

AVALIANDO O IMPACTO DAS POLÍTICAS EDUCACIONAIS EM SOBRAL

ROBERTO HSU ROCHA *
NAERCIO MENEZES-FILHO †
BRUNO KAWAOKA KOMATSU ‡

Resumo

Este artigo tem como objetivo avaliar o impacto de políticas educacionais adotadas no início dos anos 2000 sobre a qualidade do ensino no município de Sobral, no Estado do Ceará. Essas políticas reorganizaram o sistema educacional, sem incorrer em aumento expressivo dos gastos em educação. Utilizamos uma combinação do *Propensity Score Matching* e do método de diferenças em diferenças para estimar o impacto. Os resultados indicam efeitos positivos significativos no curto e no longo prazo para os anos iniciais do Ensino Fundamental. No entanto, os desafios persistem na ampliação desses resultados para os anos finais do mesmo grau de ensino.

Palavras-chave: educação básica, avaliação econômica, política pública

Abstract

This article aims to evaluate the impact of educational policies adopted in the early 2000s on the quality of education in the municipality of Sobral, in the State of Ceará. These policies have reorganized the educational system without incurring significant increase in spending on education. We use a combination of *Propensity Score Matching* and differences in differences method to estimate the impact. The results indicate significant positive effects in the short and long term for the early years of elementary school. However, challenges remain in extending these results to the final years of the same level of education.

Keywords: primary education, secondary education, economic evaluation, public policy

JEL classification: I21, I25, I28

DOI: <http://dx.doi.org/10.11606/1980-5330/ea110189>

* Universidade da Califórnia, Berkeley. E-mail: robertohsurocha@gmail.com

† CPP/INSPER e FEA/USP. E-mail: naercioamf@insper.edu.br

‡ CPP/INSPER e FEA/USP. E-mail: brunokk@insper.edu.br

1 Introdução

Atingir um ensino público de alto nível ainda é um dos maiores desafios da sociedade brasileira. Embora o acesso à escola tenha sido universalizado na última década (Veloso 2011), a qualidade do ensino ainda é um grande problema, evidenciado pela comparação internacional de notas do *Programme for International Student Assessment* (PISA), da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Na edição de 2012, o Brasil apresentou notas abaixo da média entre os 65 países participantes nas três áreas avaliadas (matemática, leitura e ciências), como mostra a Tabela 1. Os resultados brasileiros foram também muito inferiores aos de outros países em desenvolvimento como a Rússia e a Turquia.

Tabela 1: Médias do PISA 2012

Países	Matemática	Leitura	Ciências
OCDE (Média)	494	496	501
China (Xangai)	613	570	580
Rússia	482	475	496
Estados Unidos	481	498	497
Turquia	448	475	463
Chile	423	441	445
México	413	424	415
Brasil	391	410	405
Argentina	388	396	406

Fonte: OCDE. Elaboração própria.

São poucos os exemplos de sucesso no sistema público brasileiro, porém um dos mais notáveis entre eles talvez seja o de Sobral, município do Estado do Ceará. Há menos de 10 anos, Sobral demonstrava desempenho completamente regular nas avaliações da Prova Brasil, um exame nacional padronizado, mas hoje é uma referência em políticas para a educação básica (Oliveira 2013). Por meio de diversas medidas de gestão da rede pública municipal, tomadas no início dos anos 2000, o município de Sobral foi capaz de elevar a qualidade do ensino e, atualmente, possui o melhor Índice de Desenvolvimento da Educação Básica (IDEB) para os anos iniciais do Ensino Fundamental entre todos os municípios do país com mais de 25 mil habitantes. Um ponto relevante é de que esse desempenho foi realizado sem que os gastos municipais com educação tivessem aumentado de forma muito acelerada.

O presente trabalho tem como objetivo avaliar o impacto das medidas educacionais de Sobral, complementando análises qualitativas com métodos quantitativos e com o instrumental econométrico. Para tanto, trabalharemos com bases de dados da Prova Brasil (2005 a 2011), disponibilizadas pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). As bases de dados da Prova Brasil possuem, além das notas obtidas por cada aluno no exame, respostas de um questionário socioeconômico entregue a todos os participantes da avaliação. O método econométrico utilizado para tentar avaliar o impacto de mudanças na gestão da educação pública sobre a qualidade de ensino de Sobral será uma combinação de duas técnicas, o *Propensity Score Matching* (PSM) e o estimador de diferenças em diferenças. O nosso foco

são as políticas voltadas para os anos iniciais do Ensino Fundamental, portanto, apresentaremos os resultados da metodologia aplicada para o 5º ano (antiga 4ª série) do Ensino Fundamental de forma detalhada. Com relação ao 9º ano (antiga 8ª série), discutiremos somente os resultados da regressão feita para obter o estimador de diferenças em diferenças.

Além desta breve introdução, o artigo está dividido em sete partes. Apresentamos uma breve revisão da literatura na Seção 2. A Seção 3 descreve o município de Sobral e, em especial, as políticas educacionais adotadas por sua Secretaria Municipal de Educação. Na Seção 4, descreveremos a metodologia utilizada neste estudo e as duas técnicas citadas acima. A quinta parte apresenta os dados utilizados e as características descritivas observáveis dos alunos, agrupados por escola. Na Seção 6, mostraremos os resultados obtidos, além de alguns testes de robustez na Seção 7. Finalizamos na Seção 8, buscando agrupar todos os aspectos discutidos no trabalho na conclusão.

2 Revisão de Literatura

A literatura sobre avaliação de políticas educacionais voltadas à qualidade da educação é extensa, cobrindo diversas dimensões do processo educacional.

Em primeiro lugar, o simples aumento de gastos com educação não parece garantir efeitos sobre a qualidade da educação. Hanushek (2003) argumenta que o aumento de recursos governamentais em educação, sem atenção aos esquemas de incentivos internos às escolas, no geral, resulta em melhoras pequenas na qualidade da educação. No Brasil, Menezes-Filho & Amaral (2009) trazem evidências de que, após controlar por características de escolaridade municipais, formação dos professores e quantidade de horas-aula, não há correlações significantes entre os gastos municipais em educação e o desempenho escolar de alunos da 4ª e 8ª séries do Ensino Fundamental. Diaz (2012) traz evidências adicionais de que não há relação direta entre o nível de gastos em educação e o desempenho escolar, na comparação entre municípios brasileiros.

Diversos estudos mais recentes focam sobre mecanismos específicos de melhoria da qualidade do ensino. Veloso (2011) traz um levantamento de experiências internacionais sobre reformas educacionais, em que destaca três mecanismos institucionais presentes na maioria dessas reformas: mecanismos de *accountability* dos agentes envolvidos, de competição entre escolas e de descentralização do sistema educacional.

Em relação ao *accountability*, ou responsabilização de agentes envolvidos no processo educacional, uma das formas mais utilizadas de mecanismos de incentivos de agentes dentro das escolas é a de pagamentos de bônus para professores e funcionários, baseados em metas pré-determinadas. As evidências internacionais de medidas como essas são majoritariamente de efeitos positivos sobre o desempenho dos alunos (Oshiro et al. 2015, Veloso 2011). Em relação ao Brasil, Oshiro et al. (2015) avaliam os impactos de uma política desse tipo, implementada no Estado de São Paulo em 2008. Com uma metodologia da PSM associada ao estimador de diferenças em diferenças, os autores encontram efeitos positivos do programa após um ano, para alunos da 4ª série do Ensino Fundamental (atual 5º ano), equivalentes a 6,3 pontos na escala SAEB em matemáticas, e a 2,6 pontos em língua portuguesa. No entanto, após três anos, o impacto sobre as notas de matemática desaparecem, enquanto aquele

sobre língua portuguesa se torna negativo, com uma reversão à média. Entre os alunos da 8ª série, os impactos são negativos para as duas disciplinas, tanto no curto quanto no médio prazo.

Em relação ao mecanismo de competição, uma forma de ampliação de competição no sistema educacional é por meio de *vouchers*, que permitem que pais escolham e matriculem seus filhos em escolas privadas. No Brasil, no entanto, a educação pública no geral apresenta baixos níveis de competição entre escolas, com regras de alocação que não favorecem a competição, como o critério geográfico de alocação de estudantes (Tavares 2015).

O mecanismo de descentralização do sistema educacional coloca ênfase sobre a gestão local das escolas, sob a premissa de que agentes locais possuem maior conhecimento sobre os problemas e necessidades das escolas de suas áreas. Nos Estados Unidos, estudos sobre a experiência com escolas charter trazem evidências no geral positivas dos efeitos da autonomia dessas escolas em relação ao aprendizado dos alunos e sobre a taxa de conclusão dos cursos. Veloso (2011), no entanto, chama a atenção para possíveis consequências negativas, caso os gestores locais não tenham boa capacidade técnica. No Brasil, o processo de descentralização do Ensino Fundamental ocorreu por meio da municipalização, com incentivos gerados pela Constituição Federal de 1988, a Lei de Diretrizes e Bases de 1996, e especialmente pela institucionalização do Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério (FUNDEF) em 1996. Utilizando as metodologias de PSM e diferenças em diferenças, Razo et al. (2005) avaliam os efeitos desse processo na qualidade do ensino das séries iniciais do Ensino Fundamental dos Estados e concluem que os resultados foram não significantes ou negativos, dependendo da região considerada.

