

INSPER
GRADUAÇÃO EM CIÊNCIAS ECONÔMICAS

BRUNA DOMINICI CRICCI

**A POLÍTICA DE BÔNUS DO ESTADO DE SÃO PAULO MUDOU O
DESEMPENHO DOS ALUNOS NA PROVA DO SARESP E NA PROVA
BRASIL?**

SÃO PAULO
2018

BRUNA DOMINICI CRICCI

**A POLÍTICA DE BÔNUS DO ESTADO DE SÃO PAULO MUDOU O
DESEMPENHO DOS ALUNOS NA PROVA DO SARESP E NA PROVA
BRASIL?**

Monografia apresentada ao curso de Ciências
Econômicas, como requisito parcial para a obtenção
do Grau de Bacharel do Insper Instituto de Ensino e
Pesquisa

Orientador: Prof. Dr. Naercio Aquino Menezes Filho

SÃO PAULO
2018

CRICCI, Bruna Dominici

A política de bônus do Estado de São Paulo mudou o desempenho dos alunos na prova do SARESP e na Prova Brasil?

Bruna Cricci. – São Paulo, 2018. 36 págs.

Monografia (Bacharel em Economia) – Insper, 2018.

Orientador: Prof. Dr. Naercio Aquino Menezes Filho

1. Economia 2. Educação 3. Diferenças em diferenças 4. Política de bônus

BRUNA DOMINICI CRICCI

**A POLÍTICA DE BÔNUS DO ESTADO DE SÃO PAULO MUDOU O
DESEMPENHO DOS ALUNOS NA PROVA DO SARESP E NA PROVA
BRASIL?**

Monografia apresentada ao curso de Ciências
Econômicas, como requisito parcial para a obtenção
do Grau de Bacharel do Insper Instituto de Ensino e
Pesquisa

Data de Aprovação: ___/___/___

BANCA EXAMINADORA

Prof. Dr. Naercio Aquino Menezes Filho
Professor Doutor
Insper – Instituto de Ensino e Pesquisa

Prof. Dr. Rodrigo Moita
Professor Doutor
Insper – Instituto de Ensino e Pesquisa

Aos meus pais,
meu irmão,
e meus avós

AGRADECIMENTOS

Agradeço, primeiramente, à minha família por sempre me apoiar durante esses anos da graduação, por me incentivarem tanto a seguir o caminho que acredito ser o que me fará mais feliz e que me trará grande satisfação profissional e pessoal, a vida acadêmica. Também por serem os meus maiores exemplos e por me darem tanto amor.

Agradeço ao Professor Doutor Naercio Aquino Menezes Filho pela paciência, qualidade de orientador, transmissão de conhecimento e por influenciar e motivar na incessante busca pela melhora do sistema educacional brasileiro, bem como, no melhor funcionamento das políticas públicas do país.

Agradeço também ao professor Paulo Marques por seus ensinamentos e pela ajuda ao usar o programa R para analisar os dados dessa monografia.

Sou imensamente grata ao INSPER e aos seus colaboradores pela sua excelência educacional, pela infraestrutura sem igual e por sempre colaborar para um debate mais embasado em dados e com menos ideologia.

Ao Vitor Cavalcante, recém-economista que mais admiro, que me apoia e me incentiva em todas as minhas decisões e que contribui todos os dias para ser uma pessoa e uma economista melhor.

Agradeço aos meus colegas que contribuíram para o meu desenvolvimento pessoal, tal qual para meu crescimento político e econômico e que tornaram essa jornada mais leve e feliz: Luis Felipe Bento, Victor Rodrigues, Marcela Biagi, Paula Tibau e Daniella Atallah.

“Education is the most powerful weapon,
which you can use to change the world”

Nelson Mandela

RESUMO

A educação é um fator relevante para o desenvolvimento econômico e social de um país, bem como para a redução de sua desigualdade. O Brasil possui como um dos seus maiores desafios a melhora do sistema educacional, com isso políticas públicas vêm sendo implementadas em busca desse desenvolvimento. O presente trabalho tem por objetivo investigar a política de bonificação para professores do Estado de São Paulo, criada em 2008, de modo a compreender se os alunos paulistas têm um desempenho melhor na prova que vale bônus para o professor, a Prova do SARESP (“high stake exam”), quando comparado com a Prova Brasil (“low stake exam”). Os métodos utilizados foram “Diferenças em Diferenças”, Mínimos Quadrados Ordinários e Regressão Quantílica. Os resultados permitem concluir que as escolas com pior desempenho nas provas tiveram 7,76 pontos, na escala SAEB, na nota do SARESP a mais do que o crescimento das respectivas notas na Prova Brasil; nas escolas da mediana o desempenho também é maior no “high stake exam”, mas a discrepância entre as notas é considerada pequena; já nas escolas com as melhores notas, o desempenho foi maior na Prova Brasil após a introdução do bônus.

Palavras-chave: 1. Economia. 2. Educação. 3. Diferenças em diferenças. 4. Política de bônus 5. Regressão Quatílica

ABSTRACT

Education is a relevant factor for the economic and social development of a country, as well as for reducing its inequality. As one of its greatest challenges, Brazil faces the need of educational system improvement, which is being mostly addressed by the implementation of public policies. The present study aims to investigate São Paulo's state bonus policy for teachers, created in 2008, in order to understand whether the students perform better in the test that is worth bonuses for the teacher, SARESP test (high stake exam), when compared to Prova Brasil (low stake exam). The methods used were "Differences in Differences", Ordinary Least Squares and Quantile Regression. The results show that schools with the worst performance in the tests had 7.76 points, at SAEB scale, in the SARESP scores more than the growth of the respective marks in Prova Brasil; for the median schools the performance was also higher in the high stake exam, although the discrepancy between grades is small; and for the schools with the best grades, the performance was greater in Prova Brasil after the bonus implementation.

Keywords: 1. Economy. 2. Education. 3. Differences in differences. 4. Bonus Policy. 5. Quantile Regression

LISTA DE GRÁFICOS

Gráfico 1 - Evolução da Escolaridade da População Adulta e da Produtividade do Trabalho: 1950-2010.....	14
Gráfico 2 - Histogramas das notas médias das escolas dividido por prova e por ano.....	24
Gráfico 3 - Distribuição de notas por escola, dividido por conteúdo.....	24
Gráfico 4 - Variação das notas entre 2007 e 2009, por escola.....	26
Gráfico 5. Variação das notas médias das escolas, por prova e ano.....	27
Gráfico 6 - Distribuição dos alunos por escola, dividido por conteúdo, por prova e por ano.....	28

LISTA DE TABELAS

Tabela 1 - Estatística descritiva.....	23
Tabela 2 - Correlação do número médio de alunos entre as provas e os anos.....	28
Tabela 3 - Correlação da nota média em cada matéria nas respectivas provas, por ano..	29
Tabela 4 - Diferença de resultados nas provas do SARESP e da Prova Brasil.....	30
Tabela 5 - Diferença de resultados nas provas do SARESP e da Prova Brasil nos diferentes quantis.....	31

SUMÁRIO

1. INTRODUÇÃO.....	13
2. REVISÃO BIBLIOGRÁFICA.....	16
3. METODOLOGIA.....	19
4. DADOS.....	23
5. RESULTADOS.....	30
6. CONSIDERAÇÕES FINAIS.....	33
BIBLIOGRAFIA.....	35

