

# ESTIMAÇÃO DA DESIGUALDADE DE CONTINGÊNCIA ENTRE OS ESTUDANTES DO ENSINO FUNDAMENTAL DA REDE PÚBLICA DE JOÃO PESSOA – PB

Stélio Coêlho Lombardi Filho (PPGE-UFPB)

Mércia Santos da Cruz (PPGE-UFPB)

Guilherme Diniz Irffi (CAEN-UFC)

## Resumo

O objetivo do presente artigo é verificar a ocorrência de desigualdade de contingência no sentido descrito por Rawls (1971), entre os alunos do 5º e 9º ano das escolas públicas de João Pessoa. Para tanto, utilizou-se os microdados da Prova Brasil de 2011 considerando como variáveis de resultado as notas de proficiência em português e matemática. Para aferir o efeito contingência, adotou-se como estratégia empírica o método de emparelhamento por escore de propensão combinado com a técnica de decomposição de Oaxaca-Blinder. Essa escolha metodológica permite realizar comparações entre estudantes com características observáveis semelhantes e, caso ainda se observe diferença na proficiência de português e/ou matemática, decompor esse diferencial entre fatores observáveis e não observáveis. A diferença oriunda de fatores não observáveis capta o efeito contingência, isto é, o conjunto de habilidades inatas que alguns indivíduos possuem e que interfere no processo de aprendizado. A partir dos resultados, pode-se dizer que existe desigualdade de contingência tanto entre os alunos do 5º quanto do 9º ano, em português e matemática, uma vez que o maior diferencial computado entre os discentes foi decorrente de características não observáveis, mesmo considerando apenas aqueles estudantes semelhantes em suas características particulares, bem como no *background* familiar, estrutura da escola e características dos professores e diretores.

**Palavras-Chave:** Prova Brasil 2011; Emparelhamento por escore de propensão; Decomposição de Oaxaca-Blinder; Efeito contingência.

## Abstract

The aim of this article is to verify the occurrence of inequality contingency in the sense described by Rawls (1971) among students in the 5th and 9th grades at the public schools of João Pessoa. For this purpose, the microdata from Prova Brasil 2011 was used considering the proficiency grades in portuguese and mathematics as result variables. To assess the contingency effect, it was adopted as empirical strategy the Propensity Score Matching, combined with the Blinder-Oaxaca decomposition. This methodological choice allows to make comparisons among students with similar observable characteristics and, in case there is still differences in proficiency of portuguese and/or mathematics, to decompose this difference between observable and unobservable factors. The difference arising from unobservable factors captures the contingency effect, ie, the set of innate abilities that some individuals possess and that interferes in their learning process. From the results, it can be said that there is inequality contingency among students in the 5th as well as the 9th year in both portuguese and mathematics, since the biggest difference computed among students was due to unobservable characteristics, even considering only similar students in their particular characteristics, as well as family background, school structure and characteristics of teachers and principals.

**Keywords:** Prova Brasil 2011; Propensity Score Matching; Blinder-Oaxaca decomposition; Contingency effect.

**JEL:** I20; I24.

## 1. Introdução

A questão dos determinantes do desempenho educacional tem levantado discussões nas diferentes esferas do governo, entidades de classe e na sociedade como um todo. Conforme Sen (1985), o desempenho educacional é fator determinante tanto para a formação profissional do indivíduo como para seu próprio bem-estar. O desempenho escolar também vem sendo correlacionado com desigualdade de renda e divergências no *status* de saúde dos indivíduos, tornando-se fator relevante nas decisões políticas das nações desenvolvidas e subdesenvolvidas (HESHMATI, 2004).

Sobretudo desde a década de 1970, diferentes pesquisas e testes padronizados<sup>1</sup> têm objetivado mensurar o desempenho educacional e cognitivo dos estudantes em diversos países, levando a comparações intra e entre nações, no que se reporta à dispersão das notas dos estudantes. Os resultados desses testes têm motivado as diferentes sociedades a identificar os determinantes do desempenho escolar, onde cada vez mais as características do ambiente familiar, condições da escola, incentivo ao estudo dado por familiares e/ou pelo professor, bem como o esforço dos alunos e outras características individuais, vêm sendo apontados como fatores que contribuem largamente para a melhoria do aprendizado.

Adicionalmente, uma literatura mais recente investiga o papel de questões de justiça e equidade no desempenho escolar dos alunos. A partir da tríade justiça, desigualdade e desempenho, o papel da desigualdade de oportunidade vem sendo adotado como um dos preditores da captação de capital humano [PATTANAİK e XU, (1990); KRANICH, (1996); THOMSON, (2011); FIGUEIREDO e SILVA, (2011)]. Contudo, não há um consenso sobre como seria possível promover uma disseminação da educação de modo mais justo.

Considerando o conceito de justiça de Rawls (1971), existirá justiça se todos os indivíduos tiverem acesso às mesmas condições e benefícios, o que tende a garantir a equidade. Entretanto, a teoria rawlsiana considera ainda que aos indivíduos diferentes sejam dados tratamentos diferenciados. Dessa forma, essa teoria considera que alguns membros da sociedade apresentam melhores dotações como, por exemplo, maior habilidade e/ou talento nato. Sendo assim, estes indivíduos são mais privilegiados *vis-à-vis* outros sujeitos da sociedade, de tal modo que ao segundo grupo deveria ser dada atenção especial em decorrência da desigualdade de contingência<sup>2</sup>.

Mais especificamente, com relação ao papel da educação na ótica de Rawls, Gondim (2010) e Gondim e Rodrigues (2011) desenvolvem discussões mais detalhadas. Segundo esses autores, para Rawls, a educação poderia ser definida como o treinamento e o desenvolvimento de habilidades e aptidões capazes de conceder maior autonomia aos cidadãos, tornando-os mais preparados para o debate público e favorecendo o bom funcionamento de uma sociedade. Além disso, uma vez que os indivíduos são diferentes em suas dotações, caberia à sociedade alocar recursos à educação de modo a corrigir “os desvios de contingência na direção da igualdade”<sup>3</sup>.

Em vista do que foi abordado, a motivação da presente pesquisa parte de duas frentes. Do ponto de vista teórico, testar-se-á a hipótese de que os indivíduos diferentes, mesmo que submetidos às mesmas condições iniciais e ao mesmo tratamento, ainda assim apresentarão resultados diferentes, tanto do ponto de vista da construção dos níveis de capital humano como na capacidade de edificação das habilidades. Nestes termos, pode-se dizer que esse artigo se propõe a investigar o desempenho escolar a partir das notas de proficiência de português e matemática para os alunos do 5º e 9º ano das escolas públicas de João Pessoa, na

---

<sup>1</sup> Com destaque para o *Programme for International Student Assessment* (PISA).

<sup>2</sup> Hanushek (1970) apresenta uma função de produção para a educação que inclui um vetor de características inatas como insumo para o desempenho escolar.

<sup>3</sup> Rawls (1971), p. 101.

Paraíba, aferidos a partir da Prova Brasil de 2011. Para isso, foi utilizada a técnica desenvolvida por Oaxaca (1973) e Blinder (1973) para decompor o desempenho acadêmico entre componentes observáveis (*background* familiar, estrutura da escola, características do aluno, do professor e diretor) e não observáveis (desigualdade de contingência, no sentido descrito por Rawls).

Além disso, visando uma maior confiabilidade nos resultados, utiliza-se do método de Pareamento por Escore de Propensão (*Propensity Score Matching*) com enfoque no nível de esforço (se o aluno faz o dever de casa e se ele lê livros) e na questão do incentivo ao estudo (por parte da família), para controlar por características observáveis como estrutura física e organizacional da escola, nível de instrução dos professores e diretores e pelo esforço e incentivo, o desempenho escolar dos alunos.