Além desses mecanismos, há estudos que chamam a atenção para a qualidade da gestão da escola e o papel do diretor sobre o desempenho escolar dos alunos. Tavares (2015) explora a descontinuidade da regra de seleção para um programa de melhoria da gestão escolar do Estado de São Paulo, o Programa de Gestão Escolar por Resultados, implementado pelo governo do estado de São Paulo em 2008. A autora traz evidências de efeitos causais positivos da participação no programa sobre as notas de matemática de alunos com maiores dificuldades de aprendizado, sugerindo que a adoção de práticas básicas de gestão geraram melhora na qualidade da educação. Marioni et al. (2014) encontram efeitos diferenciados de dimensões da gestão escolar sobre a qualidade do ensino.

A extensão das atividades educacionais para o contraturno das escolas, a educação integral, também pode estar associada à melhoria na educação pública. Xerxenevsky (2012) examina essa questão em uma avaliação do impacto do Programa Mais Educação, do governo federal, sobre o desempenho de alunos do Rio Grande do Sul. Combinando o PSM com a estimação de diferenças em diferenças, suas estimativas indicam efeitos positivos sobre as notas em língua portuguesa, porém negativos sobre as notas em matemática, para alunos do 5º ano do Ensino Fundamental. Os efeitos sobre estudantes do 9º ano foram muito pequenos. Aquino & Kassouf (2011) utilizam o pareamento por PSM em dados em painel para estimar os efeitos do Programa Escola de Tempo Integral, do governo do estado de São Paulo, e encontram estimativas positivas, porém de pequena magnitude para língua portuguesa, e não significantes para matemática, entre alunos do 9º ano.

Como veremos na Seção 3, as políticas adotadas no município de Sobral

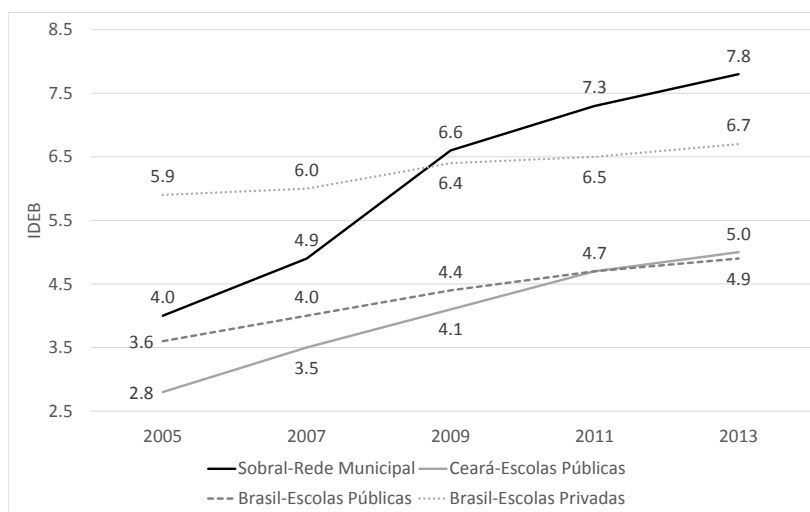
incluem medidas que combinam alguns dos mecanismos descritos anteriormente, porém foram realizadas quase simultaneamente. Por esse motivo, não é possível identificar os efeitos separados de cada dimensão. Dessa forma, nossa análise será realizada sobre o conjunto das políticas educacionais como um todo.

Uma avaliação semelhante a essa que procuramos realizar foi feita por Rocha (2014), que examina o impacto do conjunto de medidas adotadas no município do Rio de Janeiro em 2009. As políticas analisadas por ele incluem medidas de reforço escolar, educação em período integral, realização de oficinas de atividades esportivas e artísticas no contraturno, implementação de laboratórios de ciências, salas de saúde, de leitura e de informática. Além disso, também foi instituída uma política de bonificação de professores e a construção de uma plataforma online de estudos para consulta dos alunos. O autor utiliza uma metodologia de PSM combinada com a estimação por diferenças em diferenças e encontra impacto positivo de curto prazo sobre as notas de matemática e língua portuguesa de alunos do 5º ano do Ensino Fundamental. Após dois anos, ainda se verifica um efeito positivo sobre a nota de matemática, porém com uma regressão à média em língua portuguesa. Entre alunos do 9º ano, o autor encontrou impacto significativo somente após dois anos nas notas de matemática.

3 Sobral e suas Políticas Educacionais

Sobral é um município do interior do Ceará, localizado a 238 km da capital, Fortaleza. Estimativas do Instituto Brasileiro de Geografia e Estatística (IBGE) para 2013 apontam a cidade como a 5ª maior do estado com aproximadamente 198 mil habitantes. O PIB nominal medido em 2011 era de R\$ 2.436.463, colocando o município como a 4ª maior economia do estado e a maior entre as cidades do interior. Sobral poderia ser apenas mais um município brasileiro, mas os resultados obtidos pela sua rede de ensino público o colocam como uma referência em políticas de educação para todo o país (Oliveira 2013). A Figura 1 ilustra a evolução do Índice de Desenvolvimento da Educação Básica (IDEB), um índice misto de qualidade do ensino, dos anos iniciais do Ensino Fundamental para Sobral. O aumento do índice é notório comparado com a evolução no Brasil para o mesmo período.

No entanto, Sobral nem sempre foi destaque na educação pública. No fim de 2000 e início de 2001 foram feitas avaliações externas com alunos da rede de ensino municipal. Os resultados apontaram para uma situação comum em muitos municípios brasileiros. Embora quase todas as crianças de Sobral fossem para a escola, os níveis de aprendizado eram muito ruins: dos alunos avaliados, 40% não conseguiam ler. A partir desse diagnóstico, a gestão da Secretaria de Desenvolvimento da Educação (antiga Secretaria da Educação) entre 2001 e 2004 tomou uma série de medidas com o objetivo de melhorar os níveis de alfabetização de seus alunos. Como os anos iniciais do Ensino Fundamental público eram oferecidos exclusivamente por escolas municipais, essas medidas abrangeram todas as crianças que frequentavam o ensino público. O foco das políticas deixou de ser a ampliação da rede de ensino e passou a ser a qualidade de aprendizagem das crianças, com foco na alfabetização dos alunos. As medidas se organizaram em três grandes áreas: pedagógica, administrativa e de avaliação.

Figura 1: Evolução do índice de desenvolvimento da educação básica

Fonte: INEP. Elaboração Própria.

No campo pedagógico, já em 2001, antes de se tornar uma prática nacional, a Secretaria ampliou em um ano o Ensino Fundamental. Com a criação da 1ª série básica (equivalente ao atual 1º ano) que antecedia a 1ª série regular, os alunos de Sobral tinham um ano a mais para a alfabetização¹. O foco em alfabetização é visto pelos gestores de educação como a base de todo o processo de aprendizado e orientou as metas para a melhoria da qualidade do ensino (Brasil 2007).

Em parceria com institutos de educação e universidades locais, a Secretaria de Educação desenvolveu um material didático unificado. Também foi criada uma escola para professores em 2005 que oferece formação continuada para todos os docentes da rede pública municipal. Os gestores indicam que o ponto diferencial do treinamento é o seu pragmatismo. Cada vez que são orientados, os professores recebem indicações específicas para o trabalho que vão realizar durante o mês dentro das salas de aula, de acordo com o material desenvolvido pela Secretaria. A prefeitura de Sobral também passou a pagar salários maiores que o piso nacional, valorizando o magistério.

Na área administrativa, a primeira medida foi definir junto ao governo do estado as diferentes atribuições de cada esfera. Inicialmente o município passou a ser responsável pelas matrículas até a 4ª série, enquanto a esfera estadual assumia a responsabilidade pelas matrículas do restante do Ensino Fundamental. Isso permitiu que a gestão municipal concentrasse seus esforços na melhoria da qualidade das séries iniciais da educação básica.

Outra política administrativa importante foi a chamada nucleação². Esse processo consistiu em reduzir o número de escolas da rede municipal de 98 para 38, já que as escolas da zona rural apresentavam alta dispersão espacial, ocorrência de turmas multisseriadas e alta variação no número de matrículas, que dificultavam a alocação eficiente de recursos. As unidades fechadas eram da zona rural, e a maior parte delas era de escolas muito pequenas. An-

¹A Lei nº 294 de 26 de Março de 2001 institui o Ensino Fundamental em 9 anos.

²Estabelecida pela Portaria nº 001 de 2001 da Secretaria de Desenvolvimento da Educação.

tes da nucleação, as 40 menores escolas municipais correspondiam a apenas 4,4% das matrículas. Assim, mesmo fechando escolas, a Secretaria conseguiu garantir o acesso à educação para todas as crianças do município.