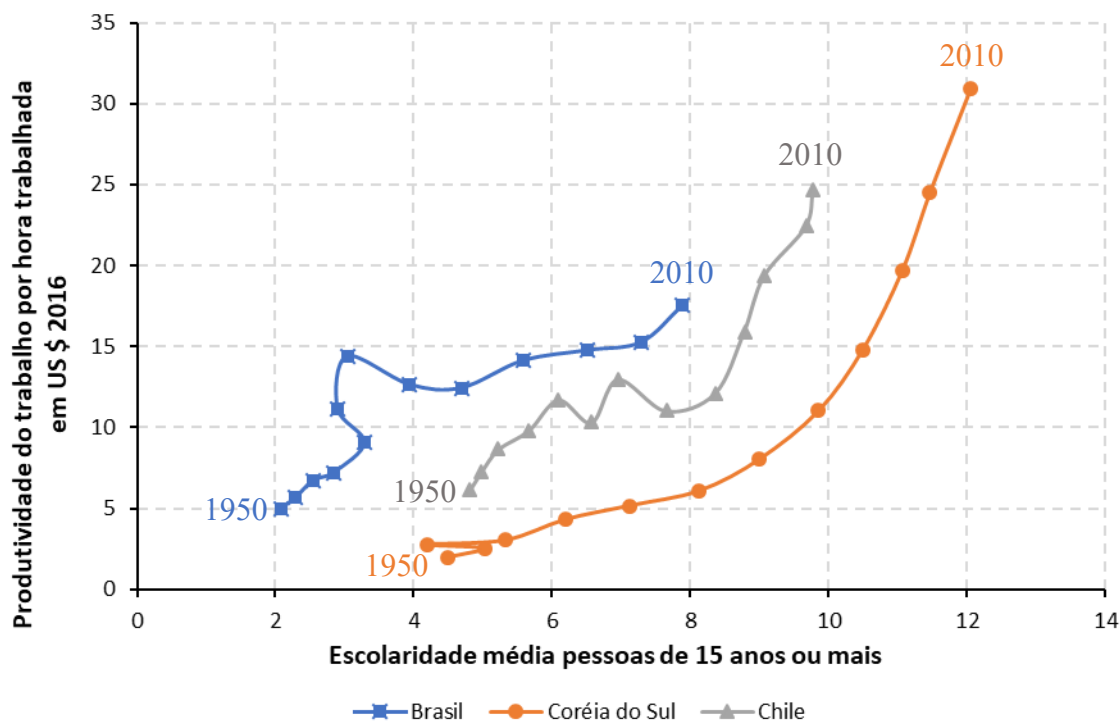
1. INTRODUÇÃO

O Estado de São Paulo criou uma política de bonificação por mérito aos docentes da rede estadual em 2008. A importância de políticas educacionais, em geral, reside em como seu desenho é feito e se os incentivos dados aos agentes são os que geram eficiência nos resultados e menores desvios de comportamento, de forma a manter o real escopo do programa. Dessa forma, o presente trabalho tem a finalidade de analisar a política educacional do Estado de São Paulo, de modo a compreender se os alunos paulistas têm um desempenho melhor na prova que vale bônus para o professor – “high stake exam” – no caso, a Prova do SARESP, quando comparado com a Prova Brasil. Assim, a hipótese dessa monografia é que os alunos terão um desempenho melhor na prova do SARESP em comparação com a Prova Brasil, uma vez que a primeira tem um valor maior para o professor, o bônus, o que cria uma discrepância face a matriz de incentivos por eles enfrentados em cada prova.

O Brasil é o décimo país mais desigual do mundo, segundo o relatório das Nações Unidas de Desenvolvimento Humano (RDH) (2016). Evidências empíricas sugerem que a educação é uma variável muito relevante para explicar a desigualdade de renda no Brasil [Menezes-Filho (2001)], bem como, responsável por grande parte das diferenças de produtividade entre os países [Hall and Jones (1998) e Romer e Weil (1992)]. No caso brasileiro, vemos que apesar de haver uma evolução da escolaridade durante o tempo, a produtividade do trabalho se manteve estagnada durante 30 anos, conforme apresentado no gráfico 1. Isto denota que o inegável desafio de crescimento e inclusão social do país está intimamente relacionado com uma reforma em seu sistema educacional.

A qualidade do ensino brasileiro ainda é baixa quando comparado com outros países. No Programa Internacional de Avaliação de Alunos (PISA) de 2015, o Brasil ficou na 63ª posição entre as 70 nações na área de ciências, 66ª na área de matemática e 59ª na área de leitura. Políticas públicas na área da educação são, portanto, relevantes para dar continuidade ao desenvolvimento econômico e social do país. As reformas educacionais devem ter como principal finalidade a progressiva melhora na qualidade do ensino, visando o melhor desempenho dos alunos.

Gráfico 1: Evolução da Escolaridade da População Adulta e da Produtividade do Trabalho: 1950-2010



Fonte: Conference Board e Barro & Lee

Modelos empíricos sugerem que incentivos individuais para professores são políticas que melhoram o desempenho dos estudantes nos exames e na qualidade de habilidades, como Lazear (2000), DN Figlio, LW Kenny (2007), Lavy (2009). Nesse contexto, diversos países criaram política públicas visando o pagamento por mérito aos professores e funcionários das instituições de ensino. No Brasil, tal incentivo foi criado apenas no Estado de São Paulo, em 2008, tendo o mesmo objetivo dos demais países, melhorar o ensino por meio de incentivos financeiros aos professores e colaboradores das escolas.

Oshiro e Scorzafave (2011) analisaram a política de bônus da rede estadual paulista, por meio da comparação de escolas afetadas pelo programa, com escolas não afetadas. Foram encontrados indícios de que o programa gerou impacto positivo apenas nas notas dos alunos da quarta série.

Esta monografia, por sua vez, busca preencher a lacuna da análise dessa política pública no que tange ao comportamento dos professores das escolas estaduais. Isto é, fazer uma análise no âmbito interno de cada escola, investigando se a resposta aos diferentes incentivos dados aos professores provocará resultados distintos em cada prova, mesmo que ambas sejam teoricamente similares. Por exemplo, por meio de incentivos aos

melhores alunos, ou desincentivo aos piores; ou mesmo a atenção dada em sala de aula para cada prova.

Este estudo busca, portanto, contribuir para a análise da política de bônus implementada no Estado de São Paulo em 2008. A avaliação dessa política pública é importante para o constante aperfeiçoamento de seu desenho, em busca do avanço da qualidade do ensino das escolas públicas estaduais paulistas. Este trabalho está dividido em seis seções, sendo a segunda uma revisão bibliográfica, que faz uma revisão dos artigos importantes que contemplam o assunto. A terceira é referente a metodologia que será adotada para analisar a hipótese, a quarta faz uma análise dos dados a serem utilizados, a quinta comenta os resultados obtidos. A sexta seção, por fim, contém a conclusão e considerações finais.

2. REVISÃO BIBLIOGRÁFICA

Em 2001, emergiu nos Estados Unidos um movimento chamado *No Child Left Behind Act* (NCLB), o qual apoiou uma reforma na área da educação com destaque em políticas de incentivo, embasada na premissa de que o estabelecimento de altos padrões e metas mensuráveis iriam aumentar a qualidade de ensino dos alunos. Um dos elementos principais dentro desse sistema são os testes padronizados, os quais servem como uma proxy do conhecimento dos alunos.

Stiggins (1999) sugere que os administradores, professores e alunos sofrem fortes pressões devido ao alto grau acadêmico exigido nos testes padronizados, o que por um lado pode acarretar em um trabalho produtivo por parte dos professores. Mas, por outro lado, o autor salienta o constante desafio que estes testes geram à profissão docente. Earl, Levin, Leithwood, Fullan & Watson (2003) por meio de entrevistas e pesquisas com educadores, formuladores de políticas, escolas e consultores de alfabetização, concluem que devido às altas metas estipuladas para as escolas, muitas vezes educadores distorcem seus esforços tentando enviesar o ensino apenas para conteúdos que levam ao aumento da pontuação dos alunos nas avaliações. Muitos professores entrevistados, de fato, reconheceram uma preparação considerável para o teste, isto é, a carga horária em sala de aula sendo usada para praticar exercícios do teste padronizado, ao invés do desenvolvimento de novas habilidades dos estudantes. Sacks (2000) e Herman (1992) acreditam que estas preparações, também chamadas de “teaching for the test”, são capazes de ocasionar um efeito “emburrecedor” nos alunos, na medida em que o pensamento crítico é negligenciado e substituído por práticas de atividades do currículo dos testes padronizados.

Tendo em vista as mudanças de comportamento dos professores e práticas como “teaching for the test”, existem estudiosos que criticam a não eficiência da política de bonificação por meio de evidências empíricas. É o caso do estudioso Roland G. Fryer (2011) que elaborou um estudo em mais de duzentas escolas públicas na cidade de Nova Iorque para compreender melhor o impacto de políticas de incentivo. Evidências levam à conclusão de que incentivos aos professores não aumentam o desempenho, a frequência ou a formatura dos alunos. Na realidade, os resultados da pesquisa apontam que se o aluno do ensino médio frequentou escolas que implementaram o programa de incentivos por três anos, sua pontuação no teste diminuirá. Análises semelhantes foram feitas por Springer et al. (2011) nas Escolas Públicas de Metro Nashville (MNPS), onde foram

realizados experimentos do Projeto de Incentivo ao Ensino (POINT) nas escolas de ensino fundamental (6º a 9º ano) durante três anos acadêmicos, de 2006 a 2009. Contudo, os autores não encontraram nenhuma diferença significativa entre o desenvolvimento dos alunos que foram designados para o grupo de tratamento – que receberam os incentivos – e os designados para o grupo de controle. Isto é, a política de incentivos não obteve os resultados esperados.

Outra grande crítica na academia com relação aos testes de desempenho, os chamados “high stake exams”, é que estes criam uma grande pressão sobre os educadores para garantir que os alunos obtenham melhores desempenhos, tal pressão cria mais incentivos para a trapaça. Amrein-Beardsley, Berliner e Rideau (2010) conduziram um estudo com 3.000 professores do estado do Arizona com o objetivo de investigar os tipos e graus de conhecimento destes nas práticas de trapaça relacionadas ao teste, no programa No Child Left Behind. Os dados evidenciaram a presença de práticas fraudulentas no programa, como também que os professores atuam de modo inteligente quando falsificam os resultados.

Este tema, todavia, não é um consenso entre os estudiosos de educação. Evidências empíricas também apontaram para resultados positivos da política de bonificação. Muralidharan e Sundararaman (2011) elaboraram uma avaliação aleatorizada com professores que participaram do programa de incentivos implementado em escolas primárias no estado indiano de Andhra Pradesh, e ao final dos dois anos de programa, os alunos em escolas que receberam a política tiveram um desempenho significativamente melhor do que as escolas de controle. Esta melhora foi vista tanto em componentes mais “conceituais” do teste, quanto em componentes mais “mecânicos”, o que leva a crer que os ganhos nas pontuações dos testes representaram aumentos efetivos na aprendizagem dos estudantes.

Atkinson, Burgess, Croxson, Gregg, Propper, Slater & Wilson (2004) realizaram uma pesquisa na Inglaterra, ainda sobre os resultados das políticas de incentivos que foram adotadas no país, e as evidências também denotam resultados positivos. Usando a metodologia “Diferenças em Diferenças” descobriram que o sistema de remuneração aos professores baseado na melhora do ensino dos alunos acarretou em um desempenho melhor nos testes realizados pelos estudantes, cerca de meio ponto por aluno. É salientado, todavia, que esse resultado não se aplica para professores de matemática, uma vez que os alunos não obtiveram progresso nessa matéria. Porém, concluem que os

docentes respondem a incentivos financeiros diretos, e que, portanto, um esquema de incentivos fortemente baseado no progresso dos alunos gerará melhores resultados nos testes. Lavy (2009) também usando a metodologia “Diferenças em Diferenças” e Regressão Descontínua concluiu que a recompensa individual para professores (política de bonificação) foi benéfica, encontrando melhora nas notas dos estudantes de Israel. Contudo, retifica que o experimento foi feito apenas durante um ano e, portanto, não foi possível medir os efeitos de longo prazo dessa política.

Dolton e Gutierrez (2011) propõe questionamentos importantes para compreendermos a política de bonificação indo além dos estudiosos apresentados acima. Eles questionam: Por que os professores são pagos até quatro vezes mais em alguns países do que em outros? Isso importa no desempenho dos alunos? A qualidade dos professores provavelmente será maior se tiverem uma remuneração mais alta na distribuição de renda do respectivo país, e os resultados dos estudantes será influenciado pela maneira com que os docentes são pagos? A partir dessas reflexões, usam dados em painel agregado em 39 países publicados pela OCDE, com uma série histórica de 10 anos, para modelarem tais relações. Foi encontrado que um aumento de cerca de 15% na remuneração dos professores acarreta em um aumento de 6 a 8% na performance dos alunos. Isso quer dizer que um aumento de 5% na posição relativa dos docentes na distribuição salarial aumentaria de 6 a 8% a proficiência dos estudantes. Ainda segundo eles, esse aumento relativo é importante, uma vez que funcionam como forma de melhorar a qualidade do ensino no longo prazo por meio de maiores incentivos aos professores e, também, via atração de melhores profissionais para esse mercado.

Por fim, importante retomar os estudiosos Oshiro e Scorzafave (2011), citados na introdução, uma vez que estudaram os efeitos da política pública de incentivos no Estado de São Paulo, a qual este trabalho também analisará. O estudo usa os métodos “Propensity Score Matching” e “Diferenças em Diferenças” para medir se os alunos do Estado de São Paulo obtiveram um desempenho melhor do que os alunos do resto do Brasil. Os resultados, por sua vez, evidenciam que na oitava série o programa não obteve sucesso, uma vez que o desempenho dos alunos paulistas foi em linha com o resto do Brasil. Contudo, os resultados da quarta série foram positivos, o efeito para Matemática foi de cerca de 6,4 pontos e o de Língua Portuguesa foi cerca de 3,7 pontos na escala SAEB.

3. METODOLOGIA

O presente trabalho pretende avaliar se a política de bônus implementada no Estado de São Paulo a partir de 2008 aumentou a discrepância de nota dos alunos entre a Prova do SARESP e a Prova Brasil, isto é, se os alunos vão, efetivamente, melhor nas provas chamadas “high stake exams”, as quais contemplam incentivos financeiros aos professores. Tais mudanças poderiam influenciar os resultados da prova do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP), a qual as notas servirão como base para o cálculo do bônus, mas o mesmo não ocorrerá com as notas dos alunos na Prova Brasil, tendo em vista que não existem incentivos nem para os professores, nem para os alunos.

O SARESP é aplicado pela Secretaria da Educação do Estado de São Paulo, que tem como finalidade diagnosticar a escolaridade básica paulista. Esta avaliação é aplicada para segunda, quarta, sexta e oitava séries (terceiro, quinto, sétimo e nono ano) do ensino fundamental e terceira série do ensino médio, apenas em escolas do governo estadual. Os alunos dos respectivos anos têm seus conhecimentos avaliados por meio de provas de Língua Portuguesa, Matemática, Ciência Humanas (História e Geografia), Ciências da Natureza e redação. Para o pagamento do bônus considera-se apenas a proficiência nas provas de Matemática e Português, por isso esta prova é chamada na academia de “high stake exam”.

Já a Prova Brasil, formalmente chamada de Avaliação Nacional do Rendimento Escolar (Anresc), é complementar ao Sistema Nacional de Avaliação da Educação Básica (Saeb) e um dos componentes de cálculo do Índice de Desenvolvimento da Educação Básica (Ideb). Este teste é aplicado em estudantes de escolas públicas urbanas em todo o Brasil da quarta e oitava séries (quinto e nono anos) do ensino fundamental, com conteúdo de Língua Portuguesa e Matemática. A Prova Brasil, por sua vez, não tem valor nem para alunos, nem para professores, uma vez que não existem (des)incentivos dados a nenhum desses grupos no caso de resultados bons (ou ruins), dessa forma, este exame é conhecido na academia como “low stake exam”.

No trabalho proposto a metodologia utilizada será “Diferenças em Diferenças”, utilizando dados relativos às avaliações dos anos 2007 e 2009, nas provas do SARESP e Prova Brasil. Esta metodologia possui um grupo de tratamento, o qual sofreu a intervenção e um grupo de controle, o qual não sofreu a intervenção e será comparável ao grupo de

tratamento. Como o nome sugere, o cálculo desse método é uma dupla subtração, sendo a primeira a diferença entre as médias da variável de resultado no período anterior e posterior a intervenção, tanto para o grupo de tratamento quanto para o grupo de controle; e a segunda diferença refere-se a diferença da primeira diferença calculada entre os dois grupos.

Para identificar o efeito causal da intervenção, o método “Diferenças em Diferenças” contempla um conjunto de hipóteses. Primeiro, a evolução da variável resposta do grupo de tratamento corresponde à trajetória dessa variável para o grupo controle na ausência do programa. Isto é relevante para que os desvios encontrados na trajetória do grupo de tratamento seja, efetivamente, efeitos causais do programa. Outra hipótese do método é que a composição de ambos os grupos não se altere de forma significativa entre os períodos anterior e posterior à intervenção. Em termos do problema, isto significa que a composição de alunos nas escolas não tenha se alterado significativamente. E a última hipótese básica do método DD é que ambos os grupos não sejam afetados de forma heterogênea por mudanças de qualquer natureza que ocorram após o programa. Ou seja, que não ocorra nenhuma mudança independente do programa tanto no grupo de controle como no grupo de tratamento, de modo que o efeito causal observado não seja inteiramente advindo da intervenção, mas também da mudança idiossincrática.

Neste trabalho tem-se como grupo de tratamento os resultados da prova do SARESP entre 2007 e 2009, uma vez que este está sob o contexto da intervenção, o programa de bônus ao professor. E o grupo de controle corresponde aos resultados da Prova Brasil entre 2007 e 2009, uma vez que não está sob intervenção, e segue uma trajetória praticamente igual a do grupo de tratamento na ausência do programa.

Dessa forma, a primeira diferença será entre as médias das provas do SARESP e da Prova Brasil nos anos de 2009 e 2007, respectivamente. E a segunda diferença será a diferença entre estas diferenças, como ilustrado a seguir:

$$\beta_{DD} = (\overline{Y_S^{09}} - \overline{Y_{PB}^{09}}) - (\overline{Y_S^{07}} - \overline{Y_{PB}^{07}}) \quad (1)$$

Em que,

β_{DD} é o estimador de Diferenças em Diferenças (DD)

$\overline{Y_S^{09}}$ é a média da nota da prova do SARESP em 2009

$\overline{Y_{PB}^{09}}$ é a média da nota da Prova Brasil em 2009

$\overline{Y_S^{07}}$ é a média da nota da prova do SARESP em 2007

$\overline{Y_{PB}^{07}}$ é a média da nota da Prova Brasil em 2007

Embutida nesse estimador está a primeira hipótese apresentada, de que a variação do grupo de controle representa a variação do grupo de tratamento na ausência do programa de bônus. Em termos do problema, isto quer dizer que dado que a prova do SARESP e a Prova Brasil contemplam conteúdos parecidos, a diferença entre as notas das provas deve ser, na média, estatisticamente igual. Portanto, o estimador de Diferenças em Diferenças vai ser responsável por capturar o efeito causal da política do bônus do Estado de São Paulo.

É possível também modelar este problema em uma regressão linear. A equação básica do modelo pode ser especificada como:

$$\overline{Y}_i = \alpha + \theta P_i + \gamma A_i + \beta P_i A_i + \varepsilon \quad (2)$$

Em que,

\overline{Y}_i é a nota média da prova

O subscrito i representa cada escola

P_i é uma dummy que assume valor um para prova do SARESP e valor zero para a Prova Brasil

A_i é uma dummy que assume valor um no período pós programa (2009) e valor zero para período pré programa (2007)

Tendo em vista a equação (2), o efeito causal do programa é medido pelo parâmetro a frente da interação entre as duas variáveis binárias, o parâmetro β . A interação é importante na regressão como forma de capturar o que ocorreu especificamente com o grupo de tratamento no período pós-programa, no caso, em 2009. Isto é, capta se a diferença de desempenho entre as provas aumentou de 2007 para 2009. Em termos do problema, isto quer dizer que se o $\hat{\beta}$ for relevante e positivo significa que a diferença de desempenho no SARESP em comparação com a Prova Brasil foi maior depois da implementação do bônus, comprovando a hipótese deste trabalho.

O modelo especificado na equação (2) será calculado pelo método Mínimos Quadrados Ordinários (MQO) e pelo método regressão quantílica. O primeiro é usado para verificar o impacto médio das variáveis independentes na variável dependente. Já o segundo, refere-se a um método de estimação que permite estudar o comportamento dos indivíduos em diferentes quantis, isto é, analisar o impacto das variáveis independentes em diversos quantis da distribuição condicional. Isto será feito para compreender se as escolas que estão melhorando seu desempenho nas provas são as melhores, a mediana, ou as escolas com pior desempenho em 2007.

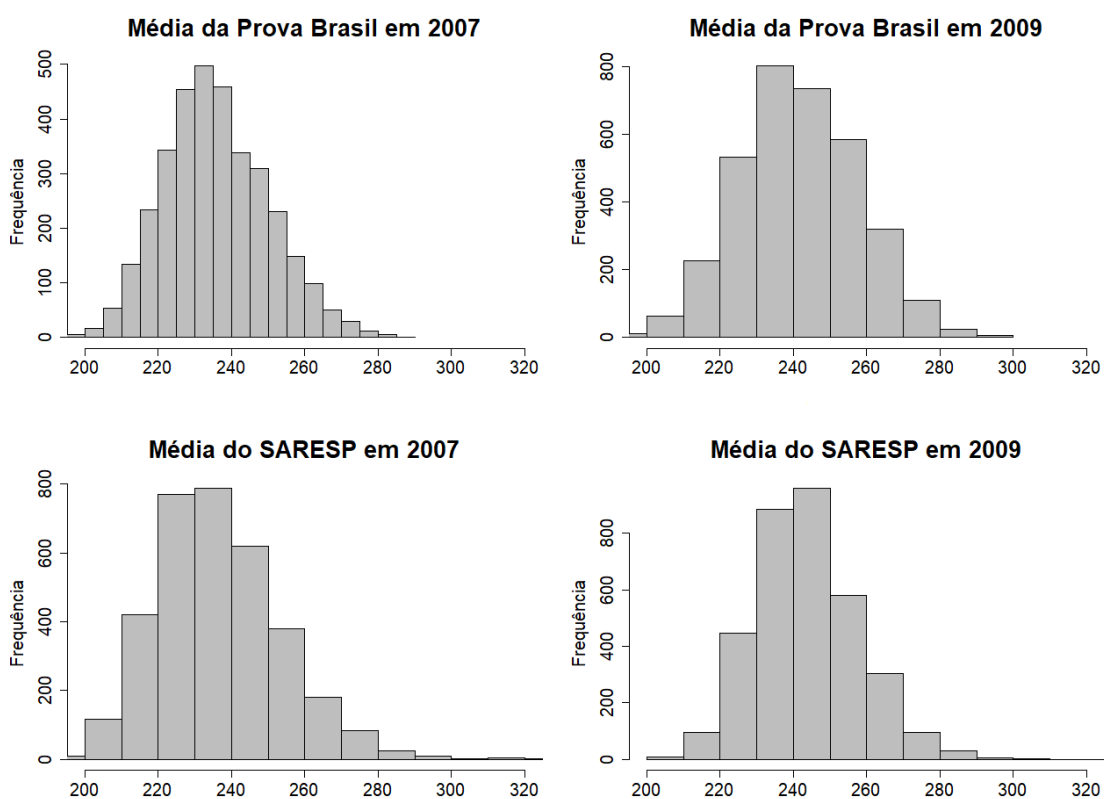
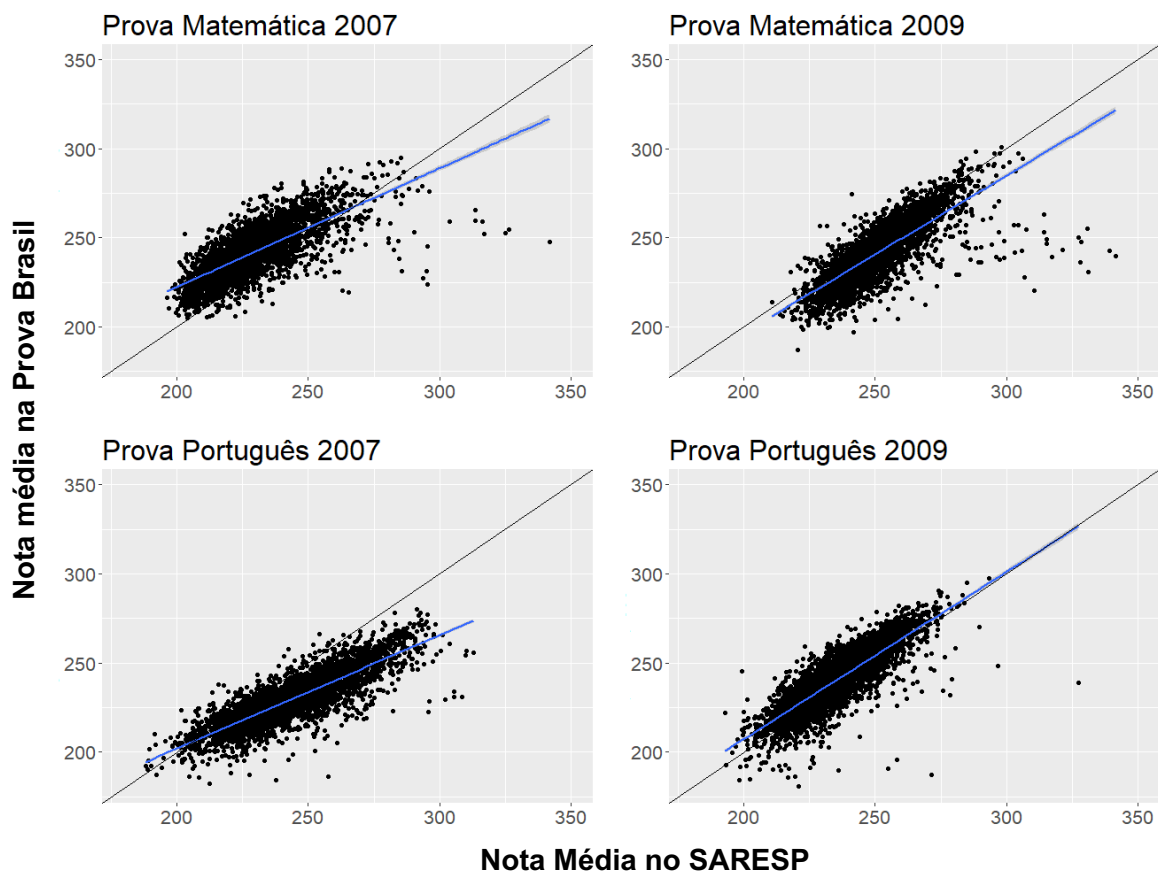
4. DADOS

Como salientado na Metodologia, esta monografia utilizará dados da Prova Brasil e do SARESP dos anos de 2007, período anterior à implementação do bônus no Estado de São Paulo, e 2009, período posterior à política. Os dados referem-se ao 9º ano do ensino fundamental e estão divididos em 3418 escolas do Estado de São Paulo. Correspondente a cada instituição tem-se a nota média de Português e Matemática e o número de alunos que realizaram as respectivas provas.

Tabela 1: Estatística descritiva

	SARESP 2007	SARESP 2009	Prova Brasil 2007	Prova Brasil 2009
Nº médio de alunos (Matemática)	116,97	119,87	113,67	102,13
Nº médio de alunos (Português)	119,13	119,87	113,67	102,14
Média em Matemática	227,77	251,28	240,85	241,78
Média em Português	245,26	236,08	230,79	241,18
Média (Português e Matemática)	236,51	243,68	235,82	241,48
Desvio Padrão (Matemática)	16,25	14,95	14,62	16,90
Desvio Padrão (Português)	19,82	14,76	14,98	16,81

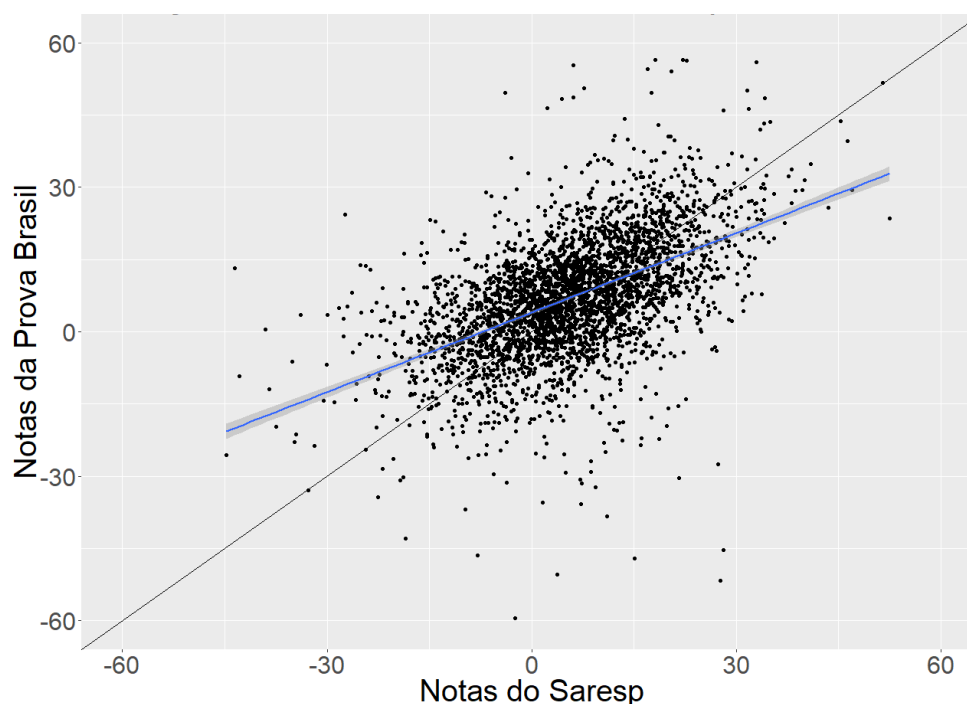
Fonte: INEP e Secretaria da Educação do Estado de São Paulo

Gráfico 2. Histogramas das notas médias das escolas dividido por prova e por ano**Gráfico 3.** Distribuição de notas por escola, dividido por conteúdo

Ao analisarmos primeiramente os histogramas com as notas médias em cada prova nos diferentes anos, graficamente não se vê grandes mudanças na distribuição das notas. Porém, pela tabela 1 entende-se que após a implementação do bônus os alunos apresentaram um desempenho melhor na prova do SARESP, em comparação com a Prova Brasil, o que vai de acordo com a hipótese inicial do presente trabalho, de que os alunos têm desempenho melhor na prova que vale bônus para o professor (“high stake exam”). Contudo, analisando mais de perto os dados desagregados da tabela, conjuntamente com o gráfico 3, verifica-se que tal procedimento ocorre com as notas da prova de Matemática, mas não sucede com as notas de Português.

No gráfico 3 observa-se que as notas de Matemática em 2007 estavam, no geral, mais altas na Prova Brasil do que na Prova do SARESP, onde a maioria dos dados se localizam no lado esquerdo do gráfico, no qual a curva de tendência (azul) está acima da curva de 45° (preta), curva que representa os pontos em que as notas da Prova Brasil e do SARESP são as mesmas. Em 2009, por sua vez, verifica-se o contrário, onde a curva de tendência está inteiramente abaixo da curva de 45°, salientando que a maior parte das escolas tiveram um desempenho melhor no SARESP do que na Prova Brasil. Essas evidências corroboram com a hipótese inicial. Por outro lado, em Português ocorre basicamente o inverso, uma vez que em 2007 as escolas tiveram um desempenho melhor no SARESP do que na Prova Brasil, mas em 2009 tal curva fica ligeiramente acima, até que converge para a curva de 45°, o que quer dizer que as notas da Prova Brasil foram levemente acima das notas do SARESP.

Gráfico 4. Variação das notas entre 2007 e 2009, por escola



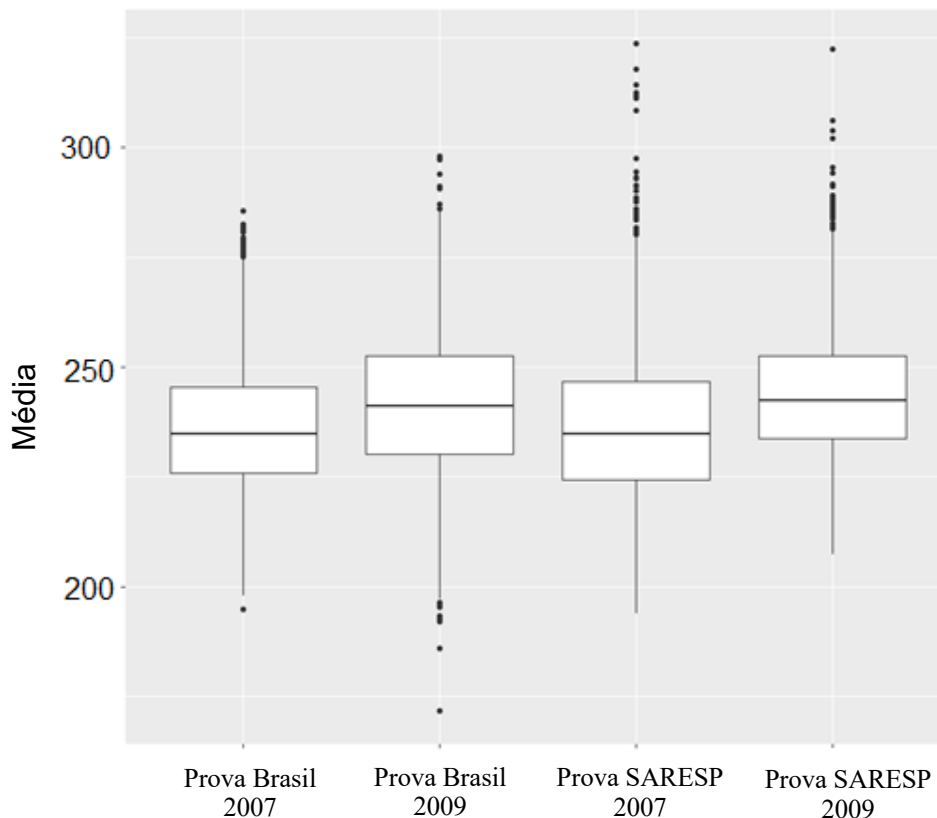
Ainda no que tange à análise de notas, o gráfico 4 expõe a variação das notas em cada escola entre 2007 e 2009, nas diferentes provas. Por tal entende-se, primeiramente, que há uma correlação positiva entre a variação das notas, isto é, quando a diferença entre os anos é positiva em uma prova, também será positiva na outra, o que faz sentido, dado que ambas contemplam conteúdos parecidos e ocorrem em momentos próximos.

Ademais, vê-se no gráfico que a reta de tendência (azul) cruza a de 45° (preta) aproximadamente no valor 10, o que revela que quando a escola teve um desempenho pior em 2009 do que em 2007 ou um desempenho de até 10 pontos positivos de diferença, a variação na Prova Brasil foi maior do que a do SARESP. Em contrapartida, a partir do momento em que a reta de tendência fica abaixo da curva de 45°, significa que em 2009 a escola que teve uma nota de mais de 10 pontos de diferença com 2007 apresentou uma discrepância maior na Prova do SARESP.

Em outras palavras, o gráfico 4 expõe que quando ocorre uma variação baixa na Prova Brasil entre 2007 e 2009, essa variação é baixa ou até menor no SARESP. Porém nas escolas que apresentam um salto maior no desempenho entre os anos (após o cruzamento das retas), a diferença da nota foi maior no SARESP do que a variação na Prova Brasil. Isto é, as escolas com maior crescimento no desempenho, tiveram um desempenho ainda melhor no SARESP, quando comparado com a Prova Brasil. Tais sugestões vão,

novamente, de acordo com a hipótese do trabalho, indicando que muitas escolas tiveram um desempenho melhor na prova que têm um valor maior para o professor.

Gráfico 5. Variação das notas médias das escolas, por prova e ano



O gráfico 5 apresenta a dispersão das notas em cada prova nos diferentes anos. Compreende-se que a Prova do SARESP apresenta mais “outliers” para cima, isto é, com notas mais altas, do que para baixo, escolas com notas muito baixas. Já na Prova Brasil há a presença de mais escolas com desempenhos bastante baixos em 2009, e apesar de existirem “outliers” para cima, mostram-se menores do que os da prova do SARESP. Tal diagnóstico sugere a importância de uma análise mais aguçada nos diferentes quantis da nota média, investigando com mais detalhes as diferentes escolas. Ademais, o gráfico 5 também nos traz indícios que corroboram com a hipótese inicial, uma vez que parece que há mais escolas com melhores desempenhos no SARESP do que na Prova Brasil.

Como mencionado, o método “Diferenças em Diferenças” contempla um conjunto de hipóteses. Dentre elas que a composição dos alunos não se altere de forma significativa entre os períodos anterior e posterior à intervenção, o que parece uma hipótese realista, tendo em vista que os alunos são da mesma escola, e realizam a prova com apenas um mês de diferença. O gráfico 6 e as tabelas abaixo fazem análise deste pressuposto, de

modo a averiguar a composição dos alunos entre os anos, que em caso de grandes mudanças, pode colocar a hipótese em cheque.

Gráfico 6. Distribuição dos alunos por escola, dividido por conteúdo, por prova e por ano

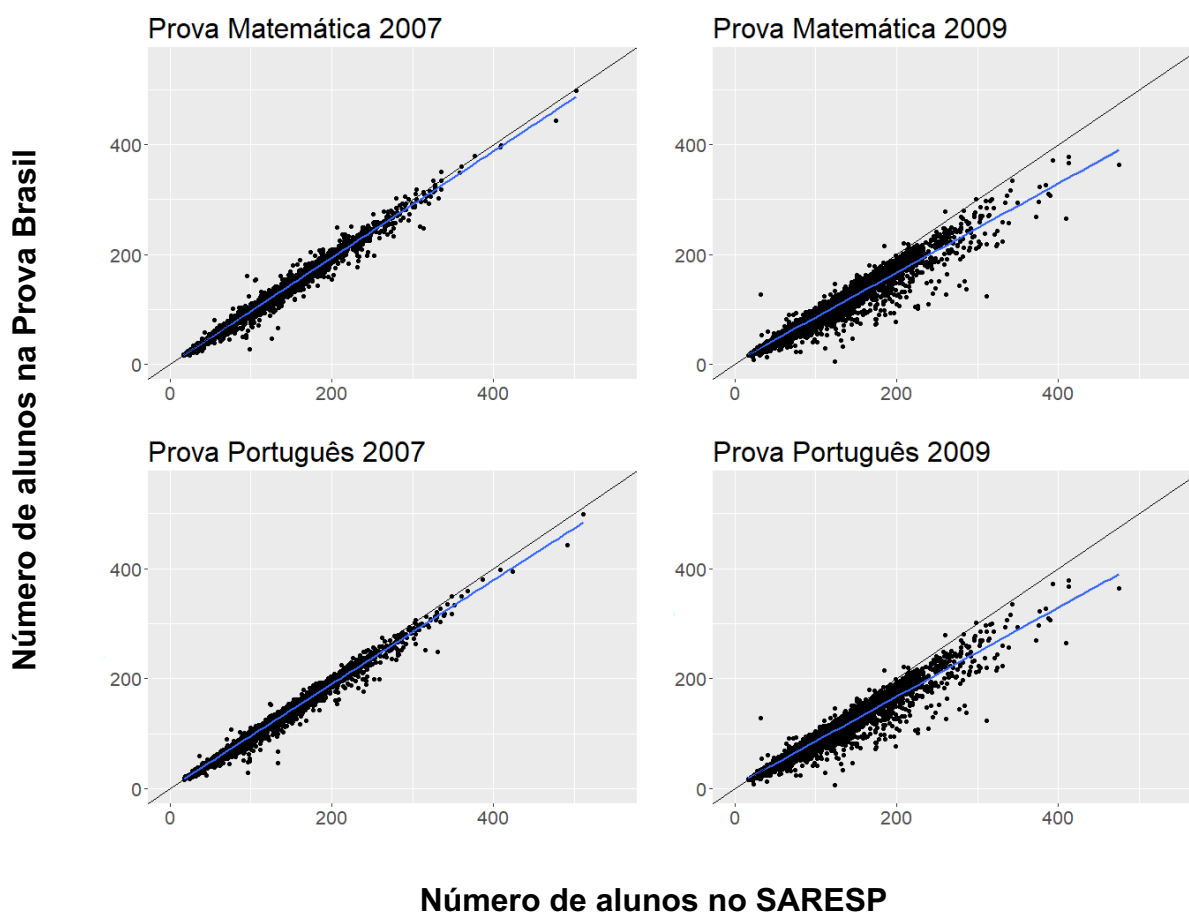


Tabela 2. Correlação do número médio de alunos entre as provas e os anos

	PROVA BRASIL 2007	SARESP 2009
SARESP 2007	0,992	0,828
PROVA BRASIL 2009	0,959	0,799

Fonte: INEP e Secretaria da Educação do Estado de São Paulo

O gráfico 6 indica que houve um aumento da quantidade de alunos que fizeram as provas de 2007 para 2009, este parece maior na prova do SARESP. Este acréscimo, por sua vez, não quer dizer que a composição dos alunos é muito diferente entre os anos. A tabela 1 ressalta que houve uma queda no desvio padrão das notas do SARESP, que não procede na Prova Brasil. Isto posto, há duas possíveis reflexões: (1) todos os alunos melhoraram sua performance, e aqueles que estavam com uma média menor em 2007, obtiveram um

desempenho ainda maior, havendo uma convergência entre as notas; (2) ocorreu uma mudança na composição dos alunos, de modo tinham mais alunos com melhores desempenhos realizando a prova, e os alunos com pior desempenho não a fizeram em 2009. A tabela 2 traz indícios sobre a segunda reflexão.

A queda na correlação do número médio de alunos entre as provas nos diferentes anos (de 0,992 para 0,799), sugere que a composição dos alunos do 9º ano do Estado de São Paulo que realizaram a prova do SARESP e a Prova Brasil em 2007 era mais parecida do que em 2009, o que indica que após a introdução do bônus, pode ter ocorrido uma mudança na composição dos alunos que pode ser oriunda de ambas as provas. Contudo, como a correlação entre os anos da prova do SARESP é menor do que o da Prova Brasil, os dados propõem que esta mudança seja mais expressiva na prova do SARESP.

Tabela 3. Correlação da nota média em cada matéria nas respectivas provas, por ano

	Matemática	Português
2007	0,74	0,84
2009	0,78	0,82

Fonte: INEP e Secretaria da Educação do Estado de São Paulo

Em contrapartida, se tivesse ocorrido efetivamente uma mudança na composição de alunos no SARESP, intuitivamente deveria haver menos correlação nas notas de 2009. Entretanto, não é isso que os dados apontam, tendo em vista que apesar da correlação da nota média em Português ter sido menor, a correlação de Matemática foi maior, o que não traz evidências claras de possível contestação da hipótese do método “Diferenças em Diferenças”.

A outra hipótese do método é referente a trajetória da variável resposta (média das provas) do grupo de tratamento (SARESP) entre 2007 e 2009, que na ausência do programa de bônus, deveria corresponder à trajetória dessa variável para o grupo de controle (Prova Brasil) entre 2007 e 2009. Tal hipótese não é possível provar com os dados, porém é provável que ocorra, dado que os alunos são os mesmos, as provas contemplam o mesmo conteúdo e ocorrem em momentos muito próximos, apenas um mês de diferença.

5. RESULTADOS

A primeira regressão efetuada nesse trabalho refere-se à equação (2), destacada na Metodologia, calculada pelo método Mínimos Quadrados Ordinários (MQO). Esta pretende comprovar a hipótese apresentada, se os alunos do Estado de São Paulo, de fato, terão um desempenho melhor na prova do SARESP do que na Prova Brasil, após a implementação do bônus. A variável dependente é a nota média das provas em cada escola, o objetivo é estimar o efeito causal da política de bônus para professores do Estado de São Paulo por meio do estimador de Diferenças em Diferenças. A interação é o parâmetro relevante da regressão, tendo em vista que esta captura se o aumento no desempenho dos alunos no SARESP de 2007 para 2009 foi maior do que o acréscimo na Prova Brasil.

Tabela 4. Diferença de resultados nas provas do SARESP e da Prova Brasil

Variável dependente:	Nota média
Prova*Ano	1.502 *** (0.528)
Dummy Prova	0.695 * (0.373)
Dummy Ano	5.669 *** (0.373)
(Intercepto)	235.812 *** (0.264)
Observações	3418

Erro padrão entre parênteses

Estatística F: p-valor < 2.2e-16

Significância: 0.001 '****' 0.01 '**' 0.05 '*'

Extraímos da tabela 4 que a nota média da Prova Brasil em 2007 (dummy ano = 0 e dummy prova = 0) é representada pelo intercepto, 235,81; 0,695 reflete a diferença entre as duas provas em 2007 e 5,66 denota o crescimento do desempenho dos alunos na Prova Brasil e, portanto, 1,502 é a diferença no desempenho dos alunos na prova do SARESP entre 2007 e 2009, quando comparado com o crescimento da Prova Brasil.

O último parâmetro reflete o coeficiente da interação entre a dummy ano e a dummy prova, o qual apesar de ser um coeficiente positivo e estatisticamente significativo na

regressão da nota média das provas, tem um valor muito baixo de apenas 1,5 pontos. Isto é, depois da implementação do bônus no Estado de São Paulo em 2008, a nota dos alunos no SARESP aumentou em 1,502 pontos a mais do que o crescimento das respectivas notas na Prova Brasil. O aumento de 1,5 ponto no SARESP representa um efeito pequeno quando comparada com a média da Prova Brasil em 2007 = 235, pois significa um aumento de menos de 1% na nota da prova. Assim sendo, os dados apontam que o efeito do bônus, apesar de estatisticamente significativo, é pequeno no “high stake exam” quando comparado com o “low stake exam”.

Tabela 5. Diferença de resultados nas provas do SARESP e da Prova Brasil nos diferentes quantis

Variável dependente:	Nota media (1) Quantil = 0.10	(2) Quantil = 0.25	(3) Quantil = 0.50	(4) Quantil = 0.75	(5) Quantil = 0.90
Prova*Ano	7.764 *** (0.779)	4.964 *** (0.667)	1.276 * (0.649)	-1.074 * (0.816)	-4.530 *** (0.949)
Dummy Prova	-2.188 *** (0.556)	-1.334 *** (0.462)	0.145 (0.463)	1.253 * (0.592)	3.762 *** (0.698)
Dummy Ano	2.694 *** (0.565)	4.421 *** (0.479)	6.341 *** (0.487)	6.974 *** (0.554)	7.829 *** (0.614)
(Intercepto)	218.388 *** (0.324)	225.784 *** (0.291)	234.629 *** (0.312)	245.472 *** (0.403)	255.038 *** (0.448)
Observações	13672				

Erro padrão entre parênteses

Significância: 0.001 ‘***’ 0.01 ‘**’ 0.05 ‘*’

A tabela 5 apresenta os resultados da regressão quantílica, nos diferentes quantis. Percebe-se que os quantis inferiores (0,10 e 0,25), refletem as escolas que tiveram o pior desempenho na Prova Brasil em 2007, visto que o intercepto apresenta os valores mais baixos quando comparados às outras regressões. A variável “Dummy Prova” mostra que as piores escolas obtiveram em 2007 notas mais altas na Prova Brasil do que na prova do SARESP, uma vez que este parâmetro é negativo. Já a variável “Dummy Ano” mostra

que estes subgrupos obtiveram os piores crescimentos na nota da Prova Brasil de 2007 para 2009, tendo em vista que apresentam os menores valores, ao contrastar com as regressões de maiores quantis. Por fim, o parâmetro da interação “Prova*Ano”, o qual os quantis inferiores apresentaram valores positivos, e quanto menor o quantil, maior o valor. Em outras palavras, isto revela que as piores escolas em 2007 experimentaram um aumento maior no desempenho dos alunos de 2007 e 2009 na prova do SARESP do que na Prova Brasil. No caso, o quantil inferior (0,10), as escolas tiveram 7,76 pontos, na escala SAEB, nas notas do SARESP a mais do que o crescimento das respectivas notas na Prova Brasil, e para o quantil 0,25 este número é 4,96. Tal resultado salienta que a política de bonificação afetou as piores escolas aumentando mais o desempenho dos alunos na prova que vale bônus para o professor (“high stake exam”), isto é, na prova do SARESP, o que vai de acordo com a hipótese inicial.

Na mediana, quantil 0,50, o parâmetro da interação é próximo ao do MQO, sendo a diferença entre o crescimento da nota do SARESP 1,27 pontos a mais do que o crescimento da Prova Brasil. Dessa forma, entende-se que na mediana, apesar do resultado ser positivo e estatisticamente significativo, o aumento é considerado pequeno. Nas regressões dos quantis superiores (0,75 e 0,90), vê-se exatamente o oposto dos quantis inferiores. Tais subgrupos representam as escolas que conquistaram as melhores notas na Prova Brasil em 2007 e estas, por sua vez, obtiveram notas ainda melhores no SARESP, visto que o parâmetro da “Dummy Prova” é positivo. Ademais, a “Dummy Ano” salienta que as melhores escolas apresentaram grandes crescimentos de 2007 para 2009 nas notas da Prova Brasil, e tal crescimento foi maior nesta prova do que no SARESP, uma vez que a interação é negativa. À vista disso, entende-se que nas escolas com melhor desempenho o bônus não acarretou em maiores notas na prova que tinha valor maior para o professor (SARESP), quando comparado com a prova que não tem valor para nenhum para ambos os grupos.

6. CONSIDERAÇÕES FINAIS

O presente trabalho procurou fazer uma análise da política educacional criada no Estado de São Paulo no final de 2008, que formulou um programa de bonificação ao corpo docente das escolas estaduais. Tal análise busca compreender se os alunos das escolas estaduais do respectivo estado têm desempenho melhor na prova que vale bônus para o professor (SARESP), quando comparado com a Prova Brasil, a qual não tem valor nem para o aluno, nem para o professor.

A metodologia proposta para fazer tal análise é “Diferenças em Diferenças”, utilizando dados relativos às notas das escolas no Estado de São Paulo nas provas do SARESP e Prova Brasil, dos anos 2007 e 2009. As regressões foram calculadas por dois métodos diferentes, Mínimos Quadrados Ordinários (MQO) e Regressão Quantílica. Os resultados apontam que as escolas, na média, obtiveram uma nota de 1,502 pontos a mais no SARESP do que na Prova Brasil em 2009. Tal valor é, por sua vez, baixo, representando menos de 1% de aumento na nota da prova, tornando-se possível concluir que apesar de haver uma diferença positiva e estatisticamente significativa, este efeito é considerado pequeno.

O método regressão quantílica, entretanto, proporciona a análise em diferentes quantis. Por conseguinte, conclui-se que as escolas que tiveram pior desempenho em 2007, apresentaram crescimento em ambas as notas, porém a diferença entre o crescimento no SARESP e a Prova Brasil em 2009 foi de 7,76 pontos na escala SAEB. Tal discrepância traz possíveis especulações quanto ao comportamento dos professores após a introdução do bônus nas escolas com pior desempenho, dentre elas que o corpo docente pode ter instruído alunos a fazerem a prova com mais esforço e cuidado, ou que gastaram mais tempo nas aulas debatendo o estilo de prova do SARESP, realizando o chamado “teaching for the test”, ou que “incentivaram” os piores alunos a não comparecerem no dia da prova. Tais afirmações são apenas hipóteses, uma vez que este trabalho não se aprofunda a ponto de estabelecer o mecanismo perfeito para tal conclusão.

Ademais, nos quantis mais altos, que representam as melhores escolas, não é notado tal comportamento, tendo em vista que a interação é negativa, e o crescimento nas notas da Prova Brasil foram, na verdade, maiores do que no SARESP, em 2009.

Este trabalho traz evidências e discussões com relação ao comportamento do corpo docente após a introdução do programa de bônus nas escolas. Bem como debate acerca de possíveis distorções que podem ser criadas no ensino com a política de bonificação, uma vez que esta cria uma distinção entre o valor das provas para o professor, o que poderá gerar problemas na mensuração da real melhora do desempenho dos alunos. Todavia, ainda há muito espaço para a discussão no tema, tal qual pesquisas que possam ajudar a elucidar os possíveis motivos que geraram a discrepância entre o crescimento da nota do SARESP quando comparado com o da Prova Brasil em algumas escolas após a implementação da política pública; ou compreender quais foram os comportamentos dos professores para que houvesse tais distorções.

7. REFERÊNCIAS BIBLIOGRÁFICAS:

AMREIN-BEARDSLEY, A.; BERLINER, D.; & RIDEAU, S. **Cheating in the first, second, and third degree: Educators' responses to high-stakes testing.** Education Policy Analysis Archives, n. 18, p. 14, 2010.

ATKINSON, Adele; BURGESS, Simon; CROXSON, Bronwyn; GREGG, Paul; PROPPER, Carol; SLATER, Helen & WILSON, Deborah. **Evaluating the Impact of Performancerelated Pay for Teachers in England.** Centre for Market and Public Organisation University of Bristol, Department of Economics, 2004.

DOLTON, Peter & MARCENARO-GUTIERREZ, Oscar. **If you pay peanuts do you get monkeys? A cross country analysis of teacher pay and pupil performance.** Economic Policy, 2009. p. 5–55.

EARL, Lorna; LEVIN, Ben; LEITHWOOD, Ken; FULLAN, Michael; WATSON, Nancy; TORRANCE, Nancy; JANTZI, Doris; MASCALL, Blair & VOLANTE, Louis. **England's National Literacy and Numeracy Strategies: Final report of the external evaluation of the implementation of the strategies.** Inglaterra: Department of Education and Employment, 2003.

FIGLIO, David N., & KENNY, Lawrence W. **Individual teacher incentives and student performance.** Journal of Public Economics, 2007, p. 901-914.

FRYER, Roland. **Teacher incentives and student achievement: Evidence from New York City Schools.** Cambridge, MA: National Bureau of Economic Research, 2011.

HALL, Robert E., & JONES, Charles I. **Why Do Some Countries Produce So Much More Output per Worker Than Others?.** NBER Working Paper 6564, 1998.

HERMAN, Joan L. **What research tells us about good assessment.** Educational Leadership, 1992, v. 48, n.8, p. 74-78.

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

LAZEAR, Edward P. **Performance Pay and Productivity**. American Economic Review, 2000. v. 90, n. 5, p. 1346-1361.

MANKIW, Nicholas G., & ROMER, David, & WEIL, David. **A Contribution to the Empirics of Economic Growth**. Quarterly Journal of Economics, 1992. v. 107, n.2.

MENEZES-FILHO, Naercio A. **Educação e desigualdade**. Microeconomia e sociedade no Brasil, 2001, p. 13-49.

MENEZES-FILHO, Naercio A. **Os Determinantes do Desempenho Escolar do Brasil**. Instituto Futuro Brasil, Ibmecc-SP e FEA-USP, 2007.

OSHIRO, Claudia H., & SCORZAFAVE, Luiz G. **Efeito do pagamento de bônus aos professores sobre a proficiência escolar no Estado de São Paulo**. 39º Encontro Nacional de Economia, 2001.

STIGGINS, Richard. **Assessment, student confidence, and school success**. Phi Delta Kappan, 1999. v. 8, n.3, p. 191-198.

SACKS, Peter. **Predictable losers in testing schemes**. School Administrator, 2000. V. 57, n. 11, p. 6-9.

SPRINGER, Matthew; HAMILTON, Laura; MCCAFFREY, Daniel; BALLOU, Dale; LE, Vi-Nhuan; PEPPER, Matthew, et al.. **Teacher pay for performance: Experimental evidence from the project on incentives in teaching**. Nashville, TN: National Center on Performance Incentives, 2010