Espera-se que mesmo considerando condições iguais – sejam elas familiares ou de oportunidade, esforço e incentivo – o desempenho nos exames padronizados seja divergente, em função da diferença de habilidades, o que não garante uma distribuição justa de contingência, haja vista que os indivíduos mais habilidosos são comparativamente mais favorecidos. Com isso, muitas dimensões das desigualdades nos rendimentos escolares dos diferentes estudantes não podem ser captadas apenas por fatores observáveis, dado que os indivíduos são dissimilares em sua própria natureza, caracterizando assim o efeito contingência.

Em suma, a partir de uma perspectiva prática, a hipótese é de que mesmo controlando por fatores observáveis, o desempenho em português e matemática dos alunos tanto do 5º quanto do 9º ano são díspares em relação aos indivíduos semelhantes. Portanto, admite-se que a diferença em termos de desempenho seja decorrente da divergência inata de cada indivíduo.

A estratégia empírica adotada para alcançar o objetivo proposto consiste em estimar um modelo por meio da técnica de *Propensity Score Matching* para gerar um suporte comum, o que tornará a amostra mais homogênea em termos de características observáveis. A partir disso, será possível comparar o desempenho dos alunos. Em seguida, emprega-se a metodologia desenvolvida por Oaxaca e Blinder (1973) para decompor o desempenho em português e em matemática entre componentes observáveis e não observáveis. A expectativa de diferenças significativas para os componentes não observados fornece evidências de que existe efeito contingência entre os alunos do 5º e 9º ano das escolas públicas de João Pessoa.

Para consecução desses objetivos, optou-se por dividir o artigo em quatro seções, além desta introdução. A próxima contempla a descrição da fonte e o tratamento dos dados. Em seguida, é descrita a estratégia empírica adotada para captar o efeito da desigualdade de contingência. A análise e discussão dos resultados são o tema da quarta seção. E, por fim, são tecidas as considerações finais.

## **2. Fonte e tratamento dos dados**

Para a realização desta pesquisa, utilizou-se os microdados da Prova Brasil de 2011, disponíveis no portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). A Prova Brasil é uma das três modalidades que compõe o Sistema de Avaliação da Educação Básica (SAEB), implementado pelo governo federal em 1990. Pela Figura 1, abaixo, é possível compreender melhor como está estruturado o referido sistema.

### **Figura 1 – Estrutura do Sistema de Avaliação da Educação Básica**



**Fonte:** INEP.

A ANEB consiste na parte amostral do SAEB, abrangendo alunos das redes públicas e privadas do país, tanto em áreas urbanas quanto em áreas rurais, matriculados no 5º e 9º ano do Ensino Fundamental e no 3º ano do Ensino Médio. O público-alvo dessa modalidade é composto por escolas que tenham de 10 a 19 estudantes matriculados no 5º ou 9º ano do Ensino Fundamental regular público, escolas que tenham 10 ou mais estudantes matriculados no 5º ou 9º ano do Ensino Fundamental regular privado, e escolas que tenham 10 ou mais estudantes matriculados no 3º ano do Ensino Médio regular público ou privado (BRASIL, 2012).

A ANRESC (Prova Brasil) é uma avaliação censitária, envolvendo alunos do 5º e 9º ano do Ensino Fundamental das escolas públicas das redes municipais, estaduais e federal. São consideradas todas as escolas com pelo menos 20 alunos matriculados, nas áreas rurais e urbanas do Brasil (BRASIL, 2012).

Por fim tem-se a ANA<sup>4</sup>, que trata-se também de uma avaliação censitária, englobando os alunos do 3º ano do Ensino Fundamental das escolas públicas. Seu objetivo principal é avaliar os níveis de alfabetização e letramento em Língua Portuguesa, alfabetização Matemática e condições de oferta do Ciclo de Alfabetização das redes públicas (BRASIL, 2012).

Isto posto, como o objetivo desse artigo é captar o possível diferencial de desempenho existente entre os estudantes do Ensino Fundamental da rede pública de João Pessoa, controlando as comparações por meio de atributos observáveis e usando o método de Oaxaca-Blinder para decompor essa diferença em fatores explicados e não explicados, tem-se que os microdados da Prova Brasil são adequados à essa finalidade. Isso ocorre, pois, seus questionários apresentam informações referentes às características dos discentes e seus domicílios, além de fornecer informações importantes acerca das escolas, do diretor e dos professores. Dessa forma, esta se configura em uma base de dados que permite comparar estudantes que possuem características similares, voltada exclusivamente à rede pública do Ensino Fundamental.

No total, os microdados da Prova Brasil de 2011 contém informações sobre 5.201.730 estudantes. Porém, conforme estipulado nos objetivos, e a fim de tornar a base mais homogênea, considerou-se apenas os discentes do município de João Pessoa. Após esse recorte e a exclusão das observações faltantes (*missing values*), a base final ficou com 1.193 observações para os alunos do 5º ano, e 2.486 observações para os alunos do 9º ano.

Como o desempenho escolar pode ser afetado por características dos alunos e do domicílio, bem como pelo ambiente familiar e situação socioeconômica, se faz necessário controlar por essas características para aferir o efeito contingência dos alunos. Sendo assim, optou-se por reportar no Quadro 1 a descrição das variáveis utilizadas, ressaltando que a escolha por controlar por determinadas características é feita a partir da literatura que identifica como as mais relevantes para o desempenho escolar.

As variáveis referentes à leitura de livros e resposta da tarefa de casa são utilizadas como *proxies* para medir esforço e dedicação respectivamente. Com relação às variáveis dos

<sup>4</sup> A Avaliação Nacional da Alfabetização somente foi incorporada ao SAEB em 2013. Tal avaliação ocorre todos os anos, enquanto que a ANEB e a ANRESC são realizadas bianualmente.

demais grupos, alusivas às condições das escolas e escolaridade e experiência dos professores e diretores, estas têm como objetivo medir os efeitos do ambiente escolar como um todo sobre o desempenho dos estudantes.

**Quadro 1 – Descrição das Variáveis**

<b>Variáveis</b>	<b>Descrição das Variáveis</b>
Proficiência_SAEB	Proficiência do aluno em língua portuguesa ou em matemática calculada na escala única do SAEB
<b>Características dos Alunos e do Domicílio</b>	
Sexo_aluno	1 se o sexo do aluno for masculino e 0 caso contrário.
Raça_aluno_branco	1 se a raça do aluno for branca e 0 caso contrário.
Trabalho_aluno	1 se o aluno trabalha fora de casa e 0 caso contrário.
Incentivo_aluno	1 se os pais do aluno o incentivam a estudar e 0 caso contrário.
Ler_livros_sempre	1 se o aluno ler livros sempre ou quase sempre e 0 caso contrário.
Faz_tarefa_sempre <sup>5</sup>	1 se o aluno faz a tarefa de casa sempre ou quase sempre e 0 caso contrário.
Início_dos_estudos	1 se o aluno iniciou os estudos antes do 1º ano e 0 caso contrário.
Reprovação_aluno	1 se o aluno nunca foi reprovado e 0 caso contrário.
Computador_com_internet	1 se na casa do aluno tem computador com internet e 0 caso contrário.
Escol_mãe_ens_superior	1 se a mãe do aluno completou a faculdade e 0 caso contrário.
Prof_corrige_tarefa_sempre <sup>6</sup>	1 se o professor corrige a tarefa de casa sempre ou quase sempre e 0 caso contrário.
<b>Características das Escolas<sup>7</sup></b>	
Salas_escola_bom	1 se o estado de conservação das salas de aula da escola é considerado bom e 0 caso contrário.
Banheiros_escola_bom	1 se o estado de conservação dos banheiros da escola é considerado bom e 0 caso contrário.
Biblioteca_escola_bom	1 se o estado de conservação da biblioteca da escola é considerado bom e 0 caso contrário
Laboratório_escola_bom	1 se o estado de conservação do laboratório da escola é considerado bom e 0 caso contrário
<b>Características dos Professores</b>	

<sup>5</sup> Ao avaliar a proficiência em português (matemática), a variável considerada foi se o estudante faz a tarefa de casa de português (matemática) sempre ou quase sempre.

