**Carta ao Avaliador 1**

Prezados Editor e Avaliador,

Em primeiro lugar, agradecemos imensamente a oportunidade de ter nosso artigo avaliado pela Pesquisa e Planejamento Econômico (PPE), bem como as importantes contribuições do parecer. Acatamos os pontos sugeridos pelo avaliador, e com base neles produzimos uma versão revisada. As modificações feitas encontram-se explanadas abaixo:

**Comentários**

**1) Literatura:**

**1.1 Separar nessa frase quais autores avaliam Pernambuco e quais São Paulo. Pag 3: “com os resultados obtidos nos estudos de Furtado e Soares (2015), Lépine (2017) e Oshiro, Scorfave e Dorigan (2015) que encontraram efeitos positivos dos programas de bonificação de Pernambuco e São Paulo sobre as notas médias dos alunos.”**

**Resposta:** acatamos a recomendação do avaliador e demarcamos os autores que avaliaram Pernambuco e São Paulo (página 4, 2º parágrafo).

**1.2 O nome do autor está incorreto em algumas citações “Scorzafave” e outras “Scorfave”.**

**Resposta:** a grafia correta é Scorzafave. Uniformizamos ao longo do texto as citações a esse artigo.

**1.3 Colocar aspas no trecho que é uma citação ou retirar as marcações que remetem à citação [...]. Pag 5: “Assim, nessa abordagem a ênfase recai sobre análises de custos e benefícios privados e sociais [...], análises de eficiência na alocação de recursos [...], definição de objetivos escolares e pós-escolares de sistemas educativos, avaliação de atingimento desses objetivos. (WALTENBENG, 2006, p. 119).”**

**Resposta:** recomendação acatada.

**1.4 Falta a citação neste trecho que tem como referência ALEXANDRE, LIMA, WALTENBERG, 2014. Pag 7: “Portanto, é possível que um professor tenha total engajamento e esforço para melhorar o aprendizado de um determinado grupo de alunos sem que esse esforço se reverta efetivamente em bons resultados. Logo, apesar de agir conforme o desejável pelo principal (secretaria escolar), o professor não receberia seu bônus, ocasionando redução de motivação futura. O oposto (pouco esforço, bons resultados e recebimento de bônus) também poderia reduzir a motivação futura. Destarte, em ambos os casos, a legitimidade do programa poderia ser minada.”**

**Resposta:** a citação foi incluída, conforme o recomendado.

**1.5 Sobre a primeira especificidade, um aspecto importante é bonificação no setor público entra no sentido de convergir os interesses dos agentes, segundo a fonte citada (ALEXANDRE, LIMA, WALTENBERG, 2014) e seria interessante que isso fosse destacado pois não ficou claro neste trecho da Pag 8 : “A inexistência de pressão por lucro no setor público torna menos evidente o objetivo comum a ser alcançado pelos agentes envolvidos em tais esferas. Dessa forma, o pagamento por desempenho de alunos tende a não se disseminar naturalmente nesses espaços e o alinhamento entre os interesses de principal e agente dificultados.”**

**Resposta:** reescrevemos o trecho citado, de maneira a explicitar as diferenças de incentivos entre funcionários de escolas privadas e de escolas públicas.

**1.6 A abordagem empírica na seção 2.1 está baseada majoritariamente em um artigo do tema (ALEXANDRE, LIMA, WALTENBERG, 2014). A recomendação é que outras fontes pudessem também estar abarcadas nesta seção.**

**Resposta:** foram incluídas outras referências bibliográficas na seção.

**1.7 Comentários: Ambos os artigos (Lapine e Oshiro, Scorzafave & Dorigan utilizaram a Prova Brasil. Pag 14: “Diferentemente do observado em Oshiro, Scorzafave e Dorigan (2015), a autora (Lépine) verifica o efeito do programa sobre um teste padronizado não diretamente relacionado ao esquema de incentivo, a Prova Brasil.”**

**Resposta:** revisamos a seção e reescrevemos muitos trechos e, assim corrigimos algumas informações.

**2 O Programa Escola de Valor:**

**2.1 Explicar melhor o programa. Apesar do entendimento que houve modificações de um ano para outro e que as informações são pouco transparentes (ou até por esse motivo), seria bastante útil uma sistematização em formato de tabela resumo as mudanças ocorridas durante vários anos.**

**2.2 Também poderia ser feita alguma exploração do contexto do programa e do histórico do estado da Paraíba.**

**2.3 Não está claro em que momento o incentivo é estabelecido pelo governo do estado, em que período do ano é lançado o edital, quando ocorre a inscrição e a avaliação.**

**Resposta para os pontos 2.1, 2.2 e 2.3:** procuramos inserir maiores informações sobre o programa, embora a falta de transparência em alguns aspectos foi limitante de uma melhor descrição do prêmio, como, por exemplo, se os critérios são de fato observados na concessão das bonificações. Incluímos no texto um histórico do PEV e como suas características se alteraram ao longo dos anos (Quadro 1).

**2.4 Na pag 17 afirma que “Quanto ao relatório de execução do PIP, ele deve estar alinhado com o Plano de Metas do IDEPB do presente ano e estar devidamente respaldado com provas documentais que comprovem os resultados obtidos pela escola naquele período, enfatizando principalmente o crescimento dos resultados no IDEPB em relação ao ano imediatamente anterior e a participação mínima dos estudantes na avaliação do IDEPB no ano.”. Na mesma página, em frase seguinte diz que o prêmio não está diretamente vinculado ao desempenho “Dessa forma, o recebimento do Prêmio Escola de Valor pelas escolas inscritas não é diretamente vinculado ao desempenho da escola nas avaliações externas...”. Aqui seria mais apropriado dizer que não está vinculado “apenas” ao desempenho.**

**2.5 Ainda sobre o comentário anterior, na Pag 28 afirma que “Contudo, visando utilizar indicadores confiáveis e mensuráveis, serão utilizados a proficiência média dos alunos no SAEB nas disciplinas de Língua Portuguesa e Matemática, a taxa de aprovação escolar e a taxa de abandono escolar referentes às séries/anos finais das primeira e segunda fases do ensino fundamental (5º e 9º anos). Esses indicadores foram escolhidos devido ao relacionamento direto que apresentam com o objetivo fim do programa, a saber, a melhora do desempenho escolar.”**

**Resposta para os itens 2.4 e 2.5:** de acordo com as informações às quais obtivemos acesso, notas são atribuídas às escolas de acordo com o IDEPB, um indicador local de desenvolvimento da educação, cuja construção engloba diversos critérios de avaliação e não somente o desempenho acadêmico. A forma como se dá a avaliação da escola, conforme falamos, não está clara. As notas das escolas não foram sequer divulgadas pela Secretaria de Educação. Além disso, a literatura sugere a não utilização de indicadores locais, como forma de evitar a manipulação de resultado por parte da escola que está sendo avaliada. Sendo assim, procuramos utilizar como variável de impacto do programa as proficiências obtidas no SAEB, um indicador nacional, bastante utilizado na literatura econômica de educação, e que mede exclusivamente o rendimento escolar. Procuramos deixar essa justificativa clara no texto do artigo.

**3 Metodologia:**

**3.1 A parte da metodologia está bastante extensa para um método já bastante conhecido. Não requer muito detalhamento, sugiro um texto mais breve, com a citação das fontes das metodologias e da contextualização ao tema do artigo.**

**Resposta:** procuramos tornar a seção mais enxuta, focando em informações mais essenciais.

**3.2 Na Pag 19 falta uma vírgula em 𝑖 (𝑌𝑖(1), 𝑌𝑖(0))**

**Resposta:** o erro apontado foi corrigido.

**3.3 No caso de DD, é importante que se apresente algum gráfico para demonstrar as tendências paralelas dos grupos de controle e tratamento antes da intervenção.**

**3.4 Explicar fontes que comprovem esta afirmação de relaxamento da hipótese de tendências paralelas. Pag 26: “Utilizou-se da técnica do pareamento aliada ao método de diferenças e diferenças no intuito de homogeneizar os grupos de tratados e controles no período pré-programa por meio do matching e, assim, permitir o relaxamento da hipótese de trajetórias paralelas do método de DD.”**

**Resposta para os itens 3.3 e 3.4:** o gráfico de evolução no tempo da variável dependente para os grupos de tratamento e controle, como forma de demonstrar a existência de tendências paralelas, depende de observações em um largo espaço de tempo para que seja possível identificar uma tendência de comportamento semelhante entre os grupos. No caso do nosso artigo, onde mensuramos os efeitos da premiação ocorrida em 2016 (apresentamos as motivações no artigo e aqui no parecer), os dados do SAEB anteriores a este ano correspondem apenas aos anos de 2015, 2013 e 2011. O SAEB como indicador de rendimento escolar é de fato consolidado na década de 2010, de modo que antes de 2011 os últimos microdados do SAEB disponíveis no site do Inep datam do ano de 2005, um lapso de tempo de 6 anos. Se considerássemos o ano de implementação do PEV (2012) como o ano da intervenção, os dados pré-tratamento seriam ainda mais escassos. Por este motivo, acreditamos que um gráfico das notas das escolas para apenas três anos não seria suficiente para mostrar a existência ou não de trajetórias paralelas das escolas tratadas e não tratadas.

Sendo assim, procuramos justificar a validade da hipótese por meio da complementaridade entre os métodos de pareamento e de diferenças em diferenças. Este método foi proposto por Heckman, Ichimura e Todd (1997,1998) e permite substituir algumas hipóteses de ambas as técnicas por hipóteses mais fracas. Por exemplo, a hipótese de seleção nos observáveis estabelece que, condicional ao vetor de variáveis observáveis X, não pode existir nenhum fator não-observável que influencie simultaneamente a decisão de participar ou não no tratamento e os resultados potenciais. Ao combinar o pareamento com o método de diferenças em diferenças, podemos permitir que fatores não-observáveis que sejam constantes ao longo do tempo influenciem simultaneamente a decisão de participar e os resultados potenciais, mesmo controlando pelo vetor de variáveis observáveis X. Por outro lado, o pareamento permite tornar os grupos de tratado e controle o mais parecidos possível em termos de variáveis observáveis, a ponto de aumentar as chances de que ambos os grupos apresentem trajetórias pré-tratamento semelhantes da variável de resultado.

Incluímos esta explicação, bem como a referência citada no corpo do texto, conforme recomendação.

**4 Aplicação do método**

**4.1 O método de DD requer que sejam comparados os resultados das variáveis de interesse Antes e Depois da intervenção. O autor utilizou os dados de 2015 (antes) e 2017 (depois) para a comparação. Contudo, a lei que institui o Programa Escola de Valor é de 2012. Ou seja, desde 2012 existe um incentivo para que escolas estaduais do estado da Paraíba melhorem em diversos aspectos (que parece variar em diferentes anos) para que possam concorrer ao prêmio. Desta forma, as escolas estaduais já estariam expostas ao tratamento (ao incentivo) em 2015, o que inviabilizaria o uso dos dados daquele ano para a comparação do “antes”. Ainda que elas não sejam contempladas em nenhum ano anterior, elas podem incentivo para realizar mudanças positivas desde 2012.**

**4.2 O grupo de controle é formado por escolas municipais e por escolas estaduais. As escolas estaduais, contudo, estão sujeitas ao incentivo. Elas apenas não se candidataram ao prêmio por algum motivo (seja por decisão do diretor da escola, porque não atingiram um nível mínimo de algum critério estipulado ou qualquer outro motivo). O que está sendo dito aqui é que, se o tratamento é o incentivo (a existência de um incentivo), ela é dada a todas as escolas estaduais desde a criação do Programa. O fato de a escola ganhar ou não o prêmio não é o tratamento. Isso inviabiliza a presença de escolas estaduais no grupo de controle, pois elas receberiam o incentivo.**

**4.3 Como mencionado na Pag 18, escolas que não foram premiados e estão no grupo de controle em um determinado ano poder ter sido premiado em anos anteriores (tendo sido portanto “tratado” anteriormente). Pag 18: “Como visto anteriormente, o programa existe desde o ano de 2012 e, desde sua criação, os critérios de seleção das escolas premiadas mudam ano a ano. Consequentemente, a relação de escolas premiadas se altera a cada premiação ocasionando grande volatilidade na entrada e saída de escolas tratadas e controles da amostra, dificultando assim a análise do tempo de exposição dos tratados ao programa e o impacto deste nos resultados educacionais.”**

**4.4 Recomendação para avaliação deste programa que fossem incluídos dados anteriores à lei que institui o programa, ou seja, anterior a 2012.**

**Resposta ao item 4 como um todo:** entendemos que o ponto levantado pelo avaliador de fato faz sentido e é bastante interessante e importante pensar no programa sob o aspecto do incentivo que é oferecido às escolas, mesmo às não premiadas. Por outro lado, o nosso objetivo durante a construção do artigo foi investigar o impacto da bonificação em si. Será que o recebimento do prêmio motiva os funcionários e professores a desempenharem melhor suas funções e, em razão disso, a qualidade e o rendimento dos alunos dessas escolas apresentam melhora? O custo dessas bonificações em termos de orçamento apresenta um retorno em qualidade da educação? A utilização da técnica de pareamento age no sentido de selecionar escolas, estaduais e municipais, igualmente elegíveis a receberem esse prêmio. Ao selecionar essas escolas, partimos do pressuposto que estas são comparáveis inclusive em termos dos incentivos gerados pela mera existência do PEV.

No tocante à análise de impacto considerando 2016 como ano da intervenção, a indisponibilidade de informações nos meios de publicidade oficiais, a falta de divulgação das pontuações e de transparência da SEE-PB, de uma maneira geral, trouxeram algumas limitações ao artigo. De fato, o programa se inicia em 2012 e desde então as informações a que tivemos acesso revelaram que os critérios de seleção das escolas premiadas mudam constantemente, ano a ano. Assim, entendemos não ser possível tratar de forma homogênea as premiações em anos subsequentes ao primeiro ano de vigência do programa, pois dadas as particularidades de cada seleção, é como se houvesse diferentes programas dentro de um. Por esta razão, o presente artigo analisa o impacto nas escolas premiadas no ano de 2016.

Sobre o item 4.3, de fato, com as mudanças nos critérios de premiação ano a ano, a relação de escolas premiadas se altera a cada premiação, fazendo com que várias escolas transitem entre os grupos das premiadas e não premiadas. Dentro de um grupo de controle é necessário diferenciar as escolas que por ventura já foram premiadas, sob o risco de tratar igualmente como grupo de controle escolas já tratadas anteriormente. Como forma de incorporar essas diferenças, modificamos o modelo empírico de diferenças em diferenças, de maneira a levar em conta o número de vezes que a escola foi premiada antes de 2016. Técnica semelhante foi empregada por Rocha e Soares (2010), os quais consideraram o tempo de exposição de municípios brasileiros ao Programa Saúde da Família (PSF). Ao invés de termos uma única variável *dummy* denotando o status de tratamento (1-tratado, 0-controle), incluímos cinco *dummies* para o número de vezes em que as escolas foram escolhidas para receber a bonificação, uma *dummy* com valor 1para quem já foi premiada uma vez, outra para quem já foi premiada duas vezes e assim por diante até a *dummy* para as escolas premiadas cinco vezes (ou seja, em todos os anos de vigência do programa, de 2012 a 2016). A interação dessas *dummies* com a variável tempo nos fornece o impacto do programa para as escolas de acordo com o número de vezes em que foram beneficiadas, comparativamente àquelas que não foram contempladas em nenhum ano. Os resultados encontrados foram interessantes e ofereceram um contraponto aos resultados encontrados por outros trabalhos que procuraram medir os efeitos desses programas de bonificação ao longo dos anos.

**5 Resultados:**

**5.1 Uma análise descritiva dos dados também é recomendável.**

**Resposta:** incluímos uma tabela com estatísticas descritivas das variáveis.

**5.2 Colocar asteriscos para os níveis de significância estatística facilita a leitura da tabela. (TABELA 1 – Teste de balanceamento entre tratados e controle no período pré-tratamento)**

**Resposta:** incluímos asteriscos juntamente aos resultados do p-valor para denotar os casos nos quais havia diferença significativa de média entre os grupos antes do pareamento.

**5.3 A equação do modelo estimado poderia ter sido apresentada.**

**Resposta:** incluímos uma equação mostrando o modelo empírico.

**5.4 A despeito dos comentários sobre a metodologia acima, é recomendável que outros grupos de comparação sejam testados para avaliar a robustez.**

**Resposta:** conforme explicamos acima, dividimos as escolas segundo a quantidade de participações, para ver como o programa impacta de forma diferenciada esses grupos.

**6 Conclusões**

**6.1 O IDEPB é um sistema de avaliação externa do estado da Paraíba, recomenda-se melhorar esse trecho. Pag. 37: “O fato do bônus estar condicionado às notas da avaliação interna do IDEPB e não às notas do Saeb, uma avaliação externa, garante maior confiabilidade aos resultados obtidos nesse estudo, visto que, evita a ocorrência de potenciais vieses, como de gamming.”**

**Resposta:** reescrevemos alguns pontos da conclusão.

**6.2 O fato de as escolas estaduais fazerem parte do controle e da existência de tal programa desde 2012 pode ter sido a razão para os resultados negativos. Ou seja, o grupo e período comparados já estariam sendo impactados positivamente pelo incentivo.**

**Resposta:** procuramos minimizar esses problemas através da estratégia que descrevemos anteriormente. Alguns resultados mudaram em relação à estratégia que adotamos na primeira versão do artigo.