A prefeitura também instituiu uma nova seleção de diretores, buscando basear suas escolhas em critérios técnicos, e que incluía os diretores que já se encontravam no cargo. O novo método estabeleceu uma seleção em 2001, por meio de um concurso de quatro etapas, incluindo entrevistas, provas e avaliação de currículos. Nessa nova forma de seleção, apenas 25% dos antigos diretores foram selecionados e continuaram na sua função. Essas medidas receberam muita resistência dos políticos, sindicatos e da população local, mas a prefeitura endossou as ideias da Secretaria de Educação, e assim as medidas foram concretizadas.

Com uma estrutura administrativa mais qualificada nas escolas, a Secretaria pôde dar mais autonomia para os gestores daquelas unidades, fazendo com que elas se tornassem os agentes principais do sistema de ensino. Os diretores passaram a ter liberdade de realocar professores entre as turmas e eram responsáveis pelas finanças de sua própria escola. Embora pudessem tomar decisões independentes, os diretores tinham que prestar contas para a Secretaria, que acompanhava de perto o que acontecia em cada escola. Se os resultados não fossem satisfatórios, a Secretaria poderia intervir nas escolas.

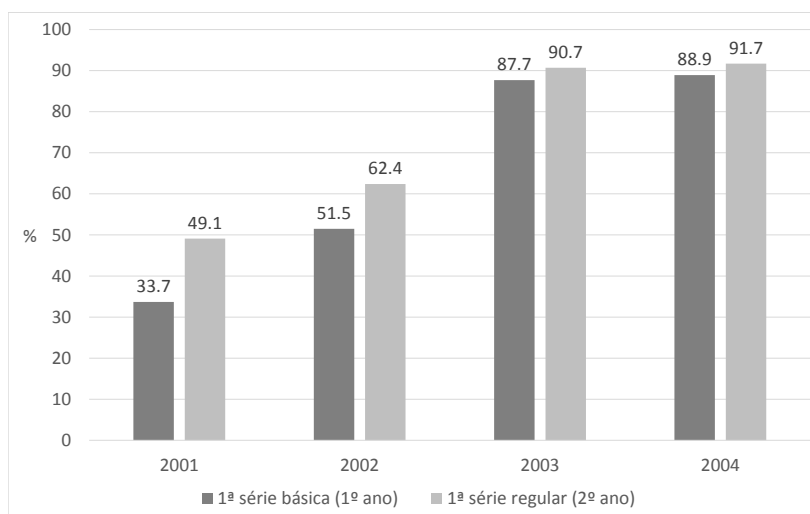
Associada às medidas administrativas e pedagógicas, o município estruturou um sistema de avaliação próprio a partir de 2001, independente das provas do governo federal e estadual. Cada aluno da rede pública de Sobral faz duas avaliações externas por ano. Assim, a Secretaria de Educação consegue acompanhar o desempenho de todas as turmas e analisar individualmente o rendimento de diretores e professores.

Com o desenvolvimento do sistema de avaliação, a prefeitura começou a pagar bônus mensais aos professores e uma premiação anual baseados nos rendimentos dos alunos e das escolas. Em 2002, as gratificações mensais eram destinadas apenas aos professores nas turmas de alfabetização. Isso fez com que os melhores professores fossem para essas turmas, mantendo o foco da Secretaria na política de alfabetização³. Atualmente, todos os professores da rede pública podem receber um bônus de até R\$500 se alcançarem as metas propostas. Com isso, o município parece ter alinhado incentivos para que professores e diretores tivessem suas ações voltadas para a qualidade dentro do sistema público de ensino.

Essas mudanças na política educacional parecem ter provocado efeitos rapidamente. A Figura 2 mostra a evolução da alfabetização entre os alunos do Ensino Fundamental de Sobral. Após as mudanças iniciadas em 2001, os níveis de alfabetização apresentaram melhoras na mesma gestão. Os alunos da 1ª série regular em 2003, ano no qual os níveis de alfabetização começam a se destacar, são aqueles que posteriormente prestariam a Prova Brasil em 2007, quando o IDEB de Sobral começa a se diferenciar do restante do país.

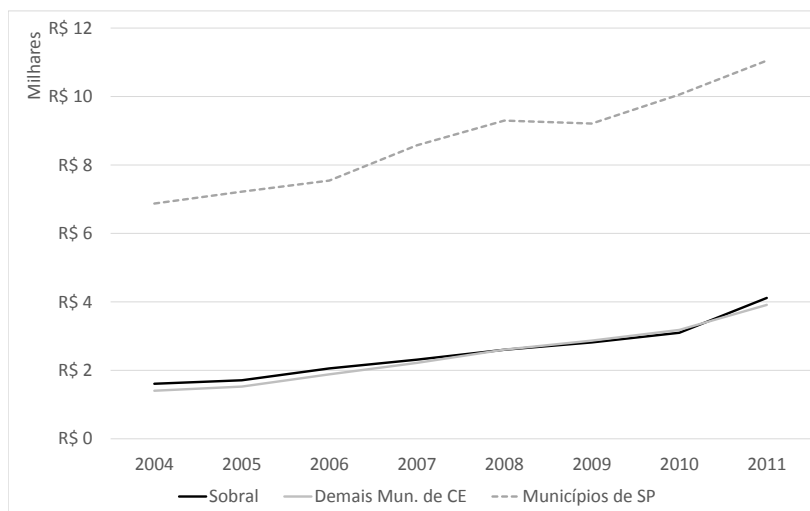
Hoje, Sobral apresenta resultados impressionantes para sua educação básica. 97% das crianças no 2º ano apresentam um nível adequado de leitura, e os resultados do IDEB são expressivos. O nível 7,8 desse indicador para os anos iniciais do Ensino Fundamental sugere que o nível educacional do município já é superior à média dos países desenvolvidos. Esses resultados foram alcançados com mudanças na gestão educacional da administração municipal,

³Gratificação instituída pela Lei nº 342 de 05 de Março de 2002.

Figura 2: Níveis de alfabetização entre 2001-2004

Fonte: MAIA (2006).

sem que se incorresse em elevações acentuadas de gastos. A Figura 3 mostra que a evolução do gasto municipal médio por aluno do Ensino Fundamental em Sobral foi muito semelhante àquela dos demais municípios do Ceará, com um nível muito inferior àquela de municípios paulistas.

Figura 3: Evolução do gasto médio municipal por aluno do Ensino Fundamental (preços constantes de 2015)

Fonte: FINBRA/STN; INEP; IBGE. Elaboração própria. O gasto por aluno foi calculado pela razão entre o total de gastos municipais no EF, deflacionado pelo Índice Nacional de Preços ao Consumidor (INPC), e o total de alunos do EF, somente nos municípios com dados disponíveis para todos os anos.

No entanto, ainda existem desafios para o sistema de ensino público de Sobral. Os resultados do IDEB para os anos finais do Ensino Fundamental não são expressivos. O município alcançou uma média de 5,1, pouco acima da

média nacional de 4,4 em 2013 e significativamente abaixo do valor para os anos iniciais (7,8). A Secretaria de Educação já vem tomando algumas medidas no sentido de municipalizar todo o Ensino Fundamental em Sobral, uma vez que nos anos finais há também a oferta do ensino por escolas estaduais. Além disso, há propostas de implementar o ensino integral.

4 Metodologia

Para a avaliação do impacto das políticas educacionais descritas na Seção 2, utilizamos uma combinação de duas metodologias econométricas. Nosso parâmetro de interesse é o efeito médio do tratamento sobre os tratados (*ATT - Average Treatment Effect on Treated*) em relação à qualidade da educação e nossa estratégia empírica consiste em uma estimação por diferenças em diferenças, combinada com um controle não paramétrico por características observáveis no período inicial. Adotamos essa estratégia, uma vez que a seleção para o tratamento (as políticas educacionais de Sobral) não é aleatória, e não há disponível um experimento natural, tal como um instrumento para as políticas do município de Sobral ou uma descontinuidade. O pareamento no período inicial nos permite construir um grupo de controle muito semelhante em características observáveis ao grupo tratado no período inicial. Além disso, a estimação por diferenças em diferenças nos permite avaliar o efeito das políticas educacionais em períodos posteriores, comparando os grupos tratado e de controle antes e depois de as políticas gerarem efeitos.

O primeiro passo do nosso exercício foi de utilizar um conjunto de características observáveis disponíveis nos dados utilizados para criar um grupo de controle comparável às escolas tratadas, utilizando o método de *Propensity Score Matching* (PSM). Para formalizar esse método, considere a variável D , que caracteriza o indivíduo i , assumindo o valor 1 quando o indivíduo é tratado e 0 quando ele não é. Assuma também que o resultado potencial da qualidade da educação para o indivíduo i seja Y_{i1} caso ele tenha sido tratado e Y_{i0} caso ele não tenha recebido o tratamento. Logo, o efeito do tratamento para o indivíduo i é $Y_{i1} - Y_{i0}$. Dessa forma, o *ATT* é definido como: $ATT = E(Y_1 | D = 1) - E(Y_0 | D = 1)$

No entanto, $E(Y_0 | D = 1)$ não é observável, e com dados observacionais como os nossos, o candidato observável mais próximo é a diferença $E(Y_1 | D = 1) - E(Y_0 | D = 0)$. Em relação ao *ATT*, essa medida carrega o viés de seleção: $E(Y_0 | D = 1) - E(Y_0 | D = 0)$.

Nesse caso, é possível utilizar uma metodologia econométrica para eliminar o viés de seleção. Para isso, são necessárias duas hipóteses, que constituem a chamada hipótese de ignorabilidade forte. Em primeiro lugar, a hipótese de independência condicional (*CIA - Conditional Independence Assumption*) que afirma que, condicional às características observáveis X , os resultados potenciais independem do tratamento: $Y_0 \perp D | X$.

Além disso, para a identificação do *ATT* também é necessária a hipótese de suporte comum: $0 < P(D = 1 | X) < 1$.

Sob essas hipóteses, podemos reescrever o *ATT* com quantidades observáveis: $ATT = E_x[E(Y_1 | D = 1, X) - E(Y_0 | D = 0, X)]$.

Rosenbaum & Rubin (1983) mostram que sob as mesmas hipóteses é possível evitar o problema de multidimensionalidade ao se trabalhar com grande número de regressores. Considerando o *propensity score* $p(X) = P(D = 1 | X)$,

sob a hipótese CIA, temos $Y_0 \perp D \mid p(X)$. O parâmetro de interesse pode ser escrito como: $ATT = E_p(x)[E(Y_1 \mid D = 1, p(X)) - E(Y_0 \mid D = 0, p(X))]$.

Dessa forma, sob as hipóteses de identificação, dado o *propensity score*, é possível criar um grupo de controle com características observáveis semelhantes às do grupo de tratamento e eliminar o viés da estimativa do *ATT*. Neste artigo, realizaremos o pareamento no ano inicial da nossa amostra, 2005, uma vez que as turmas que foram objeto das políticas de forma mais acentuada responderam à Prova Brasil pela primeira vez em 2007. Estimaremos o *propensity score* com uma função normal (*probit*) e em nossa especificação principal faremos o pareamento de escolas tratadas com os 3 vizinhos mais próximos não tratados, com reposição e com um conjunto de características observáveis selecionadas. A escolha pelo método de 3 vizinhos foi feita por conta do pequeno número de observações no grupo de tratamento. Cabe ressaltar que na regressão que define os estimadores, as observações recebem pesos, de forma que o peso total do grupo controle é igual ao do grupo de tratamento, embora o grupo de controle possua mais observações. Apresentaremos adicionalmente pareamentos por outros métodos, para testar a robustez de nossas estimativas. Além disso, também realizamos pareamentos com diferentes conjuntos de características observáveis, com base no algoritmo sugerido por Imbens (2015), para testar a sensibilidade dos resultados ao método de seleção da especificação para a estimativa do *propensity score*.

A partir dos grupos emparelhados, com dados de mais de um período é possível implementar o estimador de diferenças em diferenças, que visa realizar a estimativa do *ATT* controlando por efeitos não observáveis fixos no tempo. Seja Y^A o resultado antes do tratamento, e Y^D o resultado depois do tratamento, podemos escrever o impacto médio como:

$$DD = E(Y_1^D - Y_1^A \mid D = 1) - E(Y_0^D - Y_0^A, D = 1)$$

Esse estimador é identificado sob a hipótese de tendência comum:

$$E(Y_0^D - Y_0^A \mid D = 1, X) = E(Y_0^D - Y_0^A \mid D = 0, X)$$

O estimador de diferenças em diferenças pode ser implementado pelo método de Mínimos Quadrados Ordinários (MQO) nos grupos já definidos de tratamento e de controle de acordo com a seguinte regressão:

$$Y_{it} = \beta_0 + \gamma \text{Sobral} + \beta_1 d2007 + \beta_2 d2009 + \beta_3 d2011 + \beta_4 d2007 \times \text{Sobral} + \beta_5 d2009 \times \text{Sobral} + \beta_6 d2011 \times \text{Sobral} + X'_{it} \gamma + \varepsilon_{it} \quad (1)$$

em que *Sobral* é a variável indicadora de recebimento do tratamento (estar em uma escola de Sobral), *d2007*, *d2009* e *d2011* são variáveis indicadoras de ano, X_{it} são controles que variam ao longo do tempo e ε_{it} é o termo aleatório. Dessa regressão, os parâmetros de interesse são β_4 , β_5 e β_6 , respectivamente, o *ATT* para os anos de 2007, 2009 e 2011 com relação a 2005. As regressões são realizadas com o painel das escolas emparelhadas no passo anterior.

5 Dados

Nesta seção descreveremos os dados, com maior detalhamento sobre a construção e o significado da Prova Brasil.

A nossa variável de interesse na metodologia econométrica é a nota média dos alunos na Prova Brasil por escola. A Prova Brasil, desenvolvida e aplicada pelo INEP, é uma das partes que formam o Sistema de Avaliação da Educação Básica (SAEB). A sua primeira edição sob os moldes atuais foi em 2005, ano no qual o SAEB foi reestruturado, e veio suprir uma grande demanda de gestores, educadores e pesquisadores por informações sobre o ensino em todos os municípios e escolas. A principal diferença em relação às outras formas de avaliação que existiam previamente é o fato de ela ser censitária, ou seja, todos os alunos e escolas (que estejam de acordo com os critérios estabelecidos) participam do exame. As avaliações que existiam até 2005 eram amostrais, portanto, não geravam informações para todas as unidades de ensino.

O aluno que participa da Prova Brasil faz duas avaliações, uma de matemática e outra de português, além de responder a um questionário socioeconômico. A partir de 2007 os professores e diretores também passaram a responder os questionários socioeconômicos. Também cabe aos diretores responderem perguntas sobre condições de infraestrutura e disponibilidade de material escolar em suas escolas. As avaliações de proficiência em matemática e português são desenvolvidas a partir da Teoria de Resposta ao Item (TRI), cuja utilização vem crescendo e se espalhando por diversas áreas do conhecimento. O SAEB se vale da TRI desde 1995 quando as avaliações da educação brasileira em larga escala ainda estavam começando (Andrade et al. 2000). Tal teoria propõe uma nova forma de organizar as provas. Ela pondera de maneira diferente cada resposta dada pelo aluno, da seguinte forma. Testes que mais alunos acertam valem menos, e aqueles com menor número de acertos valem mais pontos na nota final do candidato. Usando a TRI, avaliações são feitas de forma que os resultados entre diferentes anos são comparáveis entre si, mesmo que as provas respondidas por cada aluno tenham sido diferentes. Isso é fundamental para a continuidade da avaliação do sistema de ensino, pois, de outra maneira, variações nas notas entre os anos poderiam ser explicadas por diferenças na dificuldade de cada exame.

Para todas as estimações, utilizamos dados da Prova Brasil, de escolas públicas das redes municipal e estadual. Como as políticas que são o objeto de nossa análise têm como foco os anos iniciais do Ensino Fundamental, apresentamos mais detalhadamente os dados e resultados para o 5º ano.

Para usar o estimador de diferenças em diferenças é necessário que a mesma unidade de análise seja observada antes e depois do tratamento. Dado que a Prova Brasil é aplicada somente aos alunos do 5º e 9º ano e a cada dois anos, utilizamos dados agregados ao nível de escola, pois o mesmo aluno não presta o exame mais de uma vez por ciclo educacional. Apresentaremos adicionalmente alguns resultados para o 9º ano, utilizando a mesma metodologia. Em Sobral, os anos finais do Ensino Fundamental são oferecidos tanto por escolas da rede municipal, quanto da rede estadual, porém até 2011 nenhuma escola municipal havia respondido à Prova Brasil do 9º ano. Dessa forma, a avaliação do impacto das políticas no 9º ano será realizada somente para as escolas estaduais que permaneceram em todos os anos do painel.

A Tabela 2 apresenta as características descritivas que serão utilizadas no cálculo do *propensity score*, selecionadas a partir dos questionários respondi-

dos pelos alunos ao prestarem a Prova Brasil. A tabela traz uma comparação com os dados de escolas de todo o Brasil e também apenas para as escolas da região Nordeste. Nessa tabela, foram selecionadas somente as escolas que têm alunos matriculados no 5º ano. Como o ano que tomamos como base é 2005, não utilizamos nesse trabalho as características descritivas de professores e diretores, já que os questionários para ambos só começaram a ser feitos em 2007. Foram selecionadas somente escolas que participam da amostra em todos os anos. No total, a amostra possui 24.420 escolas, das quais 6.459 situam-se na região Nordeste e 16 no município de Sobral. Selecionamos na amostra somente as escolas presentes nas quatro bases de dados (de 2005 a 2011), pois é necessário avaliar a mesma escola nos diferentes períodos.

