

IMPACTO DA BONIFICAÇÃO EDUCACIONAL EM PERNAMBUCO

CLAYTON SIRILO DO VALLE FURTADO

TUFI MACHADO SOARES

RESUMO

O artigo busca avaliar o impacto do Bônus de Desempenho Educacional, que constitui um incentivo pago aos funcionários das escolas da rede estadual de Pernambuco mediante o cumprimento de metas educacionais preestabelecidas. O objetivo principal é verificar se essa política de bonificação propiciou um melhor desempenho nas proficiências das disciplinas de Língua Portuguesa e Matemática entre 2008 e 2012. A metodologia para abordar essa questão consiste no uso de escores de propensão para produzir pareamentos, em que cada escola da rede estadual (que participa da política de bonificação) é comparada a uma escola similar da rede municipal (que não dispõe de tal política). Afere-se, como principal resultado, que as escolas estaduais obtiveram, no 9º ano do ensino fundamental, um ganho de cinco pontos em relação àquelas da rede municipal.

PALAVRAS-CHAVE ACCOUNTABILITY • AVALIAÇÃO DE POLÍTICAS EDUCACIONAIS • BÔNUS DE DESEMPENHO EDUCACIONAL • MÉTODO DAS DIFERENÇAS EM DIFERENÇAS.

IMPACTO DE LA BONIFICACIÓN EDUCACIONAL EN PERNAMBUCO

RESUMEN

El artículo tiene el propósito de evaluar el impacto del Bono de Desempeño Educacional, un incentivo que se les paga a los empleados de las escuelas de la red estadual de Pernambuco mediante el cumplimiento de metas educacionales preestablecidas. El principal objetivo es el de verificar si dicha política de bonificación propició un mejor desempeño en las asignaturas de Lengua Portuguesa y Matemáticas entre 2008 y 2012. La metodología para abordar este tema consiste en el uso de notas que producen cruces de datos, en que cada escuela de la red estadual (que participa de la política de bonificación) es comparada a una escuela similar de la red municipal (que no dispone de tal política). Como principal resultado, se verificó que las escuelas estaduais obtuvieron, en el 9º año de la educación básica, un aumento de cinco puntos en relación a las de la red municipal.

PALABRAS CLAVE ACCOUNTABILITY • EVALUACIÓN DE POLÍTICAS EDUCACIONALES • BONO DE DESEMPEÑO EDUCACIONAL • MÉTODO DE LAS DIFERENCIAS EN DIFERENCIAS.

IMPACT OF EDUCATIONAL BONUS IN PERNAMBUCO

ABSTRACT

The present article seeks to assess the impact of the Educational Performance Bonus, an incentive paid to employees of the state school network in Pernambuco, when they accomplish pre-established educational goals. The main objective of this paper is to verify whether this bonus policy led to better performance in Portuguese and Mathematics between 2008 and 2012. The methodology used to address this issue uses propensity scores to produce pairings, in which each school in the state network (participating in the bonus policy) is compared to a similar municipal school (which does not have such a policy). The main result shows that state schools have obtained, in the ninth grade, a gain of five points compared to those schools in the municipal network.

KEYWORDS ACCOUNTABILITY • ASSESSMENT OF EDUCATIONAL POLICIES • EDUCATIONAL PERFORMANCE BONUS • METHOD OF DIFFERENCES-IN-DIFFERENCES.

INTRODUÇÃO

Esse artigo analisa uma política de pagamento de incentivos salariais (bonificação), versão *high-stakes* das políticas de responsabilização em educação. Em se tratando de políticas *high-stakes* em educação, a modalidade por meio de pagamento de incentivos salariais é predominantemente adotada no Brasil, ao contrário de intervenções punitivas como as que ocorrem nos Estados Unidos, nas quais professores podem ser dispensados e escolas fechadas. De maneira geral, a concessão do pagamento do bônus às escolas considera o desempenho dos alunos a partir do cumprimento de metas acordadas previamente.

As políticas de responsabilização em educação vêm, aos poucos, se espalhando na gestão da educação brasileira. Em 2011, de acordo com Brooke (2011), pelo menos seis estados brasileiros (Amazonas, Ceará, Pernambuco, Rio de Janeiro, Minas Gerais e São Paulo) já apresentavam experiências de responsabilização e políticas de incentivos salariais e outros dois (Espírito Santo e Distrito Federal) estudavam a implementação desses modelos, havendo uma tendência de

aumento desses números com base na tônica do governo federal, que pode ser representada, dentre outros elementos, pela avaliação censitária da Prova Brasil e o Índice de Desenvolvimento da Educação Básica (Ideb).

Contudo, evidencia-se a importância de conhecer as consequências dessas políticas, pois, conforme aponta a literatura, os resultados têm sido bem variados; um resumo de diversas pesquisas pode ser encontrado em Pontual (2008).

Em se tratando de políticas de responsabilização em educação e suas consequências, o *No Child Left Behind* (UNITED STATES OF AMERICA, 2002) caracterizou-se por ser um dos maiores movimentos nos Estados Unidos e que, muito provavelmente, influenciou na modelagem dessa intervenção em outros países. A avaliação de seu impacto foi analisada no trabalho de Guisbond *et al.* (2012), que não identificaram ganhos significativos na adoção desse programa. Contudo, Hout e Elliott (2011) estabelecem alguns critérios para a validação e seleção dos estudos sobre responsabilização educacional. Tais elementos nem sempre são adotados pelas pesquisas, tornando assim suas conclusões menos robustas.

Essa modalidade de intervenção sob o enfoque da responsabilização educacional é bastante polêmica, até mesmo por conta da diversidade de resultados apontados anteriormente, e engendrou inúmeras críticas, das quais, as principais encontram-se nos trabalhos de Ravitch (2011) e Freitas (2012). Entretanto, a origem das mesmas varia desde argumentos mais ideológicos até os mais empíricos. No sentido de organizar o debate, Brooke (2013) empreendeu um enfoque que permite uma melhor compreensão dessas críticas, elucidando o caminho para os debates futuros.

Uma preocupação que se destaca refere-se ao impacto da responsabilização no estreitamento curricular, ou seja, as escolas e professores dariam maior ou quase exclusiva atenção aos conteúdos que seriam avaliados pelos testes empregados na política de responsabilização. Levantamento de pesquisas (AU, 2007) que abordaram o impacto das políticas de *accountability* no currículo aponta em diferentes direções, pois há trabalhos que indicam o estreitamento e outros a expansão do conteúdo curricular. Assim, o impacto no

currículo depende do desenho da política, do contexto no qual é aplicado e até mesmo da dimensão da matriz curricular antes e após a aplicação das políticas de *accountability*. Um estudo realizado nos estados norte-americanos pelo Conselho Nacional de Pesquisa (*National Research Council*) e financiado pelo *Hunt Institute* mostra que

[...] na ausência de padrões claros e concisos e sem um currículo a eles alinhado, os professores se apoiam nos testes estaduais para obterem diretrizes acerca do que devem ensinar. (THE HUNT INSTITUTE, 2012, p. 178)

Dessa forma, se os efeitos esperados dos programas são variados, o mesmo acontece em relação aos não desejados.

As pesquisas sobre responsabilização educacional vêm crescendo nos últimos anos. O trabalho de Corvalán (2006) reúne estudos relevantes sobre essa temática na América Latina. Já no Brasil, Brooke *et al.* (2011) abordam a utilização das políticas de incentivos salariais em alguns estados, fornecendo um bom suporte para conhecer um pouco mais sobre essa modalidade de intervenção no país.

Na literatura brasileira, existem alguns estudos sobre o desenho e o impacto das políticas de pagamentos de incentivos salariais em educação. Quanto aos diferentes desenhos, Ferraz (2009) abordou o pagamento de incentivos salariais de São Paulo e acrescentou, em sua pesquisa, o Bônus de Desempenho Educacional de Pernambuco (BDE), permitindo a análise do impacto dos diferentes desenhos de bonificação entre os dois estados. Em relação ao impacto, o estudo de Oshiro e Scorzafave (2011), realizado no estado de São Paulo, encontrou um ganho de 6,4 pontos para Matemática e 3,7 pontos para Língua Portuguesa na escala Saeb (Sistema de Avaliação da Educação Básica).

Diante da gradual implantação das políticas de bonificação, cabe uma questão: a utilização das políticas de bonificação em educação, no Brasil, tem produzido efeitos positivos no desempenho dos alunos? Por mais instigante que esse questionamento possa parecer, seria impossível esgotar esse tema tão abrangente no âmbito nacional, considerando o tamanho do Brasil e a diversidade de desenhos de bonificação utilizados.

A alternativa mais viável é o enfoque das análises em um sistema de bonificação estadual. Desse modo, a opção do presente estudo recaiu sobre o estado de Pernambuco. Alguns fatores foram decisivos para tal escolha, entre eles a presença de um sistema próprio de avaliação de larga escala desde 2000, a implementação da política do Bônus de Desempenho Educacional em 2008, além do cálculo de metas com base em um indicador sintético, o Índice de Desenvolvimento da Educação Básica de Pernambuco (Idepe), que conjuga desempenho nos testes com fluxo escolar. Alinhado a esses fatores, existe o fato de o estado ter apresentado um Ideb bem inferior às médias nacionais em 2007. Enquanto o Ideb para o total do Brasil, em 2007, era de 4,2 para os anos iniciais do ensino fundamental (EF), 3,8 para os anos finais e 3,5 para o ensino médio, em Pernambuco, esses valores correspondiam a, respectivamente, 3,6, 2,8 e 3,0 (BRASIL, 2007). Os valores baixos do Ideb, aliados à implantação dessa política de bonificação, esboçam um cenário interessante e de potencial amplitude para melhorias.

Dessa forma, o presente trabalho busca identificar o efeito produzido pelo BDE na rede estadual de Pernambuco entre 2008 e 2012, procurando verificar se foram conquistados melhores desempenhos nas proficiências de Língua Portuguesa e Matemática.

Antes de adentrar nas análises, é relevante salientar que essa intervenção foi baseada na remuneração variável por desempenho. Assim sendo, em 2008, por meio da Lei n. 13.486, de 01 de julho de 2008, instituiu-se o Bônus de Desempenho Educacional. Em seu Art. 1º, já se salienta a vinculação entre bonificação e desempenho:

[...] o Bônus de Desempenho Educacional – BDE, correspondente a uma premiação por resultados, destinado aos servidores lotados e em exercício nas unidades escolares da Rede Pública Estadual de Ensino. (PERNAMBUCO, 2008a, p. 1)

O bônus é pago para todos os funcionários das escolas estaduais que atingem o mínimo de 50% da meta preestabelecida, sendo que o seu valor é semelhante ao percentual de

cumprimento das metas; dessa forma, se a escola atingir 55% da meta, o valor do bônus será igual a esse percentual. Além disso, os funcionários devem estar alocados na escola por no mínimo seis meses no ano do cálculo da bonificação.

Pode-se dizer que esse processo de responsabilização por meio da bonificação consistiria em uma

[...] tentativa de melhorar os resultados das escolas mediante a criação de consequências para a escola ou para professores, sejam elas materiais ou simbólicas, de acordo com o desempenho dos alunos medido por procedimentos avaliativos estaduais ou municipais. (BROOKE, 2008, p. 94)

No tocante às percepções dos professores e diretores, pelo que pode ser percebido até o momento, os mesmos não participaram da construção desse processo em Pernambuco. Entretanto, o resultado de outra pesquisa indica que, nesse estado, não houve resistência por parte dos professores (SEGATTO, 2011). Ao que parece, a política de bonificação ficou confundida com uma melhoria salarial, o que geraria menos resistência daqueles que concorrem ao seu recebimento e, provavelmente, não contam com salários tão atrativos. Nesse caso, pode haver uma mobilização nas escolas para buscar o cumprimento das metas estabelecidas, o que poderia influenciar em mudanças diversas, tais como melhoria no processo de ensino-aprendizagem, gestão e relações intraescolares.

Em 2012, 56% das escolas estaduais de Pernambuco atingiram as metas e fizeram jus ao recebimento do bônus. Os pagamentos variaram entre R\$ 726 e R\$ 3.873, correspondendo a uma média de R\$ 2.976,24 por funcionário das escolas receptoras desse ano (PERNAMBUCO, 2013). Esse valor médio pode ser considerado expressivo se levarmos em conta que o teto da carreira do magistério (classe IV, faixa d, com doutorado), em 2012, era de R\$ 3.000,46. Dessa forma, há chances de o valor do bônus ser suficiente, de modo a fomentar algum tipo de motivação e mobilização nas escolas estaduais de Pernambuco.

Estudos realizados pelo Banco Mundial (FERRAZ; BRUNS, 2010; BRUNS; FERRAZ, 2011; BRUNS; FILMER; PATRINOS, 2011; BRUNS; EVANS; LUQUE, 2012) indicaram algum efeito do BDE

em Pernambuco, sobretudo no 9º ano do ensino fundamental em 2009, devendo-se dar continuidade nessas análises para verificar se os ganhos continuaram alguns anos após esses estudos. Foram verificadas algumas melhorias nas práticas pedagógicas, como, por exemplo, o aproveitamento do tempo dentro da sala de aula pelas escolas que receberam a bonificação. Ressaltam-se, ainda, um melhor aproveitamento do tempo no processo de ensino-aprendizagem e uma redução no tempo do professor fora de sala de aula, bem como no tempo consumido para organização e gerenciamento da sala.

No tocante à pressão exercida para o cumprimento das metas, salienta-se que se trata de um ingrediente inerente a essa modalidade de política. Contudo, se tal pressão sobre as escolas e seus membros não surtirem os efeitos esperados, a política se finda, uma vez que se constitui em um elemento de “desgaste” desnecessário. Obviamente, os resultados das análises de impacto também poderiam servir para subsidiar modificações no desenho dessas políticas, levando a uma espécie de aprimoramento para que os objetivos sejam atingidos com a menor quantidade possível de efeitos não desejados.

Entretanto, verificar o efeito de qualquer política não é algo simples, tendo em vista que a unidade de análise – no caso deste trabalho, a escola – pode estar submetida a mais de uma intervenção de políticas e programas, tanto do governo quanto de instituições privadas e ONGs. Por outro lado, além da diversidade de intervenções políticas que poderiam atuar nas escolas, tem-se todo um contexto socioeconômico e geográfico que pode influenciar nas comparações entre escolas. Assim, para cumprir o proposto nesse trabalho, requer-se o emprego de uma metodologia capaz de neutralizar ou, ao menos, minimizar o efeito de outras variáveis que possam influenciar no desempenho dos alunos.

METODOLOGIA

Em um estudo de impacto de uma política pública seria mais eficaz a realização de estudos experimentais, nos quais seriam escolhidos, aleatoriamente, um grupo de intervenção

(que concorreria pela bonificação salarial) e um grupo controle (que não seria submetido à política de bonificação). Essa metodologia permitiria o controle da influência de outras variáveis sob os indicadores utilizados para mensurar o impacto da adoção da política. Entretanto, é difícil a utilização desse procedimento para uma política educacional, pois, dentre as muitas questões que podem ser colocadas, está a de como excluir, *a priori*, um grupo de participar de tal política. Essa impossibilidade reforça a necessidade de projetos-piloto, já que

[...] o benefício de tal pesquisa seria inestimável, pois as evidências geradas permitiriam que a política finalmente adotada tivesse a maior probabilidade possível de alcançar seu objetivo de melhorar a aprendizagem dos alunos. (PONTUAL, 2008, p. 36)

No entanto, é raro no Brasil se pensar em avaliar determinada política antes de implantá-la em grande escala. Por esse e outros motivos, os estudos experimentais são inviáveis na avaliação de políticas educacionais e estudos observacionais.

O presente trabalho toma como fonte de análise as bases com as proficiências em Matemática e Língua Portuguesa do Sistema de Avaliação Educacional de Pernambuco (Saepe) e da Prova Brasil. Embora a ideia inicial para a análise fosse utilizar apenas as bases do Saepe, que possui aplicações anuais desde 2008, as justificativas para a utilização das análises em duas bases de dados, de origens distintas, foram se consolidando no decorrer do trabalho. A primeira justificativa é que o Saepe apresentou avaliações anuais entre 2008 e 2012, mas não teve uma avaliação em 2007. Considerando-se que o BDE teve início em 2008, seria interessante analisar o período de transição da sua implantação. Uma segunda justificativa é que o cálculo da bonificação não utiliza os resultados da Prova Brasil, o que seria, a princípio, mais adequado metodologicamente, como apontado por Hout e Elliott (2011, p. 5-6):

[para] evitar ter os resultados determinados pela inflação de escore que ocorre nos testes *high-stakes* ligados aos incentivos, os pesquisadores devem utilizar testes *low-stakes*

que não imitem os testes de *high-stakes* para avaliar como os incentivos baseados no teste afetam os resultados.

Por último, como ambas as avaliações encontram-se na mesma escala de proficiência do Saeb, torna-se possível realizar comparações, constituindo, assim, uma forma complementar de análise que permite também uma maior robustez nas conclusões.

A utilização da escola como unidade de análise deve-se ao fato de a premiação ser feita pelo desempenho da escola e não pelo desempenho dos alunos individualmente. Logo, as informações dos mesmos, tal como a proficiência, serão agregadas por escola.

Entretanto, o desempenho do aluno é determinado por inúmeros fatores (SOARES, 2007; SOARES, 2010) e os resultados da escola dependem, além das características socioeconômicas de seus alunos, de uma complexa rede de relações entre diferentes variáveis. A maior dificuldade reside em isolar o efeito das demais variáveis para que seja possível observar apenas a melhoria do desempenho proveniente das políticas de responsabilização.

Para resolver esses problemas, pelo menos em parte, utilizam-se a metodologia de escore de propensão (*propensity score*) e o subsequente pareamento das unidades escolares pelo pareamento dos escores de propensão (*matching score*). Essa técnica vem sendo aplicada no estudo de políticas de pagamentos de incentivos salariais a professores. No Brasil têm-se, por exemplo, as pesquisas realizadas sobre os efeitos do sistema de bonificação existente no estado de São Paulo (OSHIRO; SCORZAFAVE, 2011) e em Pernambuco (BRUNS; FERRAZ, 2011). Tais aplicações sugerem que essa metodologia se demonstra como uma boa alternativa para a observação do efeito do BDE em Pernambuco.

Conforme mencionado, métodos experimentais são difíceis de serem empregados em avaliações de políticas públicas educacionais, mas pode-se tentar um método observacional que se aproxime do método experimental no sentido de, tendo o grupo de tratamento (as escolas que receberam o bônus) já definido, construir um grupo de controle que funcione como um bom contrafactual do grupo tratado.

Esse grupo pode ser definido com o uso de procedimentos que minimizem o viés de seleção, por exemplo, por meio do controle de variáveis sabidamente associadas ao desempenho dos alunos ao final de um período escolar, tais como medidas da condição econômica e as proficiências prévias.

Embora o método proposto no trabalho não seja tão robusto quanto o experimental, no que tange à possível influência de outras intervenções, devemos esclarecer que o BDE era a principal política educacional atuante no estado de Pernambuco, no período analisado nesse artigo, até mesmo pela sua abrangência, visto que todas as escolas da rede estadual poderiam concorrer ao recebimento do bônus. A outra intervenção que temos conhecimento são algumas escolas da rede estadual que funcionariam em tempo integral; para evitar um possível viés, elas não entraram nos pareamentos. Além disso, como nosso grupo controle foi obtido dentro do próprio estado (rede municipal), procuramos reduzir o “risco” de o grupo controle estar submetido a outras ofertas de políticas e programas de outros estados, embora garantir-se disso seja praticamente impossível num estudo observacional e, talvez, nem mesmo em um experimental na área educacional.

O método de pareamento (*matching*) que usa os escores de propensão (*propensity score*) constitui uma alternativa frequentemente sugerida na literatura para esse fim e será usado na produção das análises acerca dos resultados do BDE. No pareamento, escolhe-se uma escola no grupo de controle (no caso, da rede municipal) com características semelhantes a cada escola no grupo de tratamento (rede estadual). Essa escola do grupo de controle funciona como o par da escola do grupo de tratamento, emulando-se, portanto, o resultado que ela teria se não participasse do programa, conforme observado por Pinto (2012, p. 85):

De outra forma, as hipóteses do pareamento postulam que, ao comparar dois indivíduos, um no grupo de controle e outro no grupo de tratamento, com as mesmas características observáveis, o único fator que diferencia os resultados destes indivíduos é a participação ou não no programa.

Ainda de acordo com Pinto (2012, p. 85), existem, pelo menos, dois pressupostos importantes para os resultados das análises de impacto baseadas na técnica do pareamento:

- ignorabilidade (não confundimento): as características observáveis utilizadas devem ser suficientes para controlar o resultado potencial na ausência de tratamento;
- sobreposição: as mesmas variáveis devem representar as características importantes das escolas nos indivíduos presentes dos grupos tratados e não tratados.

O termo *escore de propensão* apareceu pela primeira vez em 1983, em um texto de Rosenbaum e Rubin (1983) que descrevia a estimação de efeitos causais de um estudo de dados observacionais. Essa metodologia consegue reduzir as muitas variáveis que representam as características observáveis importantes a uma única dimensão (GUO; FRASER, 2010).

Uma das formas de estimar o *escore de propensão* é a utilização do modelo *logit* ou *probit* para estimar a probabilidade de participação em determinado programa ou política. Dessa maneira,

[...] o *escore de propensão* é uma pontuação de equilíbrio que representa um vetor de covariáveis. Neste contexto, um par tratado e controle compartilharia um *escore de propensão* semelhante, logo, os elementos do par poderiam ser comparáveis, embora possam diferir em valores de covariáveis específicas. (GUO; FRASER, 2010, p. 130)

O pareamento busca, então, aproximar os participantes dos dois grupos, em termos de *escore de propensão* estimado, ou seja, realiza o pareamento tentando reduzir a perda de participantes pela dificuldade de encontrar um par. Em sua forma mais simples, o pareamento utiliza alguma medida de distância entre tratados e grupo controle em relação a um conjunto de variáveis observáveis (vetor X).

A técnica de pareamento utilizada nesse trabalho é o *nearest neighbor matching* (vizinho mais próximo), por meio da qual os pares seriam formados como um caso de tratamento e outro de controle e como um critério para formação dos pares. Essa técnica emprega a menor diferença entre os *escores*

de propensão entre todas as possibilidades de pares, sendo que cada par formado remove o caso de tratamento e controle da base inicial em que estão os elementos sem pares.

Para evitar escolhas ruins, pode-se usar o *caliper matching*, que é um método de correção do pareamento que determina a distância máxima aceitável entre os escores de propensão. Outra técnica mais recente é o *optimal matching* (pareamento otimizado), que leva em consideração a distância total de todos os pares, podendo refazer pares já formados para otimizar o conjunto total dos pares (GUO; FRASER, 2010).

Após o pareamento, são realizadas as análises nos pares, com algum tipo de regressão ajustada, em que se possam observar as diferenças nos escores das covariáveis entre o grupo de tratamento e o de controle. Uma comparação entre os dois grupos, antes do pareamento, deverá indicar diferenças nas médias das variáveis observáveis. No entanto, após o pareamento, as médias deverão ficar muito próximas entre ambos.

É importante salientar que após o pareamento nenhum grupo, seja ele de controle ou intervenção, possui um desempenho inicial desfavorável. A técnica do escore de propensão e do pareamento engendra um “ponto de partida” comum entre intervenção e controle, sendo comparadas escolas com características semelhantes entre aquelas consideradas no pareamento. Assim, se o incremento da variação acontecer de forma semelhante entre controle e intervenção, isso é um indício de que a política não estaria induzindo a um esforço maior pelo grupo intervenção.

Uma vez formados os grupos de tratamento e controle, por alguma técnica de pareamento, pode-se utilizar o método das diferenças em diferenças (DD) para testar o impacto do tratamento. Em um primeiro momento, calcula-se a diferença das médias, em um intervalo de tempo, para o grupo controle e para o grupo tratamento. Em um segundo momento, faz-se a diferença dessas diferenças. Esse resultado seria o que se pode atribuir ao efeito do programa. Observe um exemplo fictício na Tabela 1.

TABELA 1 - Exemplo fictício do método das diferenças em diferenças considerando a proficiência média de duas redes públicas de ensino

GRUPO	ANO DA AVALIAÇÃO		DIFERENÇA (2012-2008)	DD
	2008	2012		
Escolas da Rede Estadual (Tratamento)	100	180	80	30
Escolas da Rede Municipal (Controle)	100	150	50	

Fonte: Elaboração dos autores.

Nesse exemplo, o impacto do programa seria de 30 pontos pelo método das diferenças em diferenças. Como vantagem, essa técnica lidaria bem com o viés de seleção associado às características não observáveis, principalmente as que não variam muito no decorrer do tempo. Algumas condições devem ser obedecidas ao se utilizar o método DD. De acordo com Foguel (2012), a composição dos grupos não deve se alterar de maneira significativa entre os períodos analisados. Além disso, é necessário garantir que os grupos não sejam afetados de maneira heterogênea por mudanças externas.

PAREAMENTO COM AS BASES DO SAEPE

As etapas escolares avaliadas, e que serão analisadas, correspondem ao 5º e 9º ano do ensino fundamental. Embora o Saepe também avalie o 3º ano do ensino fundamental, essa etapa não entra no cálculo de metas para o recebimento do BDE e, portanto, tais dados não serão aqui analisados. Por outro lado, o 3º ano do ensino médio não apresenta casos suficientes para a formação do grupo controle (rede municipal), logo, permanecerá fora das análises. Essa escolha apoia-se na composição da rede escolar pública de Pernambuco, conforme observado na Tabela 2, que apresenta a distribuição de alunos nos anos avaliados por rede de ensino.

Foram realizados dois pareamentos, um para o 5º ano e outro para o 9º ano do ensino fundamental. Antes de ser feita a regressão para o cálculo do escore de propensão, a base de dados foi organizada com os resultados dos alunos de ambas as redes nas avaliações do Saepe.

TABELA 2 – Distribuição absoluta e relativa dos alunos por série e rede de ensino em Pernambuco (2008-2012)

ANO	5º ANO EF		9º ANO EF		3º ANO EM	
	ESTADUAL	MUNICIPAL	ESTADUAL	MUNICIPAL	ESTADUAL	MUNICIPAL
	N	N	N	N	N	N
	%	%	%	%	%	%
2008	45.440	216.394	98.544	71.418	113.416	7.754
	17,4	82,6	58,0	42,0	93,6	6,4
2009	42.266	220.074	102.922	74.850	109.478	3.466
	16,1	83,9	57,9	42,1	96,9	3,1
2010	59.142	428.896	211.668	147.560	231.620	4.636
	12,1	87,9	58,9	41,1	98,0	2,0
2011	57.924	417.982	211.044	154.526	242.108	3.416
	12,2	87,8	57,7	42,3	98,6	1,4
2012	41.154	410.922	208.040	157.498	247.560	2.068
	9,1	90,9	56,9	43,1	99,2	0,8

Fonte: Elaboração dos autores a partir de microdados do Saepe (PERNAMBUCO, 2012).

As bases continuam, preliminarmente, escolas localizadas tanto na zona urbana quanto na rural. Contudo, as escolas rurais encontravam-se sobremaneira na rede municipal, não permitindo realizar o pareamento adequadamente nesse grupo, o que levou a descartá-las das análises. Além disso, foram também observados os seguintes critérios para incluir uma escola na análise:

- ter ao menos dez alunos em cada ano avaliado (5º e 9º anos);
- possuir proficiência em Língua Portuguesa e Matemática em todas as edições do Saepe entre 2008 e 2012;
- não mudar de rede de ensino entre 2008 e 2012.

A análise por meio da regressão logística que gerou os escores de propensão contou com as seguintes variáveis:

- *BDE* = variável dependente no modelo, admitindo-se o valor “0” para o grupo controle (escolas municipais) e 1 para intervenção (escolas estaduais);
- *PRF_MT_2008* = proficiência de Matemática em 2008;
- *PRF_LP_2008* = proficiência de Língua Portuguesa em 2008;
- *Nº_ALUNOS_ESC* = número de alunos na escola; o dado é relevante, pois o resultado dos alunos tende a ser maior em escolas menores (SOARES, 2004);

- *IDHM* = Índice de Desenvolvimento Humano Municipal,¹ que é um controle importante das diferenças entre os municípios;
- *ISE* = Índice Socioeconômico² dos alunos, que contribui para o controle de diferenças de origem socioeconômica; essa medida, quando agregada por escola, constitui um importante controle na observação das escolas (ALVES; SOARES, 2009).

1 Envolve a transformação das três dimensões por ele contempladas (longevidade, educação e renda) em índices que variam entre 0 (pior) e 1 (melhor) (ORGANIZAÇÃO DAS NAÇÕES UNIDAS, 2016).

2 O cálculo do ISE foi realizado no Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd-UFJF), conforme Anexo I.

Após o cálculo do escore de propensão foi realizado o pareamento das escolas que possuíam o 5º e o 9º ano do ensino fundamental. O método para formar os pares foi o *vizinho mais próximo* (GUO; FRASER, 2010). Essa técnica emprega a menor diferença (mínimos) entre os escores de propensão entre todas as possibilidades de pares entre as unidades do grupo controle e de intervenção. Cada unidade pareada terá um único par. Na formação dos pares, foi observada, ainda, uma distância máxima aceitável entre os escores de propensão (*caliper matching*). Dessa forma, pares com diferenças muito altas foram retirados, evitando, assim, o pareamento de escolas que sejam pouco semelhantes segundo o conjunto das variáveis utilizadas. As Tabelas 3 e 4 apresentam as médias das variáveis, antes e após a realização do processo de pareamento das escolas. Como se pode observar, as diferenças existentes entre os grupos de controle e intervenção antes do processo de pareamento foram reduzidas de forma significativa. Essa observação é válida tanto para o 5º ano, com 137 pares formados (73,3% do total da rede estadual submetida ao pareamento), quanto para o 9º ano do ensino fundamental, com 304 pares formados (49,4% do total da rede estadual submetida ao pareamento).³

3 Conforme a Tabela 2, o fato de a rede municipal de Pernambuco no 9º ano ser menor do que a rede estadual limitou a escolha e a formação dos pares da rede estadual; o mesmo não ocorreu com o 5º ano, cuja oferta da rede municipal é maior, facilitando assim a formação dos pares. Essa diferença também será observada no pareamento com as bases da Prova Brasil.

TABELA 3 – Valores médios das variáveis utilizadas no cálculo dos escores de propensão das escolas públicas com 5º ano do EF, com base no Saepe, antes e depois do pareamento

VARIÁVEIS UTILIZADAS NO CÁLCULO DOS ESCORES DE PROPENÇÃO	ANTES DO PAREAMENTO		APÓS O PAREAMENTO	
	CONTROLE	INTERVENÇÃO	CONTROLE	INTERVENÇÃO
IDHM	0,657	0,696	0,683	0,684
Número de alunos por escola	57,7	50,6	49,9	52,7
ISE	0,086	0,269	0,192	0,202
Proficiência em Língua Portuguesa 2008	163,3	169,8	166,1	166,3
Proficiência em Matemática 2008	166,9	173,1	169,1	169,5

Fonte: Elaboração dos autores a partir dos microdados do Saepe (PERNAMBUCO, 2012).

TABELA 4 – Valores médios das variáveis utilizadas no cálculo dos escores de propensão das escolas públicas com 9º ano do EF, com base no Saepe, antes e depois do pareamento

VARIÁVEIS UTILIZADAS NO CÁLCULO DOS ESCORES DE PROPENÇÃO	ANTES DO PAREAMENTO		APÓS O PAREAMENTO	
	CONTROLE	INTERVENÇÃO	CONTROLE	INTERVENÇÃO
IDHM	0,635	0,672	0,648	0,649
Número de alunos por escola	60,8	67,0	63,1	60,7
ISE	-0,069	0,205	0,027	0,029
Proficiência em Língua Portuguesa 2008	209,3	215,9	211,1	211,1
Proficiência em Matemática 2008	220,8	225,7	221,9	222,0

Fonte: Elaboração dos autores a partir dos microdados do Saepe (PERNAMBUCO, 2012).

PAREAMENTO COM A BASE DA PROVA BRASIL

Conforme relatado anteriormente, as bases do Saepe não contemplaram o ano de 2007, impossibilitando, assim, as análises do ano que antecede a implantação do BDE, que ocorreu em 2008. Dessa forma, foram utilizadas as bases da Prova Brasil de 2007, 2009 e 2011. Como a base da edição de 2007 apresentava apenas escolas urbanas,⁴ esse mesmo recorte foi considerado para 2009 e 2011, pois, ainda que existam escolas rurais em 2009 e 2011, não seria possível acompanhar o desempenho dessas instituições partindo de 2007.

Foram adotados os seguintes critérios para a manutenção das escolas nas análises:

- ter ao menos dez alunos em cada ano avaliado (5º e 9º anos);

⁴ As escolas rurais passaram a ser avaliadas na edição da Prova Brasil de 2009.

- possuir proficiência de Língua Portuguesa e Matemática nas edições de 2007, 2009 e 2011 da Prova Brasil;
- não mudar de rede de ensino entre 2007 e 2011.

Os procedimentos que seguem são muito semelhantes aos realizados nas bases do Saepe e, portanto, serão descritos de maneira sucinta. Na regressão logística para a geração do escore de propensão, foram utilizadas as seguintes variáveis:

- *BDE* = variável dependente no modelo, admitindo-se o valor “0” para o grupo controle (escolas municipais) e 1 para intervenção (escolas estaduais);
- *PRF_MT_2007* = proficiência em Matemática em 2007;
- *PRF_LP_2007* = proficiência em Língua Portuguesa em 2007;
- *Nº_ALUNOS_ESC* = número de alunos na escola;
- *IDHM* = Índice de Desenvolvimento Humano Municipal.

Nesse pareamento não foi empregado o ISE, pois, em 2007, não existia esse indicador nas bases da Prova Brasil. Portanto, a estimação do escore de propensão foi obtida com uma variável a menos em relação ao cálculo do escore de propensão realizado no Saepe.

Após o cálculo do escore de propensão, foi feito o pareamento das escolas. O método para formar os pares foi o mesmo utilizado no pareamento do Saepe. As Tabelas 5 e 6 mostram as médias das variáveis antes e após a realização do processo de pareamento das escolas. As diferenças existentes antes do processo de pareamento foram reduzidas de maneira perceptível. Após o pareamento, estavam formados 238 pares no 5º ano (98,3% do total da rede estadual submetida ao pareamento) e 314 no 9º ano do ensino fundamental (54,7% do total da rede estadual submetida ao pareamento).

TABELA 5 – Valores médios das variáveis utilizadas no cálculo dos escores de propensão das escolas públicas com 5º ano do EF, com base na Prova Brasil, antes e depois do pareamento

VARIÁVEIS UTILIZADAS NO CÁLCULO DOS ESCORES DE PROPENÇÃO	ANTES DO PAREAMENTO		APÓS O PAREAMENTO	
	CONTROLE	INTERVENÇÃO	CONTROLE	INTERVENÇÃO
IDHM	0,654	0,681	0,682	0,679
Número de alunos por escola	53,3	46,4	47,4	46,6
Proficiência em Língua Portuguesa 2007	156,2	161,6	161,3	161,3
Proficiência em Matemática 2007	172,7	176,5	176,5	176,3

Fonte: Elaboração dos autores a partir dos microdados da Prova Brasil (BRASIL, 2011).

TABELA 6 – Valores médios das variáveis utilizadas no cálculo dos escores de propensão das escolas públicas com 9º ano do EF, com base na Prova Brasil, antes e depois do pareamento

VARIÁVEIS UTILIZADAS NO CÁLCULO DOS ESCORES DE PROPENÇÃO	ANTES DO PAREAMENTO		APÓS O PAREAMENTO	
	CONTROLE	INTERVENÇÃO	CONTROLE	INTERVENÇÃO
IDHM	0,631	0,668	0,637	0,638
Número de alunos por escola	60,8	63,5	60,9	58,6
Proficiência em Língua Portuguesa 2007	208,4	213,0	209,6	209,4
Proficiência em Matemática 2007	219,8	222,3	220,1	220,0

Fonte: Elaboração dos autores a partir dos microdados da Prova Brasil (BRASIL, 2011).

RESULTADOS

Após a formação dos pares de escolas foi aplicado o método das diferenças em diferenças, nas análises tanto do Saepe quanto da Prova Brasil, para verificar a efetividade do BDE. Dessa forma, caso o BDE tenha tido algum impacto no desempenho dos alunos, esse “ganho” seria superior e favorável ao grupo de intervenção (rede estadual) nas séries históricas analisadas.

RESULTADOS PARA O 5º ANO DO ENSINO FUNDAMENTAL

No tocante aos pareamentos realizados no 5º ano do ensino fundamental, conforme podemos observar na Tabela 7 (base do Saepe), há uma pequena diferença das médias favorável ao grupo de intervenção. Contudo, o teste *t* apontou diferenças significativas ($p < 0,05$) apenas nos intervalos

entre 2008 e 2011 em Matemática, que apresentaram um ganho médio em torno de 4 pontos do grupo de tratamento *vis-à-vis* o grupo controle. Em Língua Portuguesa, o intervalo entre 2008 e 2012 apresentou um ganho médio em torno de 3 pontos. No entanto, são resultados muito tímidos se consideramos esse intervalo de quatro anos.

TABELA 7 - Resultados da análise de diferenças em diferenças para as escolas públicas urbanas com 5º ano do EF, com base no Saepe (137 pares)

INTERVALO	CONTROLE		INTERVENÇÃO		DIFERENÇAS DAS MÉDIAS	I.C. 95%	P-VALOR
	MÉDIA DAS DIFERENÇAS	D.P. DAS DIFERENÇAS	MÉDIA DAS DIFERENÇAS	D.P. DAS DIFERENÇAS			
MATEMÁTICA							
2008-2009	8,49	12,46	10,16	13,97	1,67	[-1,24 ; 4,57]	0,2594
2008-2010	7,73	14,91	9,79	19,74	2,06	[-1,88 ; 6,00]	0,3032
2008-2011	19,65	14,56	23,76	16,72	4,10	[0,54 ; 7,66]	0,0243*
2008-2012	21,56	16,83	24,87	15,97	3,31	[-0,46 ; 7,08]	0,0850
LÍNGUA PORTUGUESA							
2008-2009	1,67	10,22	2,11	10,59	0,43	[-1,97 ; 2,84]	0,7226
2008-2010	2,66	11,45	4,74	16,11	2,08	[-1,10 ; 5,25]	0,1980
2008-2011	8,63	12,10	11,27	13,46	2,64	[-0,31 ; 5,58]	0,0789
2008-2012	9,13	14,36	12,36	14,25	3,23	[0,10 ; 6,35]	0,0430*

Fonte: Elaboração dos autores a partir dos microdados do Saepe (PERNAMBUCO, 2012).

* Valor de p significativo até 0,05.

Com base nos resultados das bases do Saepe para o 5º ano, encontrados no limite, as mesmas análises foram feitas com a utilização das bases da Prova Brasil. Isso poderia trazer alguma informação, tendo em vista que seria possível captar os resultados do ano de 2007, antecedente à implantação do BDE. Entretanto, mesmo considerando a transição do ano de 2007 para 2009, os resultados (conforme demonstrado na Tabela 8) ficaram muito próximos dos encontrados com a base do Saepe, sendo favorável ao grupo de intervenção, com uma diferença das médias em torno de 5 pontos para Matemática no período de 2007 a 2011. Resultados mais tímidos,

com valores em torno de 3 pontos, foram observados em Língua Portuguesa no mesmo intervalo de tempo.

TABELA 8 - Resultados da análise de diferenças em diferenças para as escolas públicas urbanas com 5º ano do EF, com base na Prova Brasil (238 pares)

INTERVALO	CONTROLE		INTERVENÇÃO		DIFERENÇAS DAS MÉDIAS	I.C. 95%	P-VALOR
	MÉDIA DAS DIFERENÇAS	D.P. DAS DIFERENÇAS	MÉDIA DAS DIFERENÇAS	D.P. DAS DIFERENÇAS			
MATEMÁTICA							
2007-2009	6,91	14,72	8,80	14,75	1,89	[-0,68 ; 4,47]	0,1489
2007-2011	9,79	14,89	14,63	17,29	4,84	[2,24 ; 7,44]	0,0003*
LÍNGUA PORTUGUESA							
2007-2009	4,58	12,88	4,75	12,56	0,17	[-2,05 ; 2,39]	0,8813
2007-2011	8,45	13,72	11,42	14,96	2,97	[0,61 ; 5,32]	0,0138*

Fonte: Elaboração dos autores a partir dos microdados da Prova Brasil (BRASIL, 2011).

* Valor de p significativo até 0,05.

Portanto, os resultados dos pareamentos para o 5º ano do ensino fundamental foram positivos apenas se consideramos as diferenças entre o início e o fim das séries históricas, tanto pelo Saepe quanto pela Prova Brasil. É importante destacar que, na transição entre o período que não existia o BDE, ou seja, de 2007 para 2009, as diferenças não foram estatisticamente significativas. Outro ponto relevante é que o uso da base de dados do Saepe, que foi empregada nos cálculos do BDE, ao que parece, não provocou uma inflação de escores quando comparada à da Prova Brasil. Tal defesa justifica-se pela proximidade dos resultados obtidos entre as duas bases (Saepe e Prova Brasil). Esse achado torna indiferente a ideia de que seria melhor utilizar testes *low-stakes* para evitar a inflação de escore no estudo de impacto de uma política de responsabilização (HOUT; ELLIOTT, 2011). Pelo menos, no caso do 5º ano, os resultados aproximados indicam que não fez muita diferença utilizar uma base originada de um teste *low-stakes* ou de um teste *high-stakes*.

RESULTADOS PARA O 9º ANO DO ENSINO FUNDAMENTAL

Passando para as análises do 9º ano do ensino fundamental, temos também dois pareamentos: o primeiro com a base do Saepe e o segundo com a base da Prova Brasil.

A Tabela 9 apresenta os resultados realizados com a base do Saepe, podendo-se observar que as diferenças das médias foram favoráveis, em torno de 5 pontos, ao grupo de intervenção em todos os intervalos de tempo analisados. Além disso, os valores do teste *t* revelaram-se bastante significativos em toda a série histórica.

Todavia, no 9º ano, os ganhos gerados no intervalo inicial (2008-2009) praticamente se mantiveram. Isso também ocorre em períodos de maior amplitude, como no de 2008 a 2012, revelando que o BDE não conseguiu, no decorrer dos anos, elevar os ganhos para o grupo de intervenção e permanecendo, praticamente, no mesmo patamar do início do programa.

TABELA 9 – Resultados da análise de diferenças em diferenças para as escolas públicas urbanas com 9º ano do EF, com base no Saepe (304 pares)

INTERVALO	CONTROLE		INTERVENÇÃO		DIFERENÇAS DAS MÉDIAS	I.C. 95%	P-VALOR
	MÉDIA DAS DIFERENÇAS	D.P. DAS DIFERENÇAS	MÉDIA DAS DIFERENÇAS	D.P. DAS DIFERENÇAS			
MATEMÁTICA							
2008-2009	0,71	11,26	5,56	12,99	4,85	[2,88 ; 6,83]	< 0,0001
2008-2010	0,13	13,35	5,85	14,91	5,73	[3,53 ; 7,92]	< 0,0001
2008-2011	5,52	14,92	10,51	13,74	4,99	[2,81 ; 7,17]	< 0,0001
2008-2012	6,97	14,64	11,26	14,52	4,29	[1,95 ; 6,63]	< 0,0001
LÍNGUA PORTUGUESA							
2008-2009	3,27	11,04	7,95	11,28	4,68	[2,88 ; 6,48]	< 0,0001
2008-2010	4,40	12,25	10,14	12,20	5,74	[3,82 ; 7,67]	< 0,0001
2008-2011	5,81	14,82	11,14	12,80	5,32	[3,18 ; 7,47]	< 0,0001
2008-2012	10,80	14,73	15,97	13,51	5,17	[2,94 ; 7,40]	< 0,0001

Fonte: Elaboração dos autores a partir dos microdados do Saepe (PERNAMBUCO, 2012).

* Valor de *p* significativo até 0,05.

Também foi realizado o pareamento do 9º ano com as bases da Prova Brasil. De maneira análoga às análises do 5º ano, tal procedimento tem, como já foi dito, a função de comparar os resultados das análises, de modo a garantir resultados mais fidedignos e evitar uma possível inflação de resultados no Saepe.

Dessa forma, a Tabela 10 traz os resultados do pareamento do 9º ano na base da Prova Brasil. As diferenças das médias nos períodos avaliados ficaram em torno de 5 pontos; a exceção foi Matemática que, entre 2007 e 2009, apresentou uma diferença de 4 pontos. Resumidamente, os resultados encontrados, nessa análise, foram muito parecidos com os do Saepe, sendo que, igualmente aqui, não fez diferença utilizar uma avaliação com desenho *low-stakes* (Prova Brasil) ou *high-stakes* (Saepe) e não ocorreu uma inflação de escores na base *high-stakes*.

O efeito inicial também foi mantido em toda a série histórica, sendo que apenas a disciplina de Matemática apresentou uma discreta variação, de 1 ponto nas diferenças das médias entre o intervalo de 2007 a 2009 (4,05) e o de 2007 a 2011 (5,19). Ou seja, aqui também ocorreu uma estabilização dos ganhos gerados pelo programa de bonificação. A hipótese é que o estímulo provocado tenha alcançado um ponto de saturação, de maneira que os indivíduos submetidos à influência dessa política, uma vez que tenham empreendido esforços após a implantação do BDE (2008), não tiveram uma motivação extra para elevação dos esforços nos anos seguintes.

TABELA 10 – Resultados da análise de diferenças em diferenças para as escolas públicas urbanas com 9º ano do EF, com base na Prova Brasil (314 pares)

INTERVALO	CONTROLE		INTERVENÇÃO		DIFERENÇAS DAS MÉDIAS	I.C. 95%	P-VALOR
	MÉDIA DAS DIFERENÇAS	D.P. DAS DIFERENÇAS	MÉDIA DAS DIFERENÇAS	D.P. DAS DIFERENÇAS			
MATEMÁTICA							
2007-2009	3,01	14,03	7,07	11,80	4,05	[2,10 ; 6,00]	< 0,0001
2007-2011	4,76	14,78	9,95	13,33	5,19	[3,00 ; 7,38]	< 0,0001
LÍNGUA PORTUGUESA							
2007-2009	10,29	15,75	15,69	12,65	5,40	[3,21 ; 7,58]	< 0,0001
2007-2011	8,90	0,84	14,61	12,98	5,71	[3,68 ; 7,73]	< 0,0001

Fonte: Elaboração dos autores a partir dos microdados da Prova Brasil (BRASIL, 2011).

* Valor de p significativo até 0,05.

CONSIDERAÇÕES FINAIS

Respondendo à questão inicial deste trabalho, pode-se dizer que, com base nas análises dos pareamentos realizados tanto pela base do Saepe quanto pela Prova Brasil, os resultados foram semelhantes e positivos; além disso, corroboram na mesma direção de estudos anteriores (FERRAZ; BRUNS, 2010; BRUNS; FERRAZ, 2011) e, portanto, existem boas chances de a política de bonificação ter sido responsável pela variação positiva, ou ainda pela maior parte dos ganhos de desempenho da rede estadual no período analisado.

Os resultados positivos ocorreram de maneira mais tímida no 5º ano, com teste de significância validando apenas os intervalos com distâncias maiores de tempo (no caso do Saepe, entre 2008 e 2012, e no da Prova Brasil, entre 2007 e 2011). Isso, provavelmente, indica que a bonificação produziu um efeito menor nos resultados do 5º ano do EF.

As análises dos pareamentos do 9º ano do ensino fundamental indicaram ganhos favoráveis ao grupo de intervenção, variando em torno de 5 pontos, e com teste de significância validado em todos os intervalos analisados. Assim, por meio do pareamento, pode-se dizer que, desde o início do programa, em 2008, a bonificação demonstrou ganhos favoráveis ao grupo de intervenção.

Entretanto, esses ganhos não foram sendo incrementados no decorrer da série histórica; ao que parece, a melhoria do desempenho induzida pelo bônus, de maneira inicial, não se apresenta constante ao longo do tempo; uma vez alcançado um determinado patamar de desempenho, o desafio da escola seria manter-se nele. Dessa forma, cabe aos gestores da bonificação avaliar se o retorno de no máximo 5 pontos justificaria a continuação dessa política de bonificação no estado de Pernambuco.

Portanto, uma alternativa seria engendrar mecanismos que motivem mais os professores das escolas que não atingiram o bônus e, paralelamente, que o estado forneça suporte adequado para as escolas em maior dificuldade, seja melhorando sua infraestrutura ou, ainda, investindo na capacitação dos professores, algo já previsto no documento do Programa de Modernização da Gestão (PERNAMBUCO, 2008b).

REFERÊNCIAS

ALVES, Maria Teresa G.; SOARES, José Francisco. Medidas de nível socioeconômico em pesquisas sociais: uma aplicação aos dados de uma pesquisa educacional. *Opinião Pública*, Campinas, SP, v. 15, n. 1, p. 1-30, jun./2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-62762009000100001&lng=pt&nrm=iso>. Acesso em: 1 jul. 2013.

AU, Wayne. High-stakes testing and curricular control: a qualitative meta synthesis. *Educational Researcher*, Thousand Oaks, CA, v. 36, n. 5, p. 258-267, Aug. 2007. Disponível em: <<http://edr.sagepub.com/content/36/5/258>>. Acesso em: 6 jul. 2013.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *SAEB – 2005 primeiros resultados: médias de desempenho do SAEB/2005 em perspectiva comparada*. Brasília, DF: Inep, 2007. Disponível em: <http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/resultados/SAEB1995_2005.pdf>. Acesso em: 16 ago. 2013.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Microdados Prova Brasil*. Brasília, DF: Inep, 2011. Disponível em: <<http://portal.inep.gov.br/basica-levantamentos-acessar>>. Acesso em: 16 ago. 2013.

BROOKE, Nigel. Responsabilização educacional no Brasil. *Revista Iberoamericana de Evaluación Educativa*, Madrid, v. 1, n. 1, p. 93-109, 2008.

BROOKE, Nigel. As novas políticas de incentivo salarial para professores: uma avaliação. In: ENCONTRO DE PESQUISA EM EDUCAÇÃO DA REGIÃO SUDESTE, 10., 2011, Rio de Janeiro. *Políticas públicas, movimentos sociais: desafios à pós-graduação em educação em suas múltiplas dimensões*. Rio de Janeiro: Anped, 2011. Disponível em: <<http://www.fe.ufrj.br/anpedinha2011/ebook3.pdf>>. Acesso em: 8 ago. 2013.

BROOKE, Nigel. Políticas estaduais de responsabilização: buscando o diálogo. In: BAUER, Adriana; GATTI, Bernardete A. (Org.). *Ciclo de debates: vinte e cinco anos de avaliação de sistemas educacionais no Brasil*. Florianópolis: Insular, 2013. v. 2.

BROOKE, Nigel et al. *A avaliação externa como instrumento da gestão educacional nos estados*. Belo Horizonte: Game/FAE/UFMG, 2011.

BRUNS, Barbara; EVANS, David; LUQUE, Javier. *Achieving world-class education in Brazil: the next agenda*. Directions in development: human development. Washington, D.C.: The World Bank, 2012.

BRUNS, Barbara; FERRAZ, Claudio. *Paying teachers to perform: the effects of bonus pay in Pernambuco, Brazil*. Washington, D.C.: The World Bank, December 2011.

BRUNS, Barbara; FILMER, Deon; PATRINOS, Harry Anthony. *Making schools work: new evidence on accountability reforms*. Human development perspectives. Washington D.C.: The World Bank, 2011.

CORVALÁN, Javier. Accountability educacional: rendición de cuentas más responsabilización por los procesos y resultados de la educación. In: CORVALÁN, Javier; MCMEEKIN, Robert W. (Ed.). *Accountability educacional: posibilidades y desafíos para América Latina a partir de la experiencia internacional*. Santiago: Preal/Cide, 2006.

FERRAZ, Cláudio. Sistemas educacionais baseados em desempenho, metas de qualidade e a remuneração de professores: os casos de Pernambuco e São Paulo. In: VELOSO, Fernando (Org.). *Educação básica no Brasil: construindo o país do futuro*. Rio de Janeiro: Elsevier, 2009.

FERRAZ, Claudio; BRUNS, Barbara. Incentivando a qualidade: avaliando os efeitos de um bônus por desempenho na educação. In: WORKSHOP AVALIAÇÃO DE IMPACTO-SIEF, 1., 26-30 de abril de 2010, Rio de Janeiro. *Curso...* Rio de Janeiro: Banco Mundial, 2010. Disponível em: <http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1256762343506/6518748-1273071334136/7041933-1273164045906/17.b.Pagamento_por_desempenho_Ferraz_PORT.pdf>. Acesso em: 3 jul. 2014.

FOGUEL, Miguel N. Modelos de resultados potenciais. In: MENEZES FILHO, Naércio. (Org.). *Avaliação econômica de projetos sociais*. São Paulo: Fundação Itaú Social; Dinâmica, 2012.

FREITAS, Luiz Carlos de. Os reformadores empresariais da educação: da desmoralização do magistério à destruição do sistema público de educação. *Educação & Sociedade*, Campinas, SP, v. 33, n. 119, p. 353-377, abr./jun. 2012. Disponível em: <<http://www.cedes.unicamp.br>>. Acesso em: 12 ago. 2013.

GUISBOND, Lisa et al. A década de progresso educativo perdida sob a NCLB: que lições tirar desse fracasso político? *Educação & Sociedade*, Campinas, SP, v. 33, n. 119, p. 353-377, abr./jun. 2012. Disponível em: <<http://www.cedes.unicamp.br>>. Acesso em: 12 ago. 2013.

GUO, Shenyang Y.; FRASER, Mark W. *Propensity score analysis: statistical methods and applications*. California: Sage, 2010.

HOUT, Michael; ELLIOTT, Stuart W. *Incentives and test-based accountability in education*. Washington, D.C.: National Academy, 2011.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Programa das Nações Unidas para o Desenvolvimento. *O que é IDHM?* 2016. Disponível em: <http://www.pnud.org.br/IDH/IDHM.aspx?indiceAccordion=0&li=li_IDHM>. Acesso em: 20 jul. 2014.

OSHIRO, Cláudia Hiromi; SCORZAFAVE, Luiz Guilherme. Efeito do pagamento de bônus aos professores sobre a proficiência escolar no Estado de São Paulo. In: ENCONTRO NACIONAL DE ECONOMIA, 39., 2011, Foz do Iguaçu. *Anais...* Foz do Iguaçu: Anpec, 2011.

PERNAMBUCO. Lei n. 13.486, de 01 de julho de 2008. Institui o Bônus de Desempenho Educacional – BDE, no âmbito do Estado de Pernambuco, e dá outras providências. *Diário Oficial*, Recife, ano LXXXIV, n. 1717, 09 de setembro de 2008a.

PERNAMBUCO. Secretaria de Educação. *Programa de Modernização da Gestão Pública*. 2008b. Disponível em: <<http://www.educacao.pe.gov.br/portal/?pag=1&men=69>>. Acesso em: 11 maio 2011.

PERNAMBUCO. Secretaria de Educação. *Boletim Contextual de Avaliação da Educação – SAEPE 2008: fatores contextuais associados ao desempenho escolar*. Juiz de Fora: UFJF/CAEd, 2008c. Disponível em: <http://www.saepe.caedufjf.net/wp-content/uploads/2016/05/BOLETIM_CONTEXTUAL_SAEPE_2008.pdf>. Acesso em: 19 ago. 2013.

PERNAMBUCO. Secretaria de Educação. *Microdados SAEPE*. 2012.

PERNAMBUCO. Assessoria de Comunicação do Governo do Estado. *Profissionais da educação recebem bônus por desempenho em Pernambuco*. Recife: SEE, 2013. Disponível em: <<http://www.educacao.pe.gov.br/portal/?pag=1&cat=37&art=1351>>. Acesso em: 4 ago. 2014.

PINTO, Cristine C. X. Pareamento. In: MENEZES FILHO, Naércio (Org.). *Avaliação econômica de projetos sociais*. São Paulo: Fundação Itaú Social; Dinâmica, 2012.

PONTUAL, Teresa C. *Remuneração por mérito, desafio para a educação*. São Paulo: Fundação Lemann, 2008. Disponível em: <[http://www.fundacaolemann.org.br/arquivos/uploads/arquivos/Remunera%C3%A7%C3%A3o%20por%20m%C3%A9rito%20\(paper\)%20\(Tereza%20Pontual\).pdf](http://www.fundacaolemann.org.br/arquivos/uploads/arquivos/Remunera%C3%A7%C3%A3o%20por%20m%C3%A9rito%20(paper)%20(Tereza%20Pontual).pdf)>. Acesso em: 19 ago. 2013.

RAVITCH, Diane. *Vida e morte do grande sistema escolar americano: como os testes padronizados e o modelo de mercado ameaçam a educação*. Porto Alegre: Sulina, 2011.

ROSENBAUM, Paul R.; RUBIN, Donald B. The central role of the propensity score in observational studies for causal. *Biometrika*, Oxford, v. 70, n. 1, p. 41-55, Apr. 1983.

SEGATTO, Catarina Ianni. *Como ideias se transformam em reformas: um estudo comparativo das mudanças educacionais orientadas pelo desempenho nos estados brasileiros*. 2011. 134 f. Dissertação (Mestrado em Administração Pública e Governo) – Escola de Administração de Empresas, Fundação Getúlio Vargas, São Paulo, 2011.

SOARES, José F. O efeito da escola no desempenho cognitivo de seus alunos. *REICE – Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, Madrid, v. 2, n. 2, p. 83-104, jul./dic. 2004. Disponível em: <<http://www.ice.deusto.es/rinace/reice/vol2n2/Soares.pdf>>. Acesso em: 5 ago. 2015.

SOARES, José F. Melhoria do desempenho cognitivo dos alunos do ensino fundamental. *Cadernos de Pesquisa*, São Paulo, v. 37, n. 130, p. 135-160, jan./abr. 2007.

SOARES, Tufi M. Utilização da teoria da resposta ao item na produção de indicadores socioeconômicos. *Pesquisa Operacional*, Rio de Janeiro, v. 25, n. 1, p. 83-112, jan./abr. 2005.

SOARES, Tufi M. A expectativa do professor e o desempenho dos alunos. *Psicologia: Teoria e Pesquisa*, Brasília, DF, v. 26, n. 1, p. 157-170, jan./mar. 2010.

THE HUNT INSTITUTE. Cumprindo a promessa da reforma baseada em padrões. In: BROOKE, Nigel (Org.). *Marcos históricos na reforma da educação*. Belo Horizonte: Fino Traço, 2012.

UNITED STATES OF AMERICA. Department of Education. Office of elementary and secondary education. *No Child Left Behind: a desktop reference*. Washington, D.C.: Department of Education, 2002.

CLAYTON SIRILO DO VALLE FURTADO

Analista de Medidas Educacionais no Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF), Juiz de Fora, Minas Gerais, Brasil
clayton@caed.ufff.br

TUFI MACHADO SOARES

Professor do Departamento de Estatística e do Programa de Pós-Graduação em Educação da Universidade Federal de Juiz de Fora (UFJF). Coordenador da Unidade de Pesquisa do Centro de Políticas Públicas e Avaliação da Educação (CAEd) da UFJF, Juiz de Fora, Minas Gerais, Brasil
tufi@caed.ufff.br

ANEXO I - CÁLCULO DO ISE

O cálculo do Índice Socioeconômico dos alunos foi realizado no *software* Parscale utilizando a metodologia da TRI (Teoria da Resposta ao Item); existe literatura específica de suporte para cálculos semelhantes em Soares (2005). No caso de Pernambuco, a estimação dos valores do ISE tomou as variáveis dos questionários dos alunos, a maior parte de posse de bens e acesso aos serviços básicos, mas também foram incluídas questões acerca da escolaridade dos pais, compreendendo um total de 15 questões descritas no quadro a seguir.

QUADRO 1 - Relação dos itens do questionário dos alunos do Saepe utilizados no cálculo do ISE

QUESTÃO	OPÇÕES
06. Até que série sua mãe ou a responsável por você estudou?	1 - Nunca estudou.
	2 - 1ª a 4ª série do ensino fundamental (1º Grau)
	3 - 5ª a 8ª série do ensino fundamental (1º Grau)
	4 - 1ª a 3ª série do ensino médio (2º Grau)
	5 - Até a faculdade (ensino superior)
07. Até que série seu pai ou o responsável por você estudou?	1 - Nunca estudou
	2 - 1ª a 4ª série do ensino fundamental (1º Grau)
	3 - 5ª a 8ª série do ensino fundamental (1º Grau)
	4 - 1ª a 3ª série do ensino médio (2º Grau)
	5 - Até a faculdade (ensino superior)
08. A rua em que você mora tem calçamento?	1 - Sim; 2- Não
10. Sua casa tem energia elétrica?	1 - Sim; 2- Não
11. Sua casa possui água na torneira?	1 - Sim; 2- Não
12. Sua família recebe bolsa família?	1 - Sim; 2- Não
13. Banheiro?	1 - Nenhum; 2- Um; 3- Dois ou mais
14. Rádio?	1 - Nenhum; 2- Um; 3- Dois ou mais
15. Geladeira?	1 - Nenhum; 2- Um; 3- Dois ou mais
16. Televisão em cores?	1 - Nenhum; 2- Um; 3- Dois ou mais
17. Máquina de lavar roupa?	1 - Nenhum; 2- Um; 3- Dois ou mais
18. DVD?	1 - Nenhum; 2- Um; 3- Dois ou mais
19. Automóvel?	1 - Nenhum; 2- Um; 3- Dois ou mais
20. Além dos livros didáticos, quantos livros há em sua casa? (Não conte jornais, revistas ou gibis)	1 - 1 a 20 livros
	2 - 21 a 100 livros
	3 - Mais de 100 livros
	4 - Nenhum
21. Em sua casa tem computador?	1 - Sim, com acesso à Internet
	2 - Sim, sem acesso à Internet
	3 - Não

Fonte: Elaboração dos autores a partir do Boletim Contextual do Saepe (PERNAMBUCO, 2008c, p. 26).

Recebido em: 08 JUNHO 2016

Aprovado para publicação em: 27 OUTUBRO 2016