<sup>6</sup> Ao avaliar a proficiência em português (matemática), a variável considerada foi se o professor corrige a tarefa de casa de português (matemática) sempre ou quase sempre.

<sup>7</sup> Todas as escolas públicas de Ensino Fundamental do município de João Pessoa, contempladas pela Prova Brasil 2011, possuem dependência administrativa municipal ou estadual.

Escol_prof_ens_superior	1 se o professor concluiu o ensino superior e 0 caso contrário.
Experiência_professor	1 se o professor leciona há 10 anos ou mais e 0 caso contrário.
<b>Características do Diretor</b>	
Escol_diretor_ens_superior	1 se o diretor possui ensino superior completo e 0 caso contrário
Experiência_diretor	1 se o diretor trabalha em educação há 10 anos ou mais e 0 caso contrário

**Fonte:** Microdados Prova Brasil 2011.

Apresentada a fonte dos dados e a descrição das variáveis, a Tabela 1 contempla as estatísticas descritivas da amostra<sup>8</sup>. Como pode ser observado, cerca de 49,71% dos alunos do 5º ano são do sexo masculino, enquanto que no 9º ano tal proporção cai para 44,53%. Além disso, 26,40% e 24,54% dos discentes do 5º e 9º ano, respectivamente, declararam ser da cor branca. Em ambos os anos, pouco mais de 11% dos alunos declararam possuir algum tipo de trabalho fora de casa, e mais de 97% afirmaram que seus pais os incentivam a estudar.

Algumas diferenças consideráveis podem ser apontadas nas variáveis de leitura de livros e de resposta da tarefa de casa. Enquanto que mais de 45% dos alunos do 5º ano alegaram ler livros sempre ou quase sempre, apenas 28,16% dos estudantes do 9º afirmaram possuir tal hábito. Já no que diz respeito às tarefas de casa, quase 80% dos discentes do 5º ano afirmaram respondê-las com frequência, contra menos de 56% dos alunos do 9º ano.

Com relação às características das escolas, no geral, as principais diferenças residem nas condições da biblioteca e do laboratório. A proporção de escolas com biblioteca em boas condições é maior para os alunos do 9º ano. Já com relação às condições do laboratório, a proporção é maior para os alunos do 5º ano. Vale lembrar que algumas escolas ofertam tanto o 5º quanto o 9º ano, e todas estão localizadas no mesmo município, isto é, João Pessoa.

No que se refere aos professores, constatou-se que mais de 99% dos que ensinam para os alunos do 9º ano concluíram o ensino superior, contra 86% dos professores dos alunos do 5º ano. No quesito experiência, todavia, tem-se que, na média, os professores do 5º ano são mais experientes. Por fim, tem-se que mais de 94% dos diretores das duas séries concluíram o ensino superior e trabalham em educação há 10 anos ou mais.

<sup>8</sup> As estatísticas descritivas da base de dados completa, isto é, sem a exclusão das observações faltantes (*missing values*) podem ser consultadas no apêndice desse artigo.

**Tabela 1 - Estatísticas descritivas dos alunos do 5º e 9º ano**

	Variáveis	5º ano				9º ano			
		Média	Desvio-padrão	Mínimo	Máximo	Média	Desvio-padrão	Mínimo	Máximo
<b>Características do Aluno e do Domicílio</b>	Sexo_aluno	0,4971	0,5002	0	1	0,4453	0,4971	0	1
	Raça_aluno_branco	0,2640	0,4410	0	1	0,2454	0,4304	0	1
	Trabalho_aluno	0,1140	0,3179	0	1	0,1110	0,3142	0	1
	Incentivo_aluno	0,9723	0,1641	0	1	0,9855	0,1195	0	1
	Ler_livros_sempre	0,4568	0,4983	0	1	0,2816	0,4499	0	1
	Faz_tarefa_português_sempre	0,7955	0,4035	0	1	0,5531	0,4973	0	1
	Faz_tarefa_matemática_sempre	0,7812	0,4136	0	1	0,5133	0,4999	0	1
	Início_dos_estudos	0,7896	0,4078	0	1	0,8512	0,3560	0	1
	Reprovação_aluno	0,6454	0,4786	0	1	0,5841	0,4930	0	1
	Computador_com_internet	0,3889	0,4877	0	1	0,5133	0,4999	0	1
	Escol_mãe_ens_superior	0,1769	0,3817	0	1	0,0961	0,2948	0	1
	Prof_corrige_tarefa_port_sempre	0,7795	0,4147	0	1	0,7446	0,4362	0	1
	Prof_corrige_tarefa_mat_sempre	0,7921	0,4060	0	1	0,7418	0,4378	0	1
	Proficiência_português	191,8512	42,1511	90,9695	320,6530	245,6801	44,2513	108,8913	360,7970
	Proficiência_matemática	207,8248	41,7865	99,7824	337,1514	246,5826	45,0614	120,9378	396,0480
<b>Características da Escola</b>	Salas_escola_bom	0,7544	0,4306	0	1	0,7751	0,4176	0	1
	Banheiros_escola_bom	0,5197	0,4998	0	1	0,4360	0,4960	0	1
	Biblioteca_escola_bom	0,7225	0,4479	0	1	0,8230	0,3817	0	1
	Laboratório_escola_bom	0,5616	0,4964	0	1	0,4304	0,4952	0	1
<b>Características do Professor</b>	Escol_professor_superior	0,8625	0,3445	0	1	0,9952	0,0693	0	1
	Experiência_professor	0,7049	0,4563	0	1	0,6223	0,4849	0	1
<b>Características do Diretor</b>	Escolaridade_diretor_superior	0,9447	0,2287	0	1	0,9795	0,1418	0	1
	Experiência_diretor	0,9338	0,2488	0	1	0,9887	0,1055	0	1

Fonte: Microdados Prova Brasil 2011.

### 3. Estratégia de estimação

Para a consecução dos objetivos delineados, primeiramente foi empregado o método de Pareamento por Escore de Propensão e, posteriormente, utilizou-se a decomposição de Oaxaca-Blinder para aferir a desigualdade de contingência entre os alunos do 5º e 9º ano. Neste sentido, inicialmente, a amostra foi dividida entre os estudantes do 5º e do 9º ano do Ensino Fundamental da rede pública de João Pessoa e, a partir das notas de proficiência em português e matemática, estes estudantes foram divididos em dois grupos: aqueles com desempenho acima da média e aqueles com desempenho abaixo.

Uma vez divididos os grupos para cada ano, utilizou-se variáveis referentes às características dos alunos, dos domicílios, das escolas, dos professores e dos diretores para garantir que cada estudante tenha seu desempenho comparado apenas com outros estudantes que sejam similares nos atributos selecionados. Caso seja verificado que há diferença estatisticamente significativa no desempenho médio dos dois grupos, é possível utilizar o procedimento de Oaxaca-Blinder para decompor essa diferença em fatores explicados e não explicados. A esses fatores não explicados (não observáveis) será dado o nome de efeito contingência, isto é, o conjunto de habilidades inatas que alguns indivíduos possuem e que interfere no processo de aprendizado.