Tabela 2: Médias das variáveis descritivas dos alunos agrupadas por escola - 2005

Variáveis dos Alunos	Sobral	Nordeste	Brasil
Notas Português	163,90	158,10	172,80
Notas Matemática	167,80	166,00	180,10
Negros	76,0%	71,0%	61,9%
Possui Computador	11,8%	13,8%	21,1%
Possui Geladeira	78,5%	80,9%	90,8%
Mora com a Mãe	85,7%	82,1%	85,0%
Mãe Completou EM	20,2%	23,4%	29,1%
Pai Completou EM	25,8%	24,6%	30,5%
Incentivado a Estudar	89,9%	89,2%	92,5%
Incentivado a Fazer Lição	88,4%	86,6%	89,9%
Trabalha	16,6%	20,9%	15,6%
Reprovou	25,2%	43,2%	32,4%
Abandonou	11,0%	15,5%	11,1%
Possui Internet	6,9%	8,3%	14,0%

Fonte: INEP/MEC. Médias por escola de características médias de alunos do quinto ano em escolas por localidade. A amostra são as escolas da rede municipal e estadual do Brasil em 2005.

As médias mostram que em 2005 as escolas da rede municipal de Sobral eram semelhantes às da região Nordeste, com pequenas diferenças. Ao que parece, os perfis de estudantes das escolas sobralenses tinham menores proporções de bens (geladeira, computador e internet) e mães com menos escolaridade em comparação com a média do Nordeste. Em contraste, as variáveis relacionadas diretamente aos estudos eram comparativamente melhores: maior incentivo para fazer lição, menos estudantes trabalhavam, haviam sido reprovados ou haviam abandonado o curso.

6 Resultados

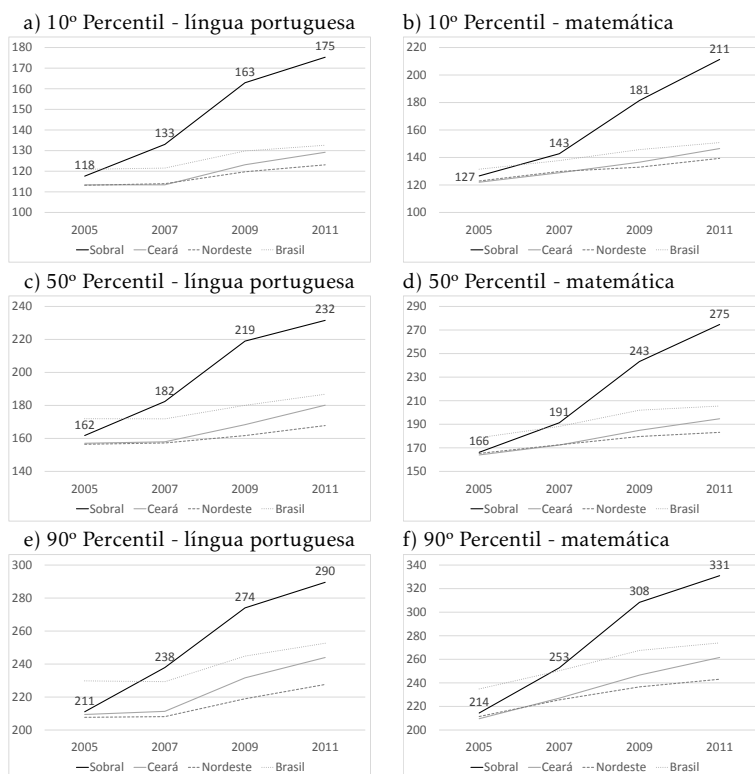
6.1 Resultados Descritivos

Começamos esta seção mostrando graficamente a evolução dos resultados dos alunos do 5º ano na Prova Brasil desde 2005 na Figura 4. A nossa análise compara os alunos de Sobral, com alunos do resto do Ceará, do Nordeste e de todo Brasil. Buscando avaliar o impacto das políticas educacionais da Secretaria de

Educação, apresentaremos resultados para diferentes percentis (10º, mediana e 90º percentil) da distribuição de notas dos alunos. Dessa forma, poderemos ver se os resultados de Sobral concentraram-se apenas nos melhores alunos ou se eles se refletem por toda a amostra. Os gráficos abaixo foram feitos a partir da base de dados de resultados dos alunos da Prova Brasil. Os dados utilizados para essa parte do trabalho são observados por aluno, e não agregados por escola.

Como podemos observar, os resultados positivos nas notas da Prova Brasil não se restringem aos melhores alunos de Sobral. Em todas as partes da distribuição há um crescimento impressionante das notas dos alunos de Sobral, em comparação com o restante do Brasil. Isso indica que as políticas educacionais são bem sucedidas em melhorar o nível de todos os estudantes dos anos iniciais do Ensino Fundamental do município, e não somente uma elite estudantil, composta pelos melhores alunos.

Figura 4: Evolução das notas na Prova Brasil - 5º ano



Fonte: INEP. Elaboração própria.

6.2 Resultados Econômétricos

Seguimos com a apresentação detalhada dos resultados para o 5º ano. Apresentamos primeiro o resultado do pareamento.

O *propensity score* foi estimado a partir do modelo *probit*. Foram incluídas no modelo as variáveis descritivas mostradas na seção anterior. O questionário da Prova Brasil permite a inclusão de diversas características observáveis,

Tabela 3: Resultados do *probit* estimado para características dos alunos do 5º ano

Variáveis Independentes	Variável Dependente	
	Escola Municipal de Sobral	
Nota: Português	0,022	(0,014)
Nota: Matemática	-0,027 **	(0,014)
Negros	2,12***	(0,77)
Possui Computador	-1,43	(1,98)
Possui Geladeira	-1,91***	(0,70)
Mora com a Mãe	0,74	(1,33)
Mãe Completou E.M.	-1,66 *	(0,90)
Pai Completou E.M.	1,22	(0,81)
Incentivado a Estudar	-2,40	(1,50)
Incentivado a Fazer Lição	0,45	(1,56)
Trabalha	0,18	(1,17)
Reprovou	-3,22***	(0,81)
Abandonou	-0,25	(1,41)
Possui Internet em Casa	-0,23	(2,59)
Constante	0,36	(2,13)
Observações	24051	
Pseudo R-quadrado	0,225	

Fonte: INEP/MEC. A amostra são as escolas da rede municipal e estadual do Brasil em 2005. Significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

porém, ao mesmo tempo em que um maior número de covariadas permite um grupo de controle mais parecido com o grupo de tratamento, dificulta encontrar controles na amostra. O *probit* foi estimado com dados de todas as escolas brasileiras das redes municipais e estaduais com turmas de 5º ano do Ensino Fundamental, presentes nos 4 anos investigados. No caso de Sobral, havia somente informações para as escolas da rede municipal, o grupo tratado. Os coeficientes estimados estão na Tabela 3.

A Tabela 4 faz uma comparação das variáveis de interesse antes e depois do *matching*. Podemos ver que as notas no grupo de controle diminuíram após o pareamento. Isso é explicado pelo fato de que as condições socioeconômicas de Sobral em 2005 eram piores que a média do país. Logo, escolas com características observáveis similares às de Sobral tendem a demonstrar notas piores do que a média brasileira, retratada pelo grupo não tratado antes do *matching*.

Tabela 4: Comparação das notas entre os grupos de tratamento e controle para os alunos do 5º ano

		Tratados	Controle	Diferença	Desvio Padrão	Estatística <i>t</i>
Língua Portuguesa	Antes do Pareamento	163,95	172,81	-8,86	4,48	-1,98
	Depois do Pareamento	163,95	164,77	-0,82	4,38	-0,19
Matemática	Antes do Pareamento	167,76	180,14	-12,38	4,46	-2,77
	Depois do Pareamento	167,76	166,47	1,29	4,05	0,32

Fonte: INEP/MEC. Médias de características médias de alunos do quinto ano em escolas dos grupos de tratamento e de controle. A amostra são as escolas da rede municipal e estadual do Brasil em 2005. O p-valor da diferença de médias foi estimado por regressão.

A Tabela 5 mostra o teste de balanceamento feito para as variáveis utilizadas no *matching*, com o objetivo de avaliar a qualidade do pareamento realizado. Foram realizados testes *t* de diferenças de médias entre os grupos. As colunas que mostram os p-valores indicam que antes do pareamento havia algumas diferenças estatisticamente significantes entre as escolas de Sobral e o restante das escolas. Após o pareamento, observamos que, em relação a praticamente todas as variáveis utilizadas, o *matching* teve resultados no sentido de reduzir as diferenças observáveis entre o grupo controle e o grupo de tratamento. Os p-valores indicam que não é possível rejeitar a hipótese nula de igualdade das médias após o pareamento.

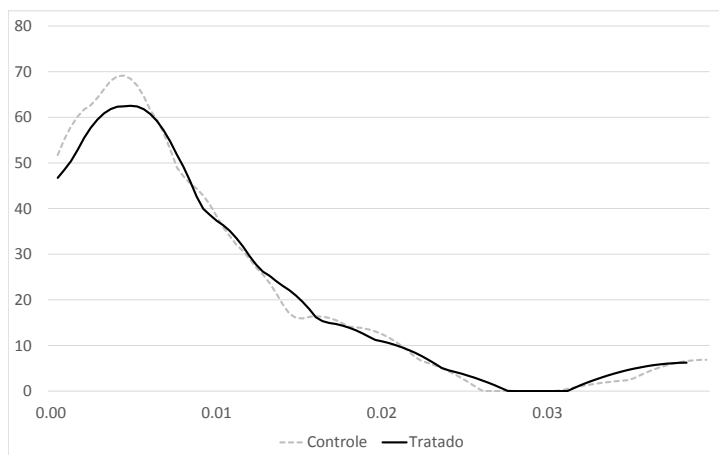
As Tabelas 4 e 5 trazem evidências de que o pareamento foi satisfatório. Para reforçar o teste de balanceamento, a Figura 5 apresenta o gráfico de densidade *Kernel* do *propensity score* para os tratados e para os controles. Os gráficos mostram distribuições praticamente idênticas para os dois grupos, de modo que há evidências de suporte comum.

O próximo passo na metodologia é a estimação da Equação 1, de diferenças em diferenças, por Mínimos Quadrados Ordinários, com dados das escolas emparelhadas. Os resultados obtidos são mostrados na Tabela 6, em que apresentamos os resultados separados por disciplina. Nas colunas (1) e (2) apresentamos as estimativas para as notas, respectivamente, de língua portuguesa e matemática, sem incluir as covariadas dos alunos. Nessa especificação obtivemos estimativas positivas e significantes para o tratamento em ambas

Tabela 5: Teste de balanceamento: características dos alunos do 5º ano

	Antes do Pareamento			Depois do Pareamento		
	Tratados	Controle	p-valor	Tratados	Controle	p-valor
Negros	76,0%	61,9%	0,001	76,0%	76,1%	0,996
Possui Geladeira	11,8%	21,1%	0,008	11,8%	11,8%	0,995
Possui Computador	78,5%	90,8%	0,000	78,5%	78,4%	0,980
Mora com a Mãe	85,7%	85,0%	0,692	85,7%	86,1%	0,792
Mãe Completou E.M.	20,2%	29,1%	0,018	20,2%	19,9%	0,923
Pai Completou E.M.	25,8%	30,5%	0,235	25,8%	23,0%	0,445
Incentivado a Fazer Lição	89,9%	92,5%	0,206	89,9%	90,9%	0,537
Incentivado a Estudar	88,4%	89,9%	0,487	88,4%	89,2%	0,724
Trabalha	16,6%	15,6%	0,641	16,6%	17,6%	0,737
Reprovou	25,2%	32,4%	0,064	25,2%	25,6%	0,903
Abandonou	11,0%	11,1%	0,980	11,0%	9,4%	0,378
Possui Internet	6,9%	14,0%	0,011	6,9%	6,7%	0,949

Fonte: INEP/MEC. Médias por escola de características médias de alunos do quinto ano em escolas dos grupos de tratamento e de controle. A amostra são as escolas da rede municipal e estadual do Brasil em 2005. O p-valor da diferença de médias foi estimado por regressão.

Figura 5: Densidades do *propensity score* estimado

Fonte: INEP/MEC. Densidade do propensity score estimado para os grupos de tratamento e de controle. A amostra são as escolas da rede municipal e estadual do Brasil em 2005.

as disciplinas, com magnitudes crescentes ao longo dos anos. Apesar de a estimativa pontual para a língua portuguesa ser crescente entre 2009 e 2011, essa diferença não possui significância estatística, de modo que o efeito do tratamento chega a um limite superior. O mesmo não ocorre com o impacto sobre as notas de matemática, que continua a crescer.

Nas colunas (3) e (4) nós incluímos as covariadas dos alunos. As estimativas no geral apresentam redução de magnitude em relação aos resultados das colunas (1) e (2), porém mantêm o sinal e a significância. Essas mudanças se devem a mudanças na composição dos grupos de tratamento e controle ao longo do tempo. Como na coluna (1), a diferença entre as estimativas do *ATT* entre 2009 e 2011 não possui significância estatística, o que sugere um limite para o impacto positivo. Entre 2005 e 2011, o tratamento gerou um crescimento médio de 40 pontos na Prova Brasil de língua portuguesa, o que representa um aumento de 28% em relação ao nível inicial.

Tabela 6: Resumo dos resultados das regressões para os alunos do 5º ano

Variáveis Independentes	Variável Dependente: Média das Notas			
	Português (1)	Matemática (2)	Português (3)	Matemática (4)
Sobral	-0,82 (2,96)	1,29 (2,69)	-0,20 (2,89)	2,04 (2,94)
2007	-4,78 (3,08)	9,34*** (3,25)	6,55 (15,86)	33,12 * (18,50)
2009	5,39 * (3,22)	21,72*** (3,46)	-2,43 (2,78)	13,54*** (3,17)
2011	14,72*** (3,19)	27,58*** (3,51)	2,19 (2,80)	14,14*** (3,32)
DD 2005-2007	26,82*** (4,13)	18,14*** (4,28)	25,87*** (3,77)	17,54*** (4,28)
DD 2005-2009	49,56*** (4,56)	55,87*** (5,12)	43,81*** (4,64)	49,64*** (5,55)
DD 2005-2011	51,92*** (4,52)	76,28*** (4,79)	46,39*** (4,34)	70,31*** (5,12)
Constante	164,77*** (2,21)	166,47*** (2,08)	74,92*** (20,63)	86,13*** (23,86)
Covariadas dos Alunos	Não	Não	Sim	Sim
Observações	256	256	256	256
R-quadrado	0,797	0,861	0,880	0,911

Fonte: INEP/MEC. A amostra são as escolas da rede municipal e estadual do Brasil em 2005 a 2011, que realizam a Prova Brasil em todos os anos. O grupo tratado são as escolas municipais de Sobral, e o grupo de controle foi construído pelo método de PSM, com escolas de 2005. Erros padrão robustos entre parênteses. Significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Com relação às notas de matemática, o impacto novamente apresenta crescimento ao longo de todo o período. No longo prazo (2005 a 2011), o impacto do tratamento é maior para matemática do que para língua portuguesa, o que sugere que a intenção dos gestores parece se confirmar. Uma vez que o aluno é de fato alfabetizado, é possível que ele seja capaz de começar a aprender mais nas demais disciplinas. O impacto de longo prazo de 70 pontos na prova de matemática representa, em relação ao nível inicial, uma melhora de 42%.

Os resultados das regressões demonstram, portanto, significativas melhoras nas notas da Prova Brasil para Sobral. Embora as regressões tenham sido feitas com amostras reduzidas, por conta do pequeno número de escolas sobralenses presentes nas bases de dados dos quatro anos da Prova Brasil, os estimadores de diferenças em diferenças de longo prazo mostraram alta significância estatística. Nossos resultados contrastam com os de avaliações anteriores de políticas educacionais (Oshiro et al. 2015, Rocha 2014) no sentido de que os efeitos parecem ser persistentes ao longo do tempo.

Na Tabela 7, apresentamos os resultados finais obtidos pela mesma metodologia, aplicados para os alunos do 9º ano do Ensino Fundamental. Os resultados das regressões mostram que não houve melhora relativa das escolas de Sobral para o 9º ano do Ensino Fundamental em nenhuma das especificações. Entre as estimativas, as únicas que foram estatisticamente significantes foram aquelas de 2005-2011 para língua portuguesa (colunas 1 e 3), e de forma negativa para matemática entre 2005 e 2011 (colunas 2 a 4). Esses resultados indicam que o impacto verificado nos anos iniciais não foi replicado nos anos finais do Ensino Fundamental, o que parece ser uma característica recorrente em avaliações para essa etapa de ensino. Contudo, é importante lembrar que até 2011 a educação dos alunos da rede pública de Sobral no 9º ano era de responsabilidade do governo do estado, e não da rede municipal, em que foram implementadas as políticas educacionais

7 Testes de Robustez

Nesta seção, apresentamos os resultados de alguns testes para avaliar a robustez dos resultados obtidos pelo modelo para o desempenho do 5º ano do Ensino Fundamental. Em primeiro lugar, parece haver evidências de que a hipótese de identificação do modelo de diferenças em diferenças não é violada, de modo que antes do tratamento ambos os grupos apresentam tendências similares em relação a algumas variáveis de resultado escolar. Como a Prova Brasil só começa a ser realizada em 2005, não é possível verificar essa hipótese com as notas desse exame. Em consequência, examinamos as tendências pré-tratamento de indicadores de taxa de abandono e de reprovação e o atraso escolar, com dados do Censo da Educação Superior do INEP.

A Figura 6 mostra as diferenças de média das três variáveis de resultado escolar para os grupos de tratamento e controle. Apesar de se verificar uma ligeira tendência de redução da taxa de reprovação e do atraso escolar entre 2004 e 2005, elas não são estatisticamente significantes e as diferenças das médias dos dois grupos são razoavelmente constantes ao longo do tempo. Esse resultado traz indícios de que as tendências anteriores às políticas educacionais implementadas em Sobral não tiveram efeitos fortes na coorte que prestou a Prova Brasil na 4ª série em 2005, de modo que os efeitos a partir de então podem ser atribuídos às mesmas políticas.

Tabela 7: Resumo dos resultados das regressões para os alunos do 9º ano

Variáveis Independentes	Variável Dependente: Média das Notas			
	Português (1)	Matemática (2)	Português (3)	Matemática (4)
Sobral	5,15 (4,32)	5,15 (4,38)	3,75 (3,71)	3,98 (4,11)
2007	11,23 ** (4,58)	4,31 (4,26)	12,22 (10,97)	16,04 (12,59)
2009	15,25*** (3,88)	4,31 (3,83)	8,32 * (4,56)	-4,32 (4,96)
2011	16,70*** (3,81)	6,44 (4,27)	22,24 * (11,33)	24,82 ** (12,42)
DD 2005-2007	-10,02 (6,76)	-11,38* (6,44)	-10,21 (6,23)	-11,53* (5,97)
DD 2005-2009	4,36 (5,51)	-2,53 (5,69)	7,76 (5,34)	3,46 (5,76)
DD 2005-2011	11,60 ** (5,39)	7,25 (5,75)	11,62 * (6,50)	0,95 (7,25)
Constante	206,17*** (2,66)	221,35*** (2,71)	181,53*** (43,76)	155,58*** (46,88)
Observações	127	127	127	127
R-quadrado	0,448	0,204	0,581	0,380

Fonte: INEP/MEC. A amostra são as escolas da rede municipal e estadual do Brasil em 2005 a 2011, que realizam a Prova Brasil em todos os anos. O grupo tratado são as escolas municipais de Sobral, e o grupo de controle foi construído pelo método de PSM, com escolas de 2005. Erros padrão robustos entre parênteses. Significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figura 6: Diferenças de média de tratados e controles - 5º ano



Realizamos adicionalmente dois testes de robustez. No primeiro deles, testamos a sensibilidade dos resultados à forma de pareamento. Em seguida, verificamos se a variação no método de especificação do modelo que gera o *propensity score* altera os resultados.

Os resultados apresentados até agora foram atingidos por meio do método de *Propensity Score Matching* com 3 vizinhos. Para a realização do primeiro teste de robustez, continuaremos utilizando o *propensity score*, assim como foi definido na Seção 4, no entanto, realizaremos o pareamento de maneira diferente para testar o método de 3 vizinhos. Testaremos o pareamento com 1, 3 e 5 vizinhos, e o método de *Kernel Based Matching* com *bandwidths* 0,001 e 0,0001.

O método de *Kernel* para pareamentos, desenvolvido por Heckman et al. (1997), permite o pareamento de um tratado para vários controles, ao calcular a média ponderada dos não tratados e compará-la com a média dos tratados. A diferença desse método é que ele consegue ponderar todas as observações do grupo dos não tratados dentro de uma janela, definida pelo *bandwidth*. Este

representa o percentual da amostra que entrará na comparação para cada unidade de tratamento. As decisões a respeito do tipo de pareamento e do *bandwidth* estão ligadas a um *trade-off* entre viés e eficiência (Caliendo & Sabine 2008). O pareamento por vizinho mais próximo apresenta viés menor e variância maior do que o pareamento baseado em *Kernel*. Já a escolha de um *bandwidth* maior, dentro do método de *Kernel*, aumenta o viés, mas reduz a variância.

Utilizamos o método de *bootstrap* para estimar os desvios padrões nas regressões de teste, um procedimento que permite inferência estatística ao gerar múltiplas amostras a partir da amostra observada. Neste trabalho, fizemos o *bootstrap* com 50 repetições.

Os resultados das estimações com métodos alternativos de pareamento estão apresentados nas Tabelas 8 e 9. As estimativas obtidas são muito similares àquelas encontradas na Tabela 6 e não se alteram de forma relevante ao longo das colunas, sugerindo que os resultados anteriores são robustos. A significância estatística de praticamente todas as estimativas foi a mesma da encontrada anteriormente sem utilizar o procedimento de *bootstrap*, embora os erros padrão obtidos agora tenham sido em média maiores do que aqueles do modelo inicial. No caso do pareamento por *Nearest Neighbor* com 1 vizinho no modelo para a nota de língua portuguesa (Tabela 8), as estimativas mantêm suas magnitudes, porém a dispersão das estimativas aumenta muito com a redução do número de observações.

As Tabelas 10 e 11 mostram formas alternativas de seleção de covariadas para a estimação do *propensity score*. Nesses testes, mantivemos o pareamento por 3 vizinhos mais próximos, porém a seleção das covariadas utilizadas foi feito com base no procedimento proposto por Imbens (2015). Na coluna 1 de cada tabela, realizamos o procedimento com base nas variáveis selecionadas em nossos exercícios principais (Tabela 6), enquanto na coluna 2, o procedimento foi feito a partir de todas as variáveis comparáveis ao longo do tempo, presentes no questionário da Prova Brasil⁴.

Em ambos os casos, os resultados não apresentam diferenças significativas em relação aos resultados principais. As estimativas para os efeitos das políticas de Sobral entre 2005 e 2007 para notas de português (Tabela 10) são muito semelhantes àquelas da coluna 3 da Tabela 6. Para os efeitos nos anos posteriores, apesar de as estimativas pontuais serem ligeiramente menores do que aquelas obtidas anteriormente, a dispersão é grande e a magnitude é semelhante àquela da estimativa pontual obtida na Tabela 6. Em relação às notas de matemática, a estimativa para o efeito entre 2005 e 2007 perde magnitude e significância estatística, porém novamente a dispersão da estimativa é alta e a ordem de grandeza da estimativa não é menor do que aquela obtida anteriormente. Portanto, as estimativas seguindo o procedimento de seleção de variáveis mostra que a maioria dos nossos resultados principais são robustas

⁴Selecionamos como variável que deveria necessariamente entrar no procedimento somente a escolaridade da mãe. As variáveis adicionais não listadas na Tabela 5 e a partir das quais selecionamos as características relevantes para o pareamento são as proporções de alunos do sexo feminino, de alunos cuja casa possuía alguns itens de infraestrutura (dois ou mais quartos, banheiro), bens duráveis diversos (televisão, rádio, lava-roupa, aspirador de pó, carro), serviço doméstico; as proporções de alunos que cujos pais participavam de reunião de pais, de alunos cujos pais davam incentivo à leitura, de alunos que entraram no sistema escolar durante a pré-escola, de alunos que só frequentaram escolas públicas, de alunos que sempre fazia lições de matemática e de português, e de alunos cujos professores corrigiam sempre as lições de matemática ou de português.

Tabela 8: Testes de robustez para método de pareamento - nota de português

Variáveis Independentes	Variável Dependente: Média de Notas de Português				
	NN - 1 (1)	NN - 3 (2)	NN - 5 (3)	Kernel (Bw=0,001) (4)	Kernel (Bw=0,0001) (5)
DD 2005-2007	21,18 ** (9,90)	25,87*** (5,61)	22,85*** (4,53)	23,20*** (3,44)	23,66*** (3,44)
DD 2005-2009	42,03*** (11,65)	43,81*** (7,95)	35,87*** (6,26)	37,75*** (4,05)	39,07*** (4,05)
DD 2005-2011	45,70*** (12,26)	46,39*** (8,37)	41,31*** (6,68)	41,66*** (4,48)	43,37*** (4,48)
Constante	44,54 (62,81)	74,92 * (45,27)	82,22 ** (33,77)	98,46*** (19,85)	110,69*** (19,85)
Observações	128	256	384	94980	63450
R-quadrado	0,903	0,880	0,864	0,846	0,857

Fonte: Prova Brasil - INEP/MEC. Elaboração própria. Regressões estimadas com controles de características médias dos alunos. Erros padrão estimados por *bootstrap* entre parênteses. Amostras definidas por emparelhamento com covariadas selecionadas (Tabelas 4 e 5), por um, três ou cinco vizinhos mais próximos (respectivamente, nas colunas 1 a 3), e por *Kernel matching*, com *bandwidth* de 0,001 e 0,0001 (respectivamente nas colunas 4 e 5). Significância: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,1$.

a variações nas características observáveis utilizadas para estimar o *propensity score*.

Tabela 9: Testes de robustez para forma de pareamento - nota de matemática

Variáveis Independentes	Variável Dependente: Média de Notas de Matemática				
	NN - 1 (1)	NN - 3 (2)	NN - 5 (3)	Kernel (Bw=0,001) (4)	Kernel (Bw=0,0001) (5)
DD 2005-2007	15,43 (10,61)	17,54*** (6,16)	16,58*** (5,57)	17,70*** (3,88)	17,66*** (3,88)
DD 2005-2009	52,57*** (15,30)	49,64*** (10,45)	43,09*** (8,97)	45,43*** (5,74)	46,03*** (5,74)
DD 2005-2011	70,26*** (13,59)	70,31*** (10,14)	63,94*** (8,61)	65,37*** (5,20)	66,25*** (5,20)
Constante	64,56 (75,09)	86,13 (53,75)	98,15 ** (43,62)	121,04*** (24,55)	127,75*** (24,55)
Observações	128	256	384	94980	63450
R-quadrado	0,935	0,911	0,900	0,886	0,896

Fonte: Prova Brasil - INEP/MEC. Elaboração própria. Regressões estimadas com controles de características médias dos alunos. Erros padrão estimados por *bootstrap* entre parênteses. Amostras definidas por emparelhamento com covariadas selecionadas (Tabelas 4 e 5), por um, três ou cinco vizinhos mais próximos (respectivamente, nas colunas 1 a 3), e por *Kernel matching*, com *bandwidth* de 0,001 e 0,0001 (respectivamente nas colunas 4 e 5). Significância: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,1$.

8 Conclusão

Embora hoje possamos afirmar que o acesso à escola está praticamente universalizado, a qualidade do ensino básico ainda é muito frágil na rede pública. No entanto, as desigualdades sociais persistem, pois aqueles que conseguem matricular seus filhos em escolas privadas de melhor qualidade, enquanto a educação pública oferecida aos menos favorecidos segue, no geral, com baixo nível de qualidade. Acabar com esse ciclo é um dos maiores desafios para que criemos uma sociedade mais democrática e igualitária. Dessa forma, políticas de sucesso na educação pública são de suma relevância.

Nesse contexto, pudemos observar o município de Sobral, cuja rede municipal de ensino alcançou resultados muito significativos nos últimos 10 anos sem aumentar expressivamente seus gastos por aluno. Por meio de mudanças na forma de gerir a rede pública e no projeto pedagógico, Sobral superou todas as metas estabelecidas para a sua educação pública. Utilizando uma metodologia de PSM associada à estimação por diferenças em diferenças, encontramos efeitos positivos de curto e longo prazo das políticas adotadas pelo município na qualidade da educação do 5º ano do Ensino Fundamental. No curto prazo, o impacto só ocorreu sobre as notas de língua portuguesa, o que é coerente com a ênfase na alfabetização. Em contraste com políticas avaliadas por estudos anteriores, no nosso caso os efeitos positivos persistiram e aumentaram após alguns anos. Os efeitos sobre as notas de língua portuguesa, apesar de se manterem positivos, pararam de crescer após dois anos. No caso da matemática, os efeitos continuaram a crescer após 4 anos, de modo que a

Tabela 10: Testes de robustez para especificação do PSM - nota de português

Variáveis Independentes	Variável Dependente: Média de Notas de Português	
	Vars. Seleccionadas (1)	Todas as Vars. (2)
DD 2005-2007	24,83*** (6,32)	22,16*** (7,87)
DD 2005-2009	42,31*** (10,83)	31,29** (12,33)
DD 2005-2011	45,08*** (10,25)	38,09*** (13,19)
Constante	48,17 (43,90)	108,99* (60,37)
Observações	256	252
R-quadrado	0,838	0,873

Fonte: Prova Brasil - INEP/MEC. Elaboração própria. Amostras definidas por emparelhamento por três vizinhos mais próximos, com termos lineares e interações obtidos pelo procedimento proposto por Imbens (2015) a partir das covariadas do exercício principal (coluna 1) e a partir do conjunto de todas as covariadas (coluna 2). Regressões estimadas com controles de médias de características dos alunos. Erros padrão estimados por *bootstrap* entre parênteses. Significância: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,1$.

magnitude da estimativa se tornou maior do que para a língua portuguesa.

Apesar disso, a cidade ainda possui o desafio de estender as melhorias na qualidade de ensino para os anos finais do Ensino Fundamental, que não apresentaram resultados similares àqueles dos anos iniciais. Todos esses resultados foram confirmados em uma série de testes de robustez, com diferentes formas de emparelhamento por PSM e de estimação dos erros padrão.

No contexto de baixa qualidade geral da educação pública no Brasil, as políticas educacionais de Sobral se posicionam no cenário nacional como um bom exemplo e podem ser usadas como referência para novos modelos na rede pública de ensino.

Tabela 11: Testes de robustez para especificação do PSM - nota de matemática

Variáveis Independentes	Variável Dependente: Média de Notas de Matemática	
	Vars. Seleccionadas (1)	Todas as Vars. (2)
DD 2005-2007	16,80*** (5,71)	12,04 (7,45)
DD 2005-2009	51,81*** (10,23)	44,46*** (12,74)
DD 2005-2011	71,35*** (11,13)	63,67*** (14,99)
Constante	56,89 (56,75)	100,15 (65,35)
Observações	256	252
R-quadrado	0,884	0,914

Fonte: Prova Brasil - INEP/MEC. Elaboração própria. Amostras definidas por emparelhamento por três vizinhos mais próximos, com termos lineares e interações obtidos pelo procedimento proposto por Imbens (2015) a partir das covariadas do exercício principal (coluna 1) e a partir do conjunto de todas as covariadas (coluna 2). Regressões estimadas com controles de médias de características dos alunos. Erros padrão estimados por *bootstrap* entre parênteses. Significância: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,1$

Referências Bibliográficas

Andrade, D. F., Tavares, H. R. & Cunha Valle, R. (2000), 'Teoria da Resposta ao Item: conceitos e aplicações', *ABE, São Paulo*.

Aquino, J. M. & Kassouf, A. L. (2011), 'A ampliação da jornada escolar melhora o desempenho acadêmico dos estudantes? Uma avaliação do programa escola de tempo integral da rede pública do estado de São Paulo', *Anais do 39º Encontro Nacional de Economia, Foz do Iguaçu, Paraná*.

Brasil (2007), *Nota Técnica: Metodologia Utilizada para o Estabelecimento das Metas Intermediárias para a Trajetória do IDEB no Brasil, Estados, Municípios e Escolas*, Ministério da Educação, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Acesso em: 6 de Agosto de 2014.

Caliendo, M. & Sabine, K. (2008), 'Some practical guidance for the implementation of propensity score matching', *Journal of Economic Surveys* 22(1), 31–72.

Diaz, M. D. M. (2012), 'Qualidade do gasto público municipal em ensino fundamental no Brasil', *Brazilian Journal of Political Economy* 32, 128–141.

Hanushek, E. A. (2003), 'The Failure of Input-based Schooling Policies', *The Economic Journal* 113(485), F64–F98.

Heckman, J. J., Ichimura, H. & Todd, P. E. (1997), 'Matching as an econometric evaluation estimator: evidence from evaluating a job training programme', *The Review of Economic Studies* 64(4), 605–654.

Imbens, G. W. (2015), 'Matching methods in practice: three examples', *Journal of Human Resources* 50(2), 373–419.

Marioni, L. S., Freguglia, R. S. & Costa, A. B. M. (2014), 'Impactos da gestão escolar sobre o desempenho educacional: uma análise longitudinal', VII Reunião da ABAVE, Avaliação e Currículo: um Diálogo Necessário.

Menezes-Filho, N. A. & Amaral, L. F. L. E. (2009), *A Relação entre Gastos Educacionais e Desempenho Escolar*, Vol. 164, INSPER Working Paper, INSPER - Instituto de Ensino e Pesquisa.

Oliveira, J. B. A. (2013), *O Sucesso de Sobral*, Instituto Alfa e Beto.

Oshiro, C. A. H., Scorzafave, L. G. & Dorigan, T. A. (2015), 'Impacto Sobre o Desempenho Escolar do Pagamento de Bônus aos Docentes do Ensino Fundamental do Estado de São Paulo', *Revista Brasileira de Economia* 69, 213–249.

Razo, R., Fernandes, C. & Soares, S. (2005), 'O Impacto da municipalização no Ensino Fundamental brasileiro: uma estimativa por escores de propensão utilizando os dados do censo escolar'.

Rocha, G. R. (2014), 'Avaliação das políticas educacionais do município do Rio de Janeiro', Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo.

Rosenbaum, P. R. & Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* 70(1), 41–55.

Tavares, P. A. (2015), 'The impact of school management practices on educational performance: evidence from public schools in São Paulo', *Economics of Education Review* 48, 1–15.

Veloso, F. (2011), A evolução recente e propostas para a melhoria da educação no Brasil, in E. Bacha & S. Schwartzman, eds, 'Brasil: a nova agenda social', LTC, Rio de Janeiro, pp. 215–253.

Xerxenevsky, L. L. (2012), 'Programa mais educação: avaliação do impacto da educação integral no desempenho de alunos no Rio Grande do Sul'.