000	0,141	0,142	0,54
Possui de 1 a 20 livros em casa	0,523	0,520	0,255	0,523	0,521	0,36
Possui de 21 a 100 livros em casa	0,156	0,157	0,564	0,157	0,158	0,78
Possui mais de 100 livros em casa	0,091	0,091	0,525	0,091	0,090	0,69
Trabalha fora	0,129	0,135	0,003	0,128	0,125	0,15
Fez maternal	0,339	0,334	0,255	0,341	0,336	0,43
Nunca foi reprovado	0,813	0,590	0,000	0,803	0,806	0,47
Mora com os pais	0,618	0,532	0,000	0,617	0,622	0,14
Mora com o pai ou a mãe	0,324	0,335	0,000	0,323	0,319	0,12
Variáveis de professores						
Experiência de até 2 anos	0,541	0,311	0,000	0,515	0,506	0,49
Experiência de 3 a 9 anos	0,223	0,397	0,000	0,242	0,257	0,17
Salário até R\$1300	0,078	0,317	0,000	0,085	0,079	0,31
Variáveis de diretores						
Salário mais de R\$1500	0,959	0,566	0,000	0,953	0,952	0,86
Experiência de até 4 anos	0,251	0,552	0,000	0,278	0,295	0,29
Experiência de 5 a 15 anos	0,558	0,375	0,000	0,553	0,532	0,25
Conselho de escola se reúne pelo menos uma vez/ano	0,987	0,823	0,000	0,985	0,978	0,22
Conselho de classe se reúne pelo menos uma vez/ano	0,986	0,797	0,000	0,984	0,985	0,88

Fonte: Elaboração própria

Tabela 8- Diferença das médias das variáveis explicativas – Oitava série – Português

	Antes do Matching			Depois do Matching		
	Tratamento	Controle	p>t	Tratamento	Controle	p>t
Variáveis de alunos						
Tem computador com internet	0,311	0,181	0,000	0,302	0,310	0,120
Escolaridade da mãe (8 a 10 anos)	0,184	0,152	0,000	0,181	0,183	0,438
Possui de 1 a 20 livros em casa	0,603	0,592	0,000	0,601	0,601	0,989
Possui de 21 a 100 livros em casa	0,183	0,193	0,000	0,185	0,187	0,391
Possui mais de 100 livros em casa	0,055	0,064	0,000	0,057	0,056	0,392
Trabalha fora	0,182	0,211	0,000	0,184	0,184	0,981
Fez maternal	0,422	0,418	0,407	0,421	0,421	0,968
Nunca foi reprovado	0,787	0,593	0,000	0,769	0,769	0,796
Mora com os pais	0,618	0,596	0,000	0,618	0,620	0,505
Mora com o pai ou a mãe	0,321	0,292	0,000	0,318	0,317	0,662
Variáveis de professores						
Salário até R\$1300	0,066	0,278	0,000	0,078	0,086	0,222
Variáveis de professores						
Salário mais de R\$1500	0,972	0,661	0,000	0,964	0,964	0,881

Fonte: Elaboração própria

Tabela 9- Diferença das médias das variáveis explicativas – Oitava série – Matemática

	Antes do Matching			Depois do Matching		
	Tratamento	Controle	p>t	Tratamento	Controle	p>t
Variáveis de alunos						
Tem computador com internet	0,311	0,182	0,000	0,302	0,308	0,224
Escolaridade da mãe (8 a 10 anos)	0,184	0,152	0,000	0,181	0,182	0,594
Possui de 1 a 20 livros em casa	0,603	0,592	0,000	0,602	0,600	0,418
Possui de 21 a 100 livros em casa	0,183	0,193	0,000	0,185	0,188	0,194
Possui mais de 100 livros em casa	0,055	0,064	0,000	0,056	0,056	0,745
Fez maternal	0,422	0,418	0,368	0,421	0,418	0,559
Nunca foi reprovado	0,786	0,593	0,000	0,768	0,769	0,752
Mora com os pais	0,618	0,596	0,000	0,617	0,621	0,099
Mora com o pai ou a mãe	0,321	0,291	0,000	0,319	0,316	0,172
Variáveis de professores						
Salário até R\$1300	0,074	0,279	0,000	0,086	0,085	0,846
Variáveis de professores						
Salário mais de R\$1500	0,972	0,661	0,000	0,965	0,964	0,823

Fonte: Elaboração própria

Por fim, pode-se também checar o que ocorreu com o Pseudo R-quadrado após o pareamento. Ao se reestimar o escore de propensão utilizando apenas informações das escolas tratadas e de controle selecionadas, o valor do Pseudo R-quadrado deve diminuir, como mostrado na Tabela 10, já que os grupos passam a ser semelhantes nessas covariadas. O poder explicativo das covariadas cai, ou seja, elas não conseguem mais explicar as diferenças existentes entre grupo de controle e tratamento e a probabilidade de participação.

Tabela 10- Pseudo R-quadrado

	Pseudo-R2			
	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
Antes do Matching	0,469	0,468	0,421	0,408
Depois do Matching	0,002	0,004	0,001	0,001

Fonte: Elaboração própria

Embora os procedimentos anteriores indiquem que o pareamento foi bem feito, as notas médias entre grupos de controle e tratamento no período pré-tratamento (2007) ainda mostraram-se estatisticamente diferentes para quarta e oitava séries (Tabela 11). Essa diferença pode indicar que características não observáveis podem ter afetado a proficiência dos alunos.

Tabela 11- Matching por vizinho mais próximo (quarta e oitava série)

Amostra		Tratados	Controles	Diferença	Desvio Padrão	T-stat
Quarta série						
Português 2007	Antes do Matching	176,653	171,731	4,922	0,447	11,020
	Depois do Matching	176,186	184,422	-8,236	0,586	-14,060
Matemática 2007	Antes do Matching	193,701	189,214	4,486	0,490	9,150
	Depois do Matching	193,010	203,288	-10,278	0,638	-16,100
Oitava série						
Português 2007	Antes do Matching	231,263	227,163	4,100	0,316	12,960
	Depois do Matching	230,133	237,260	-7,127	0,410	-17,400
Matemática 2007	Antes do Matching	242,129	239,043	3,086	0,352	8,770
	Depois do Matching	241,199	250,725	-9,526	0,450	-21,150

Fonte: Elaboração própria

Após o pareamento, o próximo passo é aplicar o estimador de Diferenças em Diferenças. Como explicado anteriormente, esse método permite eliminar o efeito de características não observadas (fixas no tempo) que podem ter afetado o resultado.

Utilizando o método de MQO, estimou-se o seguinte modelo:

$$Nota\ média_{it} = \beta_0 + \beta_1 EE_SP_i + \beta_2 Ano + \beta_3 EE_SP_i * Ano + \varepsilon_{it} \quad (20)$$

Em que *Ano* é igual a 1 para 2009 e 0 para 2007. Desta forma, o valor de β_3 representará o impacto do programa sobre o desempenho do aluno tratado.

Os resultados mostram indícios de impactos positivos apenas para a quarta série do ensino fundamental (Tabela 12). No exame de Matemática, o efeito estimado foi de cerca de 6,4 pontos na prova, o que representa 0,42 desvio padrão. Já para o exame de Português, o efeito do bônus mostra-se um pouco menor, em cerca de 3,7 pontos (0,3 desvio padrão) na prova, o que é um efeito considerável. Com relação à oitava série os coeficientes de interação não se mostraram estatisticamente significantes para nenhuma das disciplinas.

Tabela 12- Resultado do Diferenças em Diferenças

Variáveis	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
EE_SP	-8,236*** (0,596)	-10,28*** (0,678)	-7,127*** (0,43)	-9,526*** (0,468)
Ano	9,263*** (0,596)	13,60*** (0,678)	9,492*** (0,430)	0,906* (0,468)
EE_SP*Ano	3,727*** (0,843)	6,357** (0,959)	-0,504 (0,609)	0,0554 (0,662)
Constante	184,4*** (0,422)	203,3*** (0,479)	237,3*** (0,304)	250,7*** (0,331)
Observações	5.976	5.948	10.300	10.440
R-quadrado	0,137	0,2	0,128	0,074

Erro Padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

7 ANÁLISE DE ROBUSTEZ

O ponto chave desta dissertação é encontrar um grupo que represente um bom contrafactual para as escolas estaduais paulista, ou seja, que represente qual seria a evolução do desempenho escolar na rede estadual paulista na ausência da política de bonificação. A estratégia apresentada anteriormente criou um grupo de comparação baseada no escore de propensão de acordo com certas características dos alunos, professores, diretores e da escola. Os resultados encontrados indicam um possível efeito positivo do programa de bonificação sobre as notas dos alunos da quarta série.

No entanto, algumas ressalvas podem ser feitas quanto à metodologia empregada. Primeiramente, pode não ter ficado muito claro qual grupo de escolas foi usado como comparação. Tal grupo foi formado por escolas da rede estadual das outras unidades de federação e também por escolas municipais paulistas. Além disso, foram consideradas apenas as informações disponíveis nos questionários da Prova Brasil. Desta forma, não foram levadas em consideração outros fatores que podem afetar o desempenho dos alunos, como questões de política regional, por exemplo.

Na busca por uma minimização desses problemas, bem como para checar a robustez do resultado encontrado, nas seções seguintes utilizaremos novamente o método de Diferenças em Diferenças, mas usando como contrafactual grupos diferentes.

7.1 – Diferenças em diferenças por município

O modelo do programa permite que dentro de um mesmo município existam escolas que foram contempladas pelo programa (rede estadual) e escolas que não tiveram o incentivo da bonificação (outras redes). Como um mesmo município está exposto aos mesmos fatores políticos, administrativos, econômicos e geográficos, que podem afetar a qualidade do ensino, a comparação entre escolas da mesma localidade podem mostrar melhor o efeito do programa, já que se estaria controlando por efeitos locais que afetam a proficiência escolar nas duas redes. Devido a pouca disponibilidade de dados para escolas particulares e federais, o grupo de comparação será formado apenas por escolas municipais.

Escolhido o contrafactual, lançaremos mão, novamente, da hipótese de tendência comum e, ao considerar fixos os efeitos dentro de cada município, usaremos então o

método de Diferenças em Diferenças já explicitado na seção 4. O resultado encontrado mostra o impacto médio do programa para as escolas do estado de São Paulo.

A equação 20 foi acrescida de variáveis *dummy* para cada município e pode ser lida pela seguinte expressão:

$$\text{Nota média}_{it} = \beta_0 + \beta_1 EE_SP_i + \beta_2 \text{Ano} + \beta_3 EE_SP_i * \text{Ano} + \text{Município}_1 + \text{Município}_2 + \dots + \text{Município}_j + \varepsilon_{it} \quad (23)$$

No qual j representa a quantidade de municípios que serão considerados.

O estado de São Paulo contém 645 municípios. Para esta estratégia foram utilizadas cidades paulistas com pelo menos três escolas de cada rede de ensino. Ao todo, foram utilizados dados de 83 municípios para a estimação da quarta série, o que corresponde a 6.376 escolas e 42 municípios para a oitava série, o que corresponde a 3.906 escolas.

Foram utilizadas variáveis *dummy* para cada município e os resultados da nossa variável de interesse ($EE_SP * \text{Ano}$) ficaram em linha com a estratégia adotada anteriormente, tanto em direção do efeito quanto em magnitude. O efeito estimado para a quarta série foi de 3,7 pontos (0,25 desvio padrão) para notas de Leitura e 7 pontos (0,46 desvio padrão) para nota de Matemática. Para a oitava série, os coeficientes foram negativos, mas não significante estatisticamente (Tabela 13).

Tabela 13- Resultado do Diferenças em Diferenças por Município

Variáveis	Quarta série		Oitava série	
	Matemática	Português	Matemática	Português
EE_SP	-5,040*** (0,571)	-3,156*** (0,500)	-3,904*** (0,653)	-2,585*** (0,665)
Ano	12,43*** (0,562)	9,011*** (0,488)	0,0316 (0,725)	9,364*** (0,738)
EE_SP*Ano	6,978*** (0,801)	3,707*** (0,696)	-0,170 (0,908)	-1,124 (0,924)
Dummy Municípios	Sim	Sim	Sim	Sim
Constante	198,0*** (0,413)	187,4*** (0,397)	266,2*** (4,330)	247,3*** (4,408)
Observações	6.376	6.376	3.906	3.906
R-quadrado	0,331	0,324	0,239	0,243

Erro Padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

7.2 – Diferenças em diferenças: Escolas estaduais x Escolas municipais paulistas

Nesta estratégia, foram usados novamente apenas dados das escolas do estado de São Paulo, mas dessa vez o grupo de comparação é formado por todas as escolas municipais paulistas. Os gráficos abaixo apresentam as proficiências de 2005 a 2009. Segundo a hipótese de tendência comum, as trajetórias dos dois grupos deveriam ser semelhantes antes da intervenção. Não é exatamente o que se verifica nos gráficos abaixo, mas cabe notar que, ao menos para a quarta série, as trajetórias dos tratados e controles antes do programa (de 2005 a 2007) eram mais parecidas do que a de 2007 a 2009 (depois). Já para a oitava série, as trajetórias pré-tratamento para o grupo de controle é bem diferente. Cabe lembrar ainda que as notas de 2005 para escola estadual paulista podem carregar algum ruído, pois foram calculadas a partir da base amostral do Saeb.

Figura 7- Notas de Português e Matemática da Quarta série - Estadual Paulista x Municipal Paulista (2005-2009)

Fonte: Inep

Figura 8- Notas de Português e Matemática da Oitava série - Estadual Paulista x Municipal Paulista (2005-2009)

Fonte: Inep

O resultado do Diferenças em Diferenças para a quarta série mostra que o efeito é significativo tanto para Português quanto para Matemática, em linha com os resultados das metodologias adotadas anteriormente. O efeito foi de 4,09 pontos (0,28 desvio padrão) para a prova de Português e 6,27 pontos (0,41 desvio padrão) para a prova de Matemática (Tabela 14). Para a oitava série, novamente, os coeficientes não se mostraram estatisticamente significantes.

Tabela 14- Diferenças em Diferenças: Estaduais paulistas x Municipais paulistas

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
EE_SP	-7,55*** (0,452)	-9,813*** (0,529)	-2,707*** (0,499)	-4,351*** (0,518)
Ano	8,511*** (0,377)	13,249*** (0,440)	8,934*** (0,598)	-0,141 (0,621)
EE_SP*Ano	4,09*** (0,645)	6,266*** (0,753)	0,005 (0,694)	0,864 (0,719)
Constante	184,039*** (0,270)	203,452*** (0,316)	233,862*** (0,433)	246,392*** (0,449)
Observações	11.229	11.229	9.493	9.493
R-quadrado	0,113	0,17	0,090	0,013

Erro Padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

7.3 – Diferenças em diferenças: Escolas estaduais paulistas x Resto do Brasil

Nesta seção, testaremos os efeitos do programa de bonificação usando como contrafactual as escolas das outras unidades de federação do Brasil. Serão testados dois grupos de controle. O primeiro é formado por todas as escolas (todas as redes) e o segundo é formado apenas por escolas de administração dos governos estaduais.

Os gráficos a seguir mostram as trajetórias das notas de 2005 a 2009. Para as notas de Língua Portuguesa da quarta série, a média de notas do restante das escolas brasileiras apresentou trajetória bastante semelhante no período pré-tratamento, algo bem próximo do que é esperado de um bom grupo de controle para o uso do DD. Já para a nota de Matemática da quarta série, um bom comportamento foi observado pelas escolas estaduais. Para a oitava série, as trajetórias de notas de 2005 a 2007 foram muito diferentes entre os grupos tratamento e controle.

Figura 9- Notas de Português e Matemática da Quarta série - Estadual Paulista x Restante do Brasil

Fonte: Inep

Figura 10- Notas de Português e Matemática da Oitava série - Estadual Paulista x Restante do Brasil

Fonte: Inep

A estimativa para o efeito do programa é mostrada na Tabela 15. O efeito estimado para a quarta série foi positiva e estatisticamente significativa em quase 3 pontos (0,2 desvio padrão) para Língua Portuguesa e 7,7 pontos (0,5 desvio padrão) para Matemática. Já os efeitos para a oitava série foram negativos e significantes. O efeito para Português foi de -2,15 pontos (0,1 desvio padrão) e para Matemática negativo em quase 2 pontos (0,1 desvio padrão).

Tabela 15- Diferenças em Diferenças: Estaduais paulistas x Resto do Brasil

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
EE_SP	6,908*** (0,439)	6,853*** (0,500)	4,948*** (0,324)	4,126*** (0,351)
Ano	9,693*** (0,154)	11,84*** (0,175)	11,09*** (0,170)	2,708*** (0,184)
EE_SP*Ano	2,908*** (0,633)	7,674*** (0,720)	-2,152*** (0,456)	-1,985*** (0,495)
Constante	169,6*** (0,108)	186,8*** (0,122)	226,2*** (0,120)	237,9*** (0,130)
Observações	64.640	64.640	50.920	50.920
R-quadrado	0,073	0,087	0,089	0,007

Erro padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

Ao utilizar apenas as escolas estaduais, os efeitos foram semelhantes (Tabela 16). O impacto estimado é discretamente menor do que o encontrado ao utilizar todas as redes do Brasil: 2,8 pontos (0,2 desvio padrão) para Português e 7 pontos para Matemática (0,46 desvio padrão). Para a oitava série, novamente, os efeitos foram negativos e significantes: -2,3 pontos para Português (0,2 desvio padrão) e -2,1 pontos (0,1 desvio padrão) para Matemática.

Tabela 16- Diferenças em Diferenças: Estaduais paulistas x Resto do Brasil (Estaduais)

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
EE_SP	3,471*** (0,456)	3,539*** (0,514)	3,625*** (0,320)	2,488*** (0,351)
Ano	9,815*** (0,267)	12,52*** (0,301)	11,27*** (0,206)	2,854*** (0,227)
EE_SP*Ano	2,786*** (0,659)	6,996*** (0,742)	-2,335*** (0,451)	-2,132*** (0,495)
Constante	173,0*** (0,183)	190,1*** (0,207)	227,5*** (0,145)	239,6*** (0,159)
Observações	23.360	23.360	33.695	33.695
R-quadrado	0,079	0,111	0,097	0,006

Erro Padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

7.4 – Diferenças em diferenças: Escolas estaduais paulistas x Sudeste

Por fim, foi usado como contrafactual o grupo formado por escolas da região Sudeste (exceto do estado de São Paulo). É plausível supor que os estados mais próximos tenham características mais semelhantes.

Serão utilizados nesta seção quatro grupos diferentes como controle: (i) escolas de todas as redes do Sudeste (exceto São Paulo); (ii) escolas estaduais do Sudeste (exceto São Paulo); (iii) escolas de todas as redes de ensino do Rio de Janeiro e Espírito Santo e (iv) escolas estaduais do Rio de Janeiro e Espírito Santo.

A razão para a criação dos dois últimos grupos foi o comportamento muito distinto das notas de Minas Gerais da quarta série frente aos outros estados do Sudeste (Figuras 1A e 2A). A escolha foi feita baseada no comportamento das notas da quarta série, uma vez que as estimações para a oitava série até então realizadas mostraram-se significantes em poucos casos. Embora as notas para o total de escolas fluminenses também apresentarem comportamento diferente dos demais grupos, mantivemos as notas do estado devido ao bom comportamento das notas das escolas estaduais.

7.4.1 – Diferenças em diferenças: Escolas estaduais paulistas x Sudeste

Desta vez, os resultados para a quarta série mostraram-se bastante distintos do que vinha ocorrendo até agora. Quando feita a comparação contra as escolas dos três estados (MG, RJ e ES), os coeficientes que mostram o efeito do programa foram negativos e significantes tanto para o conjunto de todas as redes de ensino quanto para o conjunto contendo apenas escolas estaduais.

O gráfico que ilustra a trajetória das notas pré-tratamento mostra uma diferença considerável entre tratamento e controle, sobretudo para os grupos que eram compostos por escolas de todas as redes de ensino. Ainda que tenha que considerar algum erro na nota de 2005, a trajetória é muito diferente do que seria esperado para um grupo de controle ideal.

Figura 11- Notas de Português e Matemática da Quarta série - Estadual Paulista x Restante do Sudeste

Fonte: Inep

Figura 12- Notas de Português e Matemática da Oitava série - Estadual Paulista x Restante do Sudeste

Fonte: Inep

O efeito estimado para Português da quarta série foi significativo e negativo em 2,5 pontos (0,2 desvio padrão) enquanto o efeito para Matemática foi praticamente zero, porém, não significativo. Para a oitava série os efeitos continuaram a ser negativos e significantes em 3,8 pontos para Português e 4,4 para Matemática, o que representa 0,25 desvio padrão para ambas disciplinas.

Tabela 17- Resultado do Diferenças em Diferenças: Estaduais paulistas x Escolas do restante do Sudeste

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
EE_SP	-1,009** (0,439)	-2,091*** (0,498)	-1,478*** (0,349)	-3,162*** (0,383)
Ano	15,11*** (0,290)	19,13*** (0,329)	12,76*** (0,309)	5,124*** (0,340)
EE_SP*Ano	-2,512*** (0,632)	0,381 (0,717)	-3,822*** (0,493)	-4,402*** (0,542)
Constante	177,5*** (0,202)	195,7*** (0,229)	232,6*** (0,217)	245,2*** (0,239)
Observações	18.160	18.160	17.951	17.951
R-quadrado	0,153	0,193	0,119	0,033

Erro Padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

Resultados semelhantes a estes foram encontrados quando utilizadas as escolas estaduais. Desta vez, os efeitos estimados para a oitava série foram ainda maiores. O impacto foi negativo em 4,7 pontos para Português (0.35 desvio padrão) e 0,27 desvio padrão para Matemática (5,2 pontos), com significância estatística.

Tabela 18- Resultado do Diferenças em Diferenças: Estaduais paulistas x Escolas estaduais do restante do Sudeste

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
EE_SP	-3,492*** (0,512)	-4,616*** (0,582)	-0,640* (0,376)	-3,163*** (0,411)
Ano	14,09*** (0,488)	19,59*** (0,554)	13,64*** (0,382)	5,981*** (0,417)
EE_SP*Ano	-1,490** (0,742)	-0,0773 (0,843)	-4,696*** (0,534)	-5,258*** (0,583)
Constante	180,0*** (0,334)	198,3*** (0,379)	231,8*** (0,267)	245,2*** (0,292)
Observações	8.896	8.896	13.822	13.822
R-quadrado	0,141	0,206	0,124	0,041

Erro Padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

Os resultados negativos encontrados no Diferenças em Diferenças neste grupo se devem, em muito, às notas bastante divergentes de Minas Gerais. Embora o estado tenha sido considerado nas estimações para o restante do Brasil, o peso era bem menor. Para este grupo, no entanto, o peso do estado é de mais de 70% (ponderado pelo número de escolas), o que ajudou a construir um grupo de comparação com comportamento muito distinto do grupo de tratamento. Na próxima seção, excluiremos as notas de Minas Gerais e usaremos como contrafactual apenas as escolas do Rio de Janeiro e do Espírito Santo.

7.4.2 – Diferenças em diferenças: Escolas estaduais paulistas x Rio de Janeiro e Espírito Santo

Ao excluir as notas do estado mineiro, os resultados para a quarta série voltam a ficar parecidos com os já encontrados, ou seja, efeito positivo e significativo do programa de bonificação sobre a proficiência dos alunos da quarta série. A análise da trajetória de notas mostrou comportamento mais semelhante dos grupos de controle para a quarta série, principalmente naqueles grupos formados apenas por escolas da rede estadual. Já para a oitava série, as trajetórias dos controles entre 2005 e 2007 foram bem diferentes do tratamento.

Figura 13- Notas de Português e Matemática da Quarta série - Estadual Paulista x RJ e ES
Fonte: Inep

Figura 14- Notas de Português e Matemática da Oitava série - Estadual Paulista x RJ e ES
Fonte: Inep

A tabela 19 traz os resultados das regressões em que o coeficiente de interesse mostrou aumento de 2,7 pontos (0,2 desvio padrão) em Língua Portuguesa e de 7,2 (0,5 desvio padrão) pontos para Matemática. Já para a oitava série os efeitos foram negativos em quase 3 pontos (0,2 desvio padrão) para Português e de -4,3 pontos (0,25 desvio padrão) para Matemática. Todos se mostraram significativos.

Tabela 19- Resultado do Diferenças em Diferenças: Estaduais paulistas x Rio de Janeiro e Espírito Santo

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
EE_SP	9,882*** (0,396)	12,28*** (0,435)	11,85*** (0,459)	5,037*** (0,474)
Ano	-0,497 (0,451)	0,181 (0,495)	2,319*** (0,415)	4,137*** (0,429)
EE_SP*Ano	2,720*** (0,648)	7,232*** (0,711)	-2,911*** (0,586)	-4,314*** (0,605)
Constante	177,0*** (0,277)	193,5*** (0,305)	228,8*** (0,325)	237,9*** (0,335)
Observações	10.235	10.235	11.521	11.521
R-quadrado	0,107	0,170	0,100	0,014

Erro Padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

Quando comparados apenas a rede de ensino estadual, o impacto estimado para a oitava série é negativo em -3,5 pontos (0,2 desvio padrão) para Português e -5,7 pontos (0,37 desvio padrão) para Matemática. Para a quarta série o efeito foi o maior encontrado até agora, 6 pontos (0,4 desvio padrão) para Português e quase 10 pontos (0,7 desvio padrão) para Matemática.

Tabela 20- Resultado do Diferenças em Diferenças: Estaduais paulistas x Estaduais Rio de Janeiro e Espírito Santo

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
EE_SP	6,555*** (0,846)	9,634*** (0,931)	12,45*** (0,646)	6,473*** (0,661)
Ano	2,654*** (0,671)	3,281*** (0,739)	7,022*** (0,519)	8,576*** (0,531)
EE_SP*Ano	6,046*** (0,981)	9,881*** (1,080)	-3,516*** (0,735)	-5,751*** (0,752)
Constante	173,8*** (0,576)	190,4*** (0,634)	224,1*** (0,456)	233,5*** (0,467)
Observações	5.166	5.166	9.133	9.133
R-quadrado	0,139	0,238	0,119	0,035

Erro Padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

Com exceção do grupo de comparação da seção 7.4.1 formado pelas escolas de Minas Gerais, Rio de Janeiro e Espírito Santo, as demais estimações mostraram efeito positivo e significativo para as notas da quarta série e negativo para as notas da oitava série. Ainda que a análise gráfica deva ser feita com alguma ressalva (devido a possíveis ruídos nos dados de 2005), ficou evidente a dificuldade de encontrar um bom grupo de controle para a oitava série.

7.5 Diferenças triplas (DDD)

O método de Diferenças em Diferenças exige que as trajetórias das notas das escolas estaduais paulistas sejam semelhantes às dos grupos de controle. Todavia, ainda podem haver fatores (não relacionados ao programa de bonificação) que afetem as notas dos grupos de forma desigual como, por exemplo, as diferenças inerentes entre as administrações estaduais e municipais. Para eliminar essas possíveis diferenças, podemos combinar mais de um grupo de controle por meio das diferenças triplas (DDD).

Nesta seção, comparamos a diferença entre notas das escolas estaduais e municipais paulistas com a diferença das notas das escolas estaduais e municipais fora do estado de São Paulo que será estimada pela equação 20.

7.5.1 – Diferenças triplas (DDD): Estadual SP vs Municipal SP x Estadual Brasil vs Municipal BR

Quando comparado à diferença de estaduais e municipais dos demais estados do Brasil, o efeito estimado para o programa (ANO_SP_EE) foi de 5,4 pontos (0,4 desvio padrão) para Matemática e 4 pontos (0,3 desvio padrão) para Português. Já para as notas de oitava série, mais uma vez, os efeitos não foram significantes, conforme mostrado na tabela a seguir.

Tabela 21- Diferenças triplas: Estadual SP vs Municipal SP x Estadual Brasil vs Municipal BR

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
SP	16,16*** (0,337)	18,307*** (0,386)	9,758*** (0,553)	11,08*** (0,599)
EE	5,14*** (0,224)	4,955*** (0,257)	3,426*** (0,244)	4,241*** (0,265)
SP_EE	-12,689*** (0,568)	-14,768*** (0,649)	-6,133*** (0,647)	-8,592*** (0,701)
Ano	9,793*** (0,184)	11,685*** (0,21)	10,929*** (0,269)	2,627*** (0,292)
Ano_SP	-1,282*** (0,471)	1,564*** (0,539)	-1,995*** (0,767)	-2,769*** (0,83)
Ano_EE	0,022 (0,324)	0,834 (0,371)	0,344 (0,346)	0,227 (0,374)
Ano_SP_EE	4,068*** (0,811)	5,432*** (0,927)	-0,34 (0,901)	0,637 (0,976)
Constante	167,879*** (0,129)	185,145*** (0,148)	224,104*** (0,191)	235,313*** (0,207)
R-quadrado	0,1238	0,1445	0,1	0,022
Observações	72.036	72.036	53.351	53.351

Erro Padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

7.5.2 - Diferenças triplas (DDD): Estadual SP vs Municipal SP x Estadual SE vs Municipal SE

Quando usados os dados para a região Sudeste, o estimador de diferenças triplas mostrou novamente resultados positivos e significativos para a quarta série nas duas disciplinas (5,7 para Matemática e 5,8 para Português). Para a oitava série os efeitos foram negativos, sendo significativo apenas para Português (1,4 para Matemática e 2,3 para Português).

Tabela 22- Diferenças triplas: Estadual SP vs Municipal SP x Estadual SE vs Municipal SE

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
SP	7,969*** (0,381)	9,176*** (0,439)	-0,168 (0,593)	1,192* (0,651)
EE	3,911*** (0,416)	3,979*** (0,478)	-2,236*** (0,45)	0,003 (0,494)
SP_EE	-11,461*** (0,635)	-13,792*** (0,731)	-0,471 (0,709)	-4,354*** (0,779)
ANO	15,787*** (0,357)	19,046*** (0,411)	11,307*** (0,504)	3,746*** (0,553)
ANO_SP	-7,276*** (0,536)	-5,797*** (0,616)	-2,373*** (0,828)	-3,888*** (0,909)
ANO_EE	-1,696*** (0,601)	0,546*** (0,692)	2,328*** (0,639)	2,234*** (0,702)
ANO_SP_EE	5,786*** (0,911)	5,719*** (1,048)	-2,323** (0,994)	-1,371 (1,092)
Constante	176,07*** (0,251)	194,276*** (0,289)	234,03*** (0,356)	245,201*** (0,39)
R-quadrado	0,1411	0,1834	0,114	0,030
Observações	25.556	25.556	20.382	20.382

Erro padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

7.5.3 - Diferenças triplas (DDD): Estadual SP vs Municipal SP x Estadual RJ+ES vs Municipal RJ+ES

Na comparação com as escolas capixabas e fluminenses, o efeito estimado foi maior tanto para Português quanto para Matemática. Em cerca de 9,5 pontos para Matemática e 8,1 pontos para Português, 0,5 e 0,6 desvio padrão, respectivamente. E mostraram significância estatística. Já para as notas de oitava série, mais uma vez, os efeitos foram negativos, mas não significantes.

Tabela 23- Diferenças triplas: Estadual SP vs Municipal SP x Estadual RJ+ES vs Municipal RJ+ES

	Quarta série		Oitava série	
	Português	Matemática	Português	Matemática
SP	6.170*** (0.420)	9.125*** (0.48)	0.911 (0.635)	4,606*** (0,664)
EE	-4.033*** (0.686)	-3.969*** (0.785)	-8.818*** (0.653)	-8,321*** (0,683)
SP_EE	-3.516*** (0.822)	-5.844*** (0.94)	6.111*** (0.836)	3,97*** (0,874)
ANO	10.588*** (0.455)	12.825*** (0.521)	11.261*** (0.63)	3,736*** (0,659)
ANO_SP	-2.077*** (0.591)	0.424 (0.676)	-2.326*** (0.887)	-3,877*** (0,928)
ANO_EE	-4.033*** (1.000)	-3.192*** (1.144)	1.194 (0.924)	2,737*** (0,967)
ANO_SP_EE	8.123*** (1.190)	9.457*** (1.362)	-1.189 (1.174)	-1,874 (1,228)
Constante	177.869*** (0.321)	194.327*** (0.367)	232.951*** (0.446)	241,786*** (0,467)
R-quadrado	0,1242	0,1855	0,1186	0,037
Observações	17.631	17.631	13.952	13.952

Erro Padrão entre parênteses

*** p<0,01, ** p<0,05, * p<0,1

Fonte: Elaboração própria

8 CONSIDERAÇÕES FINAIS

As reformas educacionais têm como um dos temas centrais programas de responsabilização dos agentes escolares pela melhoria do desempenho dos alunos. Dentro deste contexto, este trabalho procurou investigar se o programa de bonificação de professores implementado em 2008 no estado de São Paulo teve impacto sobre a proficiência dos estudantes.

Boa parte da literatura internacional indica que programas de premiação de docentes levam a uma melhora do desempenho dos alunos, porém, há estudos que não mostram impacto sobre as notas, ou até mesmo, que mostram efeitos negativos, como o de Fryer (2011).

Para este trabalho, utilizou-se como método principal o *Propensity Score Matching* (PSM) para buscar grupos comparáveis mais homogêneos e, depois, uma estimação de Diferenças em Diferenças (DD) para eliminar efeitos fixos não observáveis. Procedimentos de checagem mostraram boa qualidade do pareamento.

Foram encontrados indícios de impacto positivo do programa de incentivo de professores apenas nas notas da quarta série. O efeito para Matemática foi de cerca de 6,4 pontos e o de Língua Portuguesa, em cerca de 3,7 pontos na escala SAEB. Para a oitava série não houve evidência de efeito significativo.

A fim de checar a robustez dos resultados, foram usados os métodos de Diferenças em Diferenças (DD) e Diferenças triplas (DDD) para outros grupos de controle. Os resultados se mostraram robustos a inclusão de *dummies* municipais, ou seja, ao se restringir o grupo de controle as escolas de dentro da mesma cidade e ao usar como grupo de controle todas as escolas municipais do estado. Também foram testadas como controle escolas do restante do Brasil, do restante da região Sudeste e depois do Rio de Janeiro e Espírito Santo. Com exceção da comparação do restante da região Sudeste, todos os resultados foram robustos para a quarta série. Os resultados também foram robustos quando utilizado o método de diferenças triplas.

Embora não seja o escopo deste trabalho, levantamos algumas hipóteses que poderiam ajudar a explicar as diferenças de notas entre as disciplinas e entre os ciclos do fundamental.

Com relação à diferença entre Português e Matemática, devemos considerar a diferença de aprendizado nas duas disciplinas. Talvez o esforço do professor dentro de sala de aula seja mais efetivo para o aprendizado de Matemática do que de Língua Portuguesa, uma vez que, para esta, a leitura fora de sala de aula é bastante importante. Outra hipótese é que o aumento em Matemática pode ter sido potencializado pela melhoria da Leitura, uma vez que para a solução dos problemas de Matemática é requerido interpretação do problema.

No que diz respeito às diferenças entre quarta e oitava séries, um ponto que salta aos olhos é o número de escolas do primeiro ciclo consideravelmente menor do que o de oitava série. O processo de descentralização do ensino ocorrido desde meados de 1990 levou a migração da administração estadual para municipal de algumas escolas do primeiro ciclo do ensino fundamental, o que explica o menor número de escolas tratadas para a quarta série. Isso pode ter alguma relação com o fato do programa funcionar melhor para este grupo de estudantes.

Também devemos levar em consideração que a qualidade do ensino está muito aquém do desejado, e muitos alunos passam de ano sem ter alcançado um nível desejado de aprendizado para sua série. Isso acumula um déficit de aprendizado ao passar dos anos e é provável que o déficit em relação ao desejável de crianças ao chegar na oitava série seja maior do que na quarta. Embora o esforço do professor possa ser maior, é mais difícil conseguir uma melhora desses alunos, principalmente quando os assuntos abordados vão ficando mais complexos.

Este trabalho apresentou alguns resultados sobre o efeito do programa de bonificação de professores sobre a proficiência, mas ainda há um espaço muito grande para discussões sobre o tema. Estudos posteriores podem ajudar a elucidar a diferença dos resultados para as diferentes séries e, quem sabe, ajudar a propor alguma forma de incentivo que seja mais adequado ao grupo de professores responsáveis pelo desenvolvimento do aprendizado dos alunos de quinta a oitava séries.

Referências Bibliográficas

ABADIE,A.;DIAMOND,A.;HAINMUELLER,J. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. **National Bureau of Economic Research**. v. 12831, 2007.

_____. DRUKKER, D; IMBENS, G. Implementing matching estimators for average treatment effects in Stata. **The Stata Journal**. College Station, v.4, n.3, p.290-311, 2001.

ANDRADE, E. “*School Accountability*” no Brasil: experiências e dificuldades. **Revista de Economia Política**, v. 28, n.3, p. 443-453, 2008.

_____. Alternativa de política educacional para o Brasil: *School accountability*. **Revista de Economia Política**, v. 29, n.4, p. 454-472, 2009.

ATKINSON, A.; BURGESS,S.; CROXSON.B; GREGG,P.;PROPPER,C.;SLATER, H.;WILSON,D. Evaluating the impact of performance-related pay for teachers in England. **Labour Economics** v.16, n.3, p.251–26, 2009.

BROOKE,N. O futuro das políticas de responsabilização educacional no Brasil. **Cadernos de Pesquisa**, v.36, n.128, p.377-401, 2006.

CARNOY, M.; LOEB, S. Does External Accountability Affect Student Outcomes? A Cross-State Analysis. **Educational Evaluation and Policy Analysis**, v.24, n.24, p. 305-331,2002.

DOLTON,P.; MARCENARO-GUTIERREZ,O.D. If you pay peanuts do you get monkeys? A cross country analysis of teacher pay and pupil performance, **Economic Policy** January 2011 pp. 5–55, 2001.

FERNANDES,R; GREMAUD,A. Qualidade da educação: avaliação, indicadores e metas. In: VELOSO,F. (Org), **Educação Básica no Brasil**. Rio de Janeiro: Elsevier, 2009.

FERRAZ,C. Sistemas Educacionais Baseados em Desempenho, Metas de Qualidade e a Remuneração de Professores: os Casos de Pernambuco e São Paulo. In: VELOSO,F. (Org), **Educação Básica no Brasil**. Rio de Janeiro: Elsevier, 2009.

FIGLIO,D.N.;GETZLER,L.S. Accountability, Ability and Disability: Gaming the system. **National Bureau of Economic Research**, v.0307, 2002.

_____. KENNY, L.W. Individual teacher incentives and student performance. *Journal of Public Economics*,v.91, p.901–914, 2007.

FRYER, R. Teacher incentives and student achievement: Evidence from new york city public schools. **National Bureau of Economic Research**, v. 16850, 2011.

GLAZERMAN, S.; MCKIE, A.; CAREY, N. An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year One Impact Report, **Mathematica Policy Research**, 2009

GLEWWE,P.; ILIAS,N.;KREMER,M. Teacher Incentives. *National Bureau of Economic Research*, v.9671, 2003.

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

_____; RAYMOND,M; Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, v.24, n.2, p.297-327, 2005.

HOLMSTROM,B.;MILGROM,P. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization*, v. 7, p. 24-52, 1991.

JACOB, B.A. Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. **Journal of Public Economics**, v.89, n.5-6, p.761-796, 2005.

LAUEN,D.D.. The Effects of School Wide Bonuses on Student Achievement: Regression Discontinuity Evidence from North Carolina. **Society for Research on Educational Effectiveness Conference**, 2011

LAVY,V. Performance Pay and Teachers' Effort, Productivity, and Grading Ethics. **American Economic Review**, v.99.n.5.p.1979–2011, 2009

MENEZES-FILHO, N e TAVARES, P. Noise in Education Pay for Performance Programs: Evidence using Independent Measures of School Outcomes. **Meeting of the Brazilian Econometric Society**, 2011

MINISTÉRIO DA EDUCAÇÃO - INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **SAEB – 2005 – Primeiros Resultados: Médias de desempenho do SAEB/2005 em perspectiva comparada**. 2007 disponível em: http://www.inep.gov.br/download/saeb/2005/SAEB1995_2005.pdf
ROSENBAUM, P.R., RUBIN, D.B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, Great Britain, v.70, n.1, p.41-55, 1983.

MURALIDHARAN, K; SUNDARARAMAN, V. Teacher performance pay: Experimental evidence from India. **National Bureau of Economic Research**, v. 15323, 2009

SÃO PAULO. Lei Complementar 1078 de 17 de dezembro de 2008. Institui Bonificação por Resultados - BR, no âmbito da Secretaria da Educação, e dá providências correlatas. **Diário Oficial [do] Estado de São Paulo** no. 239, volume 118 de 18 de dezembro de 2008.

SÃO PAULO (ESTADO) SECRETARIA DA EDUCAÇÃO. **Programa de Qualidade da Escola**. 2009a. Nota Técnica disponível em: <http://idesp.edunet.sp.gov.br/Arquivos/NotaTecnicaPQE2008.pdf>

_____. **Matrizes de referência para a avaliação Saresp**. 2009b. Disponível em: http://saresp.fde.sp.gov.br/2009/pdf/Saresp2008_MatrizRefAvaliacao_DocBasico_Completo.pdf

_____. **Programa de Qualidade da Escola**. 2010. Nota Técnica disponível em: http://idesp.edunet.sp.gov.br/Arquivos/Nota_tecnica_2009.pdf

SPRINGER, M.G.; HAMILTON, L; MCCAFFREY, D.; BALLOU, D.F.; LE, V; PEPPER, M; LOCKWOOD, J. R.; STECHER, B.M. Teacher Pay for Performance Experimental Evidence from the Project on Incentives in Teaching. **National Center on Performance Incentives**, 2010

TODOS PELA EDUCAÇÃO. **Números do Brasil**. Disponível em: <http://www.todospelaeducacao.org.br/educacao-no-brasil/numeros-do-brasil/brasil>. Acesso em: 08 março 2011.

VELOSO,F. Experiências de Reforma Educacional nas Últimas Duas Décadas: o que Podemos Aprender?. In: VELOSO,F. (Org), **Educação Básica no Brasil**. Rio de Janeiro: Elsevier, 2009.

VIGDOR, J.L. **Teacher Salary Bonuses in North Carolina**. National Center for analysis of longitudinal data in education research. Working Paper n.15, 2008.

Apêndice

Tabela 1A- Matching por vizinho mais próximo (NN 2) - quarta e oitava série

	Amostra	Tratados	Controles	Diferença	Desvio Padrão	T-stat
Quarta série						
Português 2007	Antes do Matching	176,65	171,73	4,92	0,45	11,02
	Depois do Matching	176,64	184,65	-8,01	0,61	-13,24
Matemática 2007	Antes do Matching	193,70	189,21	4,49	0,49	9,15
	Depois do Matching	193,64	203,02	-9,38	0,67	-13,97
Oitava série						
Português 2007	Antes do Matching	231,26	227,16	4,10	0,32	12,96
	Depois do Matching	231,24	237,41	-6,17	0,46	-13,36
Matemática 2007	Antes do Matching	242,13	239,04	3,09	0,35	8,77
	Depois do Matching	242,12	250,71	-8,58	0,50	-17,17

Nearest Neighbor (2) com Caliper (0,002)

Fonte: Elaboração própria

Tabela 2A- Matching por vizinho mais próximo (NN 5) - quarta e oitava série

	Amostra	Tratados	Controles	Diferença	Desvio Padrão	T-stat
Quarta série						
Português 2007	Antes do Matching	176,65	171,73	4,92	0,45	11,02
	Depois do Matching	176,64	184,83	-8,19	0,55	-14,92
Matemática 2007	Antes do Matching	193,70	189,21	4,49	0,49	9,15
	Depois do Matching	193,70	203,11	-9,41	0,60	-15,70
Oitava série						
Português 2007	Antes do Matching	231,26	227,16	4,10	0,32	12,96
	Depois do Matching	231,27	237,38	-6,10	0,41	-14,93
Matemática 2007	Antes do Matching	242,13	239,04	3,09	0,35	8,77
	Depois do Matching	242,13	250,88	-8,75	0,45	-19,46

Nearest Neighbor (5) com Caliper (0,005)

Fonte: Elaboração própria

Figura 1A: Notas de Português da Quarta série - Estadual Paulista x Estados do Sudeste
Fonte: Inep

Figura 2A: Notas de Português da Quarta série - Estadual Paulista x Estaduais dos estados do Sudeste
Fonte: Inep