Com base nessa descrição, pode-se dizer que a estratégia de estimação do efeito contingência se dá por meio da combinação da estimação do método de Pareamento por Escore de Propensão seguido pela decomposição de Oaxaca-Blinder. Diante disso, optou-se por apresentar em mais detalhes cada um desses métodos.

#### 3.1 Pareamento por Escore de Propensão

Na literatura de avaliação de impacto de programas sociais, o método de pareamento tem como objetivo reduzir o viés de autoseleção presente na estimação do impacto de um dado tratamento. Tal método busca construir um grupo de controle consistente e o mais similar possível ao grupo de tratamento, partindo da concepção de que quanto mais parecido forem esses grupos em suas características observáveis, menor será o viés.

Como ressaltado em Menezes-Filho (2012, p. 85): “De acordo com as hipóteses desse método, cada membro do grupo de tratamento teria um par no grupo de controle que representa o resultado que ele teria obtido caso não fosse tratado”. Assim, o pareamento assume que ao controlar os grupos com base em um conjunto de atributos observáveis, o grupo de controle passaria a representar um contrafactual adequado ao grupo de tratamento.

Apesar do referido método ser usado sobretudo para avaliar os efeitos de políticas públicas e programas sociais, pode-se empregá-lo basicamente em qualquer situação onde seja possível separar as unidades de observação em grupos com base em uma dada variável de resultado. Feito essa divisão, o método consiste simplesmente em comparar (parear) essas unidades usando como critério o conjunto de características observáveis selecionadas pelo pesquisador. Por fim, estima-se o efeito médio do tratamento nos tratados (*Average Treatment of the Treated* – ATT) através da diferença de média da variável de resultado entre os grupos.

É importante ressaltar, todavia, que quanto maior for número de características selecionadas, isto é, quanto maior a dimensão do vetor  $X$ , mais difícil acaba sendo parear as observações (MENEZES-FILHO, 2012)<sup>9</sup>. Visando solucionar essa dificuldade, Rosenbaum e Rubin (1983) propuseram o método de Pareamento por Escore de Propensão (*Propensity*

---

<sup>9</sup> Tal problema ocorre, sobretudo, se as variáveis forem contínuas.



*Score Matching* – PSM), que consiste em utilizar uma função de  $X$  que resuma todas as características presentes nesse vetor.

Isto posto, o escore de propensão pode ser entendido como a probabilidade de receber o tratamento condicionada às variáveis de controle incluídas em  $X$ . Mais formalmente, segundo Becker e Ichino (2002), o escore de propensão é dado por:

$$p(X) \equiv \Pr(D = 1|X) \quad (1)$$

Em que  $D = \{0,1\}$  indica a exposição ao tratamento e  $X$  é o vetor de características observáveis. Na prática, a estimação do escore de propensão é geralmente feita por meio de procedimentos paramétricos, tais como os modelos *logit* ou *probit* (MENEZES-FILHO, 2012). No caso do modelo *logit*, que será o procedimento adotado nesse trabalho, o escore de propensão será estimado como:

$$\hat{p}(x) = \frac{\exp(x\hat{\beta})}{1 + \exp(x\hat{\beta})} \quad (2)$$

Onde  $\hat{\beta}$  é o estimador do parâmetro  $\beta$  e  $x$  é o vetor contendo as variáveis de controle selecionadas.

Conforme Khandker, Koolwal e Samad (2010), a validade do PSM necessita que duas hipóteses sejam atendidas: independência condicional e presença de suporte comum. A hipótese de independência condicional, também chamada de não-confundimento, postula que apenas os fatores observáveis afetam a adesão ao programa. Com isso:

$$Y_1, Y_0 \perp D|X \quad (3)$$

Como ressaltado pelos autores, essa é uma hipótese forte que não pode ser diretamente testada, uma vez que depende das propriedades específicas e do desenho de cada programa.

A segunda hipótese, presença de suporte comum, requer que haja sobreposição nos escores de propensão entre as unidades dos grupos de tratado e controle (KHANDKER, KOOLWAL e SAMAD, 2010). De modo mais formal, é preciso que:

$$0 < \Pr[D = 1|X] < 1 \quad (4)$$

Essa hipótese garante que as unidades no grupo de tratamento possuem um par no grupo de controle que permita sua comparação (MENEZES-FILHO, 2012).

Para empregar o referido método nessa pesquisa, será criada uma variável dicotômica que assume valor 1, caso o resultado do aluno na Prova Brasil tenha sido acima da média dos estudantes de João Pessoa, e 0 caso contrário<sup>10</sup>. Tal variável dicotômica será regredida através de um modelo *logit*, tendo como variáveis explicativas as que foram reportadas no Quadro 1. Após a estimação do escore de propensão, obtém-se uma região de suporte comum que permitirá comparar estudantes semelhantes em seus atributos observáveis. Uma vez concluída essa primeira etapa, isto é, estimado o escore de propensão e pareado a amostra, pode-se partir para a decomposição de Oaxaca-Blinder, de modo a mesurar as parcelas do diferencial total de desempenho dos alunos (em português e matemática) que pode ser atribuído aos fatores observáveis e não observáveis.

<sup>10</sup> Mais especificamente, será criada uma variável dicotômica desse tipo tanto para português quanto para matemática, para o 5º e 9º ano.

### 3.2 Decomposição de Oaxaca-Blinder

A técnica de decomposição contrafactual desenvolvida por Oaxaca (1973) e Blinder (1973) tem sido vastamente empregada para estudar diferenças de média entre grupos, em relação a uma dada variável de resultado. Essa decomposição se deve aos fatores explicados (observáveis) e aos fatores não explicados (não observáveis), sendo muitas vezes empregada como forma de mensurar discriminação (JANN, 2008).

Nesse artigo, a variável de resultado utilizada será a proficiência em português e matemática dos alunos do 5º e 9º ano. Já as variáveis explicativas serão selecionadas com base nos resultados do escore de propensão. Tendo em vista como a decomposição de Oaxaca-Blinder funciona, esta será utilizada para captar o diferencial de desempenho entre os estudantes, examinando a ocorrência de desigualdade de contingência entre alunos com características observáveis semelhantes.

Para entender como funciona tal procedimento, suponha dois grupos ( $A$  e  $B$ ) e uma variável de resultado ( $Y$ ). A diferença de média de  $Y$  entres esses dois grupos pode ser calculada da seguinte forma:

$$R = E(Y_A) - E(Y_B) \quad (5)$$

Em que  $E(Y_A)$  e  $E(Y_B)$  denotam, respectivamente, o valor esperado da variável de resultado para os grupos  $A$  e  $B$ . Considere agora um modelo de regressão linear do tipo:

$$Y_i = X_i' \beta_i + i \quad (6)$$

Onde  $X$  é um vetor de características determinantes da variável de resultado,  $\beta$  um vetor de parâmetros desconhecidos a ser estimado,  $i$  o termo de erro estocástico,  $i \in (A, B)$  e, por hipótese, que  $E(i) = 0$ . Substituindo (6) em (5) e simplificando os termos:

$$R = E(Y_A) - E(Y_B) = E(X_A)' \beta_A - E(X_B)' \beta_B \quad (7)$$

Assuma agora que existe um vetor de coeficientes não discriminatórios ( $\beta^*$ ), responsável por determinar a contribuição das diferenças entre os regressores. Caso ocorra discriminação contra o grupo  $B$ , mas não haja discriminação em favor do grupo  $A$ , então  $\beta^* = \beta_A$  e a decomposição de Oaxaca-Blinder consiste em estimar:

$$\hat{R} = (\bar{X}_A - \bar{X}_B)' \hat{\beta}_A + \bar{X}_B' (\hat{\beta}_A - \hat{\beta}_B) \quad (8)$$

Em que  $\bar{X}_A$  e  $\bar{X}_B$  representam os valores médios das características dos grupos  $A$  e  $B$ , respectivamente, sendo usados como estimativas para  $E(X_A)$  e  $E(X_B)$ , e  $\hat{\beta}_A$  e  $\hat{\beta}_B$  são os parâmetros estimados por Mínimos Quadrados Ordinários de  $\beta_A$  e  $\beta_B$ <sup>11</sup>.

A primeira parte da equação (8), ou seja,  $(\bar{X}_A - \bar{X}_B)' \hat{\beta}_A$ , corresponde à parcela do diferencial da variável de resultado que é explicada pelas diferenças observáveis nas características dos grupos. Nesse componente, considera-se que ambos os grupos são valorados da mesma forma e, caso seu valor seja positivo, então os indivíduos do grupo  $A$  possuem maiores valores de  $Y$  devido aos seus atributos mensuráveis.

Já a segunda parte, isto é,  $\bar{X}_B' (\hat{\beta}_A - \hat{\beta}_B)$ , capta o potenciais efeitos das características não observáveis, geralmente atribuído à discriminação.<sup>12</sup> Nessa parcela, assume-se que os

<sup>11</sup> Da mesma forma, caso exista discriminação em favor do grupo  $A$ , mas sem discriminação contra o grupo  $B$ , tem-se que  $\beta^* = \beta_B$  e a estimação será dada por:  $\hat{R} = (\bar{X}_A - \bar{X}_B)' \hat{\beta}_B + \bar{X}_A' (\hat{\beta}_A - \hat{\beta}_B)$ .

indivíduos do grupo *B* apresentam as mesmas características observáveis médias do grupo *A*. Logo, caso esse componente seja positivo, então os indivíduos do grupo *A* recebem maiores valores de *Y* devido aos seus atributos não observáveis, ou devido a uma possível discriminação contra os indivíduos do grupo *B*.

Nesse artigo, a variável de resultado utilizada será a proficiência em português e matemática dos alunos do 5º e 9º ano. Já as variáveis explicativas serão selecionadas com base nos resultados do escore de propensão. Tendo em vista como a decomposição de Oaxaca-Blinder funciona, esta será utilizada para captar o diferencial de desempenho entre os estudantes, examinando a ocorrência de desigualdade de contingência entre alunos com características observáveis semelhantes.

#### 4. Análise e discussão dos resultados

Essa seção contempla a análise e discussão dos resultados, que tem como objetivo aferir se existe desigualdade de contingência entre os alunos do 5º e 9º ano, considerando as notas de proficiência em português e matemática. Para tanto, inicialmente, será calculado o escore de propensão para cada série e disciplina, o que permitirá identificar os fatores determinantes do desempenho dos alunos. Além disso, através desse escore, também será possível parear a amostra, tornando-a mais homogênea. Em seguida, serão estimados modelos de decomposição de Oaxaca-Blinder com e sem pareamento para testar a hipótese de desigualdade de contingência.

##### 4.1 Estimação do escore de propensão

Na Tabela 3 encontram-se os coeficientes estimados para o cálculo do escore de propensão dos alunos do 5º ano. O referido método foi empregado com o intuito de estimar a probabilidade dos estudantes de participar do grupo cujo desempenho na Prova Brasil foi acima da média. Como ressaltado na seção da metodologia, tal procedimento foi realizado através de um modelo de regressão *logit*, cujas variáveis explicativas são aquelas definidas no Quadro 1.

**Tabela 3 – Estimação do escore de propensão (5º Ano)**

Variáveis Explicativas	Português			Matemática		
	Coefficiente	Erro-padrão	dy/dx	Coefficiente	Erro-padrão	dy/dx
Sexo_aluno	-0,5805***	0,1253	-0,1438	0,2539**	0,1242	0,0631
Raça_aluno_branco	0,0606	0,1403	0,0151	-0,0464	0,1382	-0,0115
Computador_com_internet	0,1892	0,1305	0,0471	0,3598***	0,1286	0,08890
Escol_mãe_ens_superior	-0,0544	0,1666	-0,0136	0,0609	0,1647	0,0151
Incentivo_aluno	0,8224**	0,3978	0,1989	2,1004***	0,5440	0,4196
Ler_livros_sempre	-0,1713	0,1252	-0,0427	-0,3250***	0,1229	-0,0808
Trabalho_aluno	-0,5548***	0,1990	-0,1374	-0,2194	0,1924	-0,0548
Início_dos_estudos	0,3578**	0,1518	0,0892	0,3826**	0,1487	0,0953
Reprovação_aluno	0,7593***	0,1309	0,1876	0,6777***	0,1290	0,1677

<sup>12</sup> Nesse trabalho, tais efeitos serão abordados como efeito da desigualdade de contingência no sentido de Rawls.

Faz_tarefa_sempre	0,2424	0,1601	0,0605	0,0971	0,1485	0,0242
Prof_corrige_tarefa_sempre	-0,1068	0,1533	-0,0266	–	–	–
Salas_escola_bom	0,3568*	0,1871	0,0889	0,2077	0,1829	0,0518
Banheiros_escola_bom	-0,4783***	0,1685	-0,1186	-0,1737	0,1632	-0,0432
Biblioteca_escola_bom	0,0892	0,1655	0,0223	0,3596**	0,1631	0,0896
Laboratório_escola_bom	0,0289	0,1557	0,0072	-0,0445	0,1536	-0,0111
Escol_diretor_ens_superior	0,0031	0,3036	0,0008	-0,3778	0,3028	-0,0920
Experiência_diretor	0,8046*	0,2899	0,1956	0,8014***	0,2831	0,1957
Escol_prof_ens_superior	-0,0390	0,1889	-0,0097	-0,0583	0,1852	-0,0145
Experiência_prof	-0,2746*	0,1450	-0,068	-0,0949	0,1433	-0,02357
Intercepto	-1,8578***	0,5618	–	-3,3811***	0,6699	–

**Fonte:** Resultados da pesquisa

**Nota:** \*Estatisticamente significativa a 10%; \*\*Estatisticamente significativa a 5%; \*\*\*Estatisticamente significativa a 1%. A variável Prof\_corrige\_tarefa\_sempre não foi utilizada no cálculo do escore de propensão da prova de matemática para que a propriedade de suporte comum pudesse ser satisfeita.

Os coeficientes estimados na Tabela 3 indicam que, de modo geral, o fato do aluno ser homem reduz a probabilidade, em mais de 14 pontos percentuais (p.p.), dele obter um desempenho acima da média na prova de português. Já em matemática, o resultado é mais favorável aos estudantes do sexo masculino, indicando que estes possuem uma probabilidade maior de obter um desempenho acima da média. Em ambas as disciplinas, a variável que apresentou efeito marginal mais elevado foi aquela referente ao incentivo dos pais ao estudo: cerca de 20 p.p. em português e 42 p.p. em matemática. Outros resultados expressivos foram captados pela variável que indica se o aluno trabalha ou não fora de casa, pela variável que diz respeito à ocorrência de reprovação por parte do aluno, pela variável de experiência do diretor e pela variável de início dos estudos<sup>13</sup>.

Os resultados também mostraram que as características da escola não parecem ter relação direta significativa com o desempenho dos estudantes. Todavia, tais efeitos já eram esperados, pois os mesmos já foram observados em várias referências da literatura<sup>14</sup>. Apenas a existência de uma boa biblioteca, em relação à proficiência em matemática, e a existência de boas salas de aula, em relação à proficiência em português, apresentaram impactos positivos.

A escolaridade do professor não apresentou significância estatística nem para o 5º e nem para o 9º ano, enquanto que a experiência do professor e a qualidade dos banheiros da escola (para os alunos do 5º ano), apesar de significantes, indicaram efeito marginal negativo. Resultados similares foram encontrados em Souza (2005) e Araújo e Almeida (2013). Uma possível explicação para os efeitos observados com essas variáveis referentes aos professores foi dada por Hanushek (2007), que afirma que não necessariamente serão os professores mais experientes ou mais capacitados que melhorarão o desempenho dos alunos, mas sim aqueles com melhor didática e maior esforço ao preparar suas aulas.

Adiante, na Tabela 4, encontram-se os coeficientes estimados para os alunos do 9º ano. De forma similar para com os estudantes do 5º ano, o modelo indica que os estudantes do sexo feminino levam vantagem em português, enquanto que os estudantes do sexo masculino apresentam vantagens em matemática. Diferentemente dos resultados anteriores, a presença no domicílio de um computador com acesso à internet, e também a experiência dos professores, encontram-se positivamente relacionadas com o desempenho dos discentes, em ambas as matérias. Também vale ressaltar que a variável de incentivo dos pais ao estudo, que

<sup>13</sup> A importância da creche e da pré-escola na proficiência escolar foi analisada por Curi e Menezes-Filho (2006).

<sup>14</sup> Ver Hanushek (2007) para uma melhor compreensão do impacto da escola e dos professores sobre o desempenho dos estudantes.

mostrou-se importante para um bom desempenho dos estudantes do 5º ano, não apresentou significância estatística para os modelos referentes aos alunos do 9º ano.

**Tabela 4 – Estimação do escore de propensão (9º Ano)**

Variáveis Explicativas	Português			Matemática		
	Coefficiente	Erro-padrão	dy/dx	Coefficiente	Erro-padrão	dy/dx
Sexo_aluno	-0,3302***	0,0865	-0,0807	0,5830***	0,0832	0,1443
Raça_aluno_branco	0,0686	0,0982	0,0167	-0,0449	0,0935	-0,0112
Computador_com_internet	0,1821**	0,0865	0,0445	0,1551*	0,0821	0,0387
Escol_mãe_ens_superior	0,1409	0,1478	0,0341	0,1471	0,1400	0,0366
Incentivo_aluno	0,4107	0,3587	0,1020	-0,0229	0,3338	-0,0057
Ler_livros_sempre	0,2198**	0,0967	0,0532	–	–	–
Trabalho_aluno	-0,5210***	0,1358	-0,1292	-0,3106**	0,1290	-0,0775
Início_dos_estudos	0,3494***	0,1182	0,0864	0,2740**	0,1140	0,0684
Reprovação_aluno	0,6513***	0,0865	0,1590	0,7490***	0,0833	0,1851
Faz_tarefa_sempre	0,0633	0,0874	0,0155	0,4892***	0,0838	0,1215
Prof_corrige_tarefa_sempre	0,1811*	0,0982	0,0445	0,0118	0,0964	0,0030
Salas_escola_bom	0,1009	0,1172	0,0247	0,1024	0,1005	0,0256
Banheiros_escola_bom	0,0868	0,1058	0,0212	–	–	–
Biblioteca_escola_bom	0,0157	0,1181	0,0038	0,1449	0,1075	0,0362
Laboratório_escola_bom	0,1414	0,0936	0,0345	-0,1340	0,0866	-0,0334
Escol_diretor_ens_superior	-0,1605	0,2972	-0,0387	-0,2514	0,2930	-0,0621
Experiência_diretor	-0,0956	0,4102	-0,0232	0,0889	0,3699	0,0222
Escol_prof_ens_superior	0,2301	0,5961	0,0569	-0,0354	0,5480	-0,009
Experiência_prof	0,2090*	0,0897	0,0512	0,2742***	0,0846	0,0684
Intercepto	-1,2263	0,8912	–	-1,2393	0,8222	–

**Fonte:** Resultados da pesquisa

**Nota:** \*Estatisticamente significativa a 10%; \*\*Estatisticamente significativa a 5%; \*\*\*Estatisticamente significativa a 1%. As variáveis Ler\_livros\_sempre e Banheiros\_escola\_bom não foram utilizadas no cálculo do escore de propensão da prova de matemática para que a propriedade de suporte comum pudesse ser satisfeita.

Para a disciplina de português, as variáveis que apresentaram maiores efeitos marginais foram as relacionadas à questão da reprovação e ao fato do estudante trabalhar ou não. Com relação à primeira variável, os resultados apontaram que caso o aluno nunca tenha reprovado, sua probabilidade de obter um desempenho acima da média aumenta em quase 16 p.p. Já em relação à segunda variável, o efeito é de uma redução de pouco menos de 13 p.p. para aqueles que trabalham. Em matemática, a variável de reprovação novamente apresentou o maior efeito marginal (18,51 p.p.). Em seguida, vieram as variáveis de sexo do aluno e de resposta da tarefa de casa.

Os resultados observados indicam que as seguintes variáveis são as principais determinantes para que os estudantes apresentem desempenho acima da média: sexo do aluno, presença de computador com internet no domicílio, incentivo dos pais ao estudo, hábito de leitura, o fato do aluno trabalhar ou não, o momento em que o aluno iniciou seus estudos, se o aluno reprovou alguma vez, se o aluno costuma fazer a tarefa de casa, a experiência do diretor da escola e a experiência do professor. Tais resultados estão em consonância com aqueles encontrados em Menezes-Filho (2007), Araújo e Almeida (2013), Souza (2005) e Barros et al. (2001), bem como em outros trabalhos da literatura. No cálculo da decomposição de Oaxaca-

Blinder, serão essas as variáveis explicativas consideradas, juntamente com as variáveis de raça do aluno, condições das salas de aula e da biblioteca da escola, e escolaridade do professor. Apesar dessas últimas não terem se mostrado relevantes, as mesmas são sempre incluídas nos trabalhos empíricos que buscam analisar os determinantes do desempenho escolar

#### 4.2 Estimação da decomposição de Oaxaca-Blinder

Como já mencionado, o interesse dessa pesquisa é verificar a ocorrência de desigualdade de contingência entre os estudantes do 5º e 9º ano das escolas públicas de João Pessoa. Para tal propósito, empregou-se o método de Oaxaca-Blinder para decompor o diferencial de desempenho escolar em fatores observáveis e não observáveis, construindo grupos amostrais a partir das seguintes informações: se o aluno iniciou os estudos antes do 1º ano, se os pais do aluno o incentivam a estudar, se o aluno faz a tarefa de casa sempre ou quase sempre, se o aluno nunca foi reprovado e se o aluno apresentou desempenho acima da média. A escolha por esses grupos se justifica pela importância que tais variáveis têm apresentado em diversos estudos que buscam identificar os determinantes do desempenho escolar [MENEZES-FILHO, (2007); CURI e MENEZES-FILHO (2006); ARAÚJO e ALMEIDA, (2013), BARROS et al., (2001); SOUZA, (2005)].

Na Tabela 5 é possível observar os resultados da decomposição de Oaxaca-Blinder sem pareamento. Primeiramente, nota-se que o grupo amostral construído pela variável de início dos estudos apresentou significância estatística em todos os resultados. De acordo com esse modelo, os estudantes do 5º ano que iniciaram seus estudos na creche ou pré-escola apresentam, em média, desempenho superior de aproximadamente 10,50 pontos. É importante observar que a maior parte desse diferencial é atribuído aos fatores não explicados, isto é, não observáveis diretamente<sup>15</sup>. Já para os estudantes do 9º ano, o diferencial é um pouco maior: 11,98 pontos em português e 11,65 em matemática, onde novamente a maior parcela dos diferenciais parece ser decorrente de características não observáveis.

**Tabela 5 – Resultados da decomposição de Oaxaca-Blinder**

<b>O aluno iniciou os estudos antes do 1º ano<sup>(*)</sup></b>				
	<b>5º Ano</b>		<b>9º Ano</b>	
	<b>Português</b>	<b>Matemática</b>	<b>Português</b>	<b>Matemática</b>
Diferença total	-10,5212***	-10,4954***	-11,9811***	-11,6540***
Diferença explicada	-3,7519***	-3,2618***	-3,8101***	-3,8868***
Diferença não explicada	-6,7693***	-7,2336***	-8,1710***	-7,7672***
<b>Os pais do aluno o incentivam a estudar<sup>(*)</sup></b>				
	<b>5º Ano</b>		<b>9º Ano</b>	
	<b>Português</b>	<b>Matemática</b>	<b>Português</b>	<b>Matemática</b>
Diferença total	-20,7402***	-34,4383***	-19,3218***	-9,3992
Diferença explicada	-4,7587**	-4,1434**	-7,2957***	-5,7595**
Diferença não explicada	-15,9815***	-30,2949***	-12,0261*	-3,6397

<sup>15</sup> Lembrando que os fatores observáveis considerados aqui são os seguintes: sexo do aluno, raça do aluno, presença de computador com internet no domicílio, incentivo dos pais ao estudo, hábito de leitura, o fato do aluno trabalhar ou não, o momento em que o aluno iniciou seus estudos, se o aluno reprovou alguma vez, se o aluno costuma fazer a tarefa de casa, as condições das salas de aula e da biblioteca da escola, a experiência do diretor, a experiência e a escolaridade do professor.

**O aluno faz a tarefa de casa sempre ou quase sempre<sup>(\*)</sup>**

	5º Ano		9º Ano	
	Português	Matemática	Português	Matemática
Diferença total	-11,4035***	-7,4356***	-8,5517***	-16,1854***
Diferença explicada	-4,4387***	-0,8922	-5,0863***	-3,0816***
Diferença não explicada	-6,9648***	-6,5434***	-3,4654**	-13,1038***

  

	5º Ano		9º Ano	
	Português	Matemática	Português	Matemática
Diferença total	-20,9156***	-16,9656***	-21,9110***	-22,6848***
Diferença explicada	-4,2031***	-2,3066***	-4,2194***	-2,6081***
Diferença não explicada	-16,7125***	-14,6590***	-17,6916***	-20,0767***

  

	5º Ano		9º Ano	
	Português	Matemática	Português	Matemática
Diferença total	-65,7505***	-64,9746***	-72,6118***	-71,1269***
Diferença explicada	-2,5317***	-2,5193***	-2,6863***	-3,4886***
Diferença não explicada	-63,2188***	-62,4553***	-69,9255***	-67,6383***

  

	5º Ano	9º Ano
<b>Número de observações</b>	<b>1.955</b>	<b>3.144</b>

**Fonte:** Resultados da pesquisa

**Nota:** \*Estatisticamente significativa a 10%; \*\*Estatisticamente significativa a 5%; \*\*\*Estatisticamente significativa a 1%. <sup>(\*)</sup>Valores negativos para as diferenças indicam vantagem para os estudantes que iniciaram os estudos antes do 1º ano, para os estudantes cujos pais os incentivam a estudar, para os alunos que fazem a tarefa de casa sempre ou quase sempre, para aqueles que nunca reprovaram e para os discentes que apresentaram desempenho acima da média. Valores positivos indicam vantagens para os casos contrários.

Na segunda estimação – utilizando a variável de incentivo dos pais ao estudo como forma de segregar os estudantes – os resultados sugerem que tal incentivo apresenta impactos positivos no desempenho escolar. Na prova de português, tanto no 5º quanto no 9º ano, o diferencial de desempenho calculado ficou em torno de 20 pontos. Em matemática, apenas o resultado do 5º ano mostrou-se estatisticamente significativo, indicando um diferencial de 34,38 pontos. Mais uma vez, é notório o importante papel desempenhado pelos fatores não explicados.

Os resultados do terceiro modelo indicam que os discentes que costumam fazer a tarefa de casa tendem a apresentar melhor desempenho do que aqueles que não fazem-na com tanta frequência, ou nunca a fazem. Para os alunos do 5º ano, na prova de português, a diferença total encontrada foi de 11,40 pontos, em que quase 7 pontos foram atribuídos aos fatores não explicados. Um resultado que merece destaque é o diferencial observado no desempenho da prova de matemática dos alunos do 9º ano: cerca de 16,18 pontos. Novamente, os resultados apontam para uma maior importância relativa dos componentes não observáveis.

Mais adiante, segregando os alunos entre os que nunca reprovaram e os que já reprovaram ao menos uma vez, nota-se que os alunos do primeiro grupo apresentam um melhor desempenho tanto em português quanto em matemática, para ambos os anos. Da mesma forma como foi observado nos demais modelos, a parcela do diferencial total decorrente dos fatores não observáveis é majoritariamente dominante.

Por fim, o último modelo estimado buscou mensurar o diferencial de desempenho entre os estudantes acima e abaixo da média, verificando que esse diferencial é decorrente não apenas dos fatores observáveis, mas também dos fatores não observáveis. Como pode ser visto na Tabela 5, todos os resultados desse modelo mostraram-se estatisticamente

significativos. Para os estudantes do 5º ano, os diferenciais de desempenho nas provas de português e matemática foram de, respectivamente, 65,75 e 64,97 pontos. Já para os estudantes do 9º ano, os resultados foram de 72,61 e 71,13 pontos. Em ambos os anos, praticamente toda a parcela dos diferenciais pode ser atribuído aos fatores não explicados.

Os resultados obtidos com os cinco modelos acima descritos corroboraram o argumento apresentado neste artigo de que a diferença de desempenho existente entre boa parte dos estudantes é decorrente do efeito contingência, ou seja, de características particulares, inatas, que geralmente não podem ser observadas diretamente. Isso fica evidente ao se observar que mesmo controlando e considerando variáveis reconhecidamente importantes na determinação do desempenho estudantil, o que inclui também as *proxies* de esforço e incentivo, a maior parte dos diferenciais ainda recai sobre os fatores não observáveis.

#### 4.3 Estimação da decomposição de Oaxaca-Blinder com pareamento

Como ressaltado anteriormente, é possível tornar a amostra mais homogênea empregando a técnica de Pareamento por Escore de Propensão. Dessa forma, as comparações realizadas são feitas entre estudantes com características observáveis similares, tornando os resultados encontrados mais robustos. Na Tabela 6 encontram-se os resultados da decomposição de Oaxaca-Blinder com PSM, considerando as mesmas variáveis explicativas e de segregação apresentadas na Tabela 4 e adotadas nos modelos da Tabela 5.

**Tabela 6 – Resultados da decomposição de Oaxaca-Blinder com PSM**

<b>O aluno iniciou os estudos antes do 1º ano<sup>(*)</sup></b>				
	<b>5º Ano</b>		<b>9º Ano</b>	
	<b>Português</b>	<b>Matemática</b>	<b>Português</b>	<b>Matemática</b>
Diferença total	-12,2964***	-11,7686***	-12,1739***	-12,0067***
Diferença explicada	-4,0706***	-3,9190***	-4,4636***	-4,0399***
Diferença não explicada	-8,2258***	-7,8496***	-7,7103***	-7,9668***
<b>Os pais do aluno o incentivam a estudar<sup>(*)</sup></b>				
	<b>5º Ano</b>		<b>9º Ano</b>	
	<b>Português</b>	<b>Matemática</b>	<b>Português</b>	<b>Matemática</b>
Diferença total	-21,2761***	-34,0601***	-18,6134**	-11,8045
Diferença explicada	-3,0849	-1,1184	-4,3859*	-5,8221**
Diferença não explicada	-18,1912***	-32,9417***	-14,2275*	-5,9824
<b>O aluno faz a tarefa de casa sempre ou quase sempre<sup>(*)</sup></b>				
	<b>5º Ano</b>		<b>9º Ano</b>	
	<b>Português</b>	<b>Matemática</b>	<b>Português</b>	<b>Matemática</b>
Diferença total	-8,2297***	-6,1734**	-7,5974***	-16,7835***
Diferença explicada	-3,9697***	-0,9064	-5,0017***	-3,0876***
Diferença não explicada	-4,2600	-5,2670*	-2,5957**	-13,6959***
<b>O aluno nunca foi reprovado<sup>(*)</sup></b>				
	<b>5º Ano</b>		<b>9º Ano</b>	
	<b>Português</b>	<b>Matemática</b>	<b>Português</b>	<b>Matemática</b>
Diferença total	-20,7808***	-17,6740***	-21,5461***	-23,7631***
Diferença explicada	-3,9652***	-2,1009**	-4,1424***	-2,2953***



Diferença não explicada	-16,8156***	-15,5731***	-17,4037***	-21,4678***
<b>O aluno apresentou desempenho acima da média<sup>(*)</sup></b>				
	<b>5º Ano</b>		<b>9º Ano</b>	
	<b>Português</b>	<b>Matemática</b>	<b>Português</b>	<b>Matemática</b>
Diferença total	-67,7975***	-65,6693***	-71,6925***	-71,6389***
Diferença explicada	-1,9836***	-2,5714***	-2,7885***	-3,8139***
Diferença não explicada	-65,8139***	-63,0979***	-68,9040***	-67,8250***
<b>Número de observações</b>	<b>1.191</b>	<b>1.214</b>	<b>2.505</b>	<b>2.510</b>

**Fonte:** Resultados da pesquisa

**Nota:** \*Estatisticamente significativa a 10%; \*\*Estatisticamente significativa a 5%; \*\*\*Estatisticamente significativa a 1%. <sup>(c)</sup>Valores negativos para as diferenças indicam vantagem para os estudantes que iniciaram os estudos antes do 1º ano, para os estudantes cujos pais os incentivam a estudar, para os alunos que fazem a tarefa de casa sempre ou quase sempre, para aqueles que nunca reprovaram e para os discentes que apresentaram desempenho acima da média. Valores positivos indicam vantagens para os casos contrários.

De modo geral, os resultados da decomposição de Oaxaca-Blinder com e sem PSM foram muito similares, com pequenas diferenças na magnitude dos coeficientes. Comparando o primeiro modelo de ambas as decomposições, nota-se que com PSM o diferencial total nas notas de português e matemática é mais elevado tanto no 5º quanto no 9º ano. Dessa forma, os resultados indicam que aqueles alunos que iniciaram seus estudos na creche ou pré-escola apresentam, em média, desempenho superior de aproximadamente 12 pontos, sendo que grande parcela desse diferencial é ocasionado por fatores não explicados.

No segundo modelo, em que a segregação dos grupos foi feita através da variável de incentivo dos pais aos estudos, observa-se que a parcela do diferencial total atribuída aos fatores não observáveis foi maior no modelo com PSM. Também é interessante observar que, em ambos os casos, o coeficiente de diferença total da disciplina de matemática para os alunos do 9º ano não apresentou significância estatística, o que sugere que o desempenho nessa matéria não é diretamente afetado pelo incentivo dos pais.

Com relação ao terceiro modelo, cuja variável de controle usada para separar os grupos foi referente ao esforço dos alunos em fazer a tarefa de casa, tem-se que a diferença total foi menor para os casos com PSM, à exceção de matemática para estudantes do 9º ano. Foi percebido, conforme já esperado, a existência de um diferencial favorável aos estudantes que costumam fazer suas tarefas, em que a maior parte desse diferencial, em quase todas as análises, deu-se principalmente em decorrência de fatores não explicados.

O quarto modelo separou a amostra com base na ocorrência ou não de reprovação por parte dos alunos. Todos os coeficientes desse modelo apresentaram significância estatística, e os maiores diferenciais totais encontrados deram-se para os alunos do 9º ano. Em média, os resultados dos alunos que nunca foram reprovados é mais de 20 pontos superior em relação aos que já foram reprovados. Novamente, a maior parte do diferencial total recai sobre as características não observáveis.

Por último, tem-se o modelo que busca decompor a diferença de desempenho entre os estudantes que obtiveram desempenho acima e abaixo da média, considerando apenas os estudantes dentro da região de suporte comum. Os resultados foram similares ao modelo sem PSM, indicando que grande parte do diferencial existente é devido aos fatores não explicados.

Como pôde ser visto nesta seção de resultados, a decomposição de Oaxaca-Blinder com PSM também apresenta evidências favoráveis ao efeito contingência descrito por Rawls (1971), como forma de explicar o diferencial de desempenho entre os discentes. Isso fica claro pois, mesmo controlando as comparações para que as essas fossem realizadas entre indivíduos similares, e considerando o efeito da escola, do domicílio, dos professores, do

diretor e incluindo variáveis de esforço e incentivo, ainda existe um diferencial não explicado que é oriundo das particularidades e capacidades inatas de cada estudante.

## 5. Considerações finais

São vários os trabalhos na literatura que buscam identificar os determinantes do desempenho estudantil e a maior parte dessas pesquisas identifica nas características individuais dos alunos, bem como no *background* familiar, os principais fatores que explicam os resultados observados. Mais recentemente, as questões de justiça e equidade vêm sendo debatidas nesse contexto, e o papel da desigualdade de oportunidade tem ganhado cada vez mais espaço. Todavia, apesar de todo o avanço já realizado, ainda não há consenso sobre como seria possível oferecer educação de qualidade de modo mais abrangente e justo.

Rawls (1971) apresentou uma teoria considerando que alguns indivíduos são, por natureza, diferentes em suas dotações, isto é, diferentes em suas habilidades e talentos natos. Desse modo, existiria uma desigualdade de contingência que explicaria o porquê de indivíduos com acesso aos mesmos recursos e benefícios apresentarem resultados tão diferentes. Tendo isso em vista, o presente estudo teve por objetivo verificar a ocorrência de desigualdade de contingência entre os alunos do 5º e 9º ano das escolas públicas de João Pessoa, a partir do desempenho em português e matemática aferidos pela Prova Brasil de 2011.

Para isso, adotou como estratégia empírica a decomposição de Oaxaca-Blinder combinado com o método de Pareamento por Escore de Propensão, para garantir que as comparações fossem realizadas apenas entre estudantes com características observáveis semelhantes.

A partir dos resultados aferidos pela decomposição de Oaxaca-Blinder, pode-se dizer que existe desigualdade de contingência entre os alunos do 5º e 9º, tanto em português quanto em matemática, uma vez que se observaram diferenças de mais de 65 pontos, na média, atribuídas majoritariamente às características não observáveis. Portanto, mesmo utilizando as proxies de esforço e incentivo foi possível observar diferenças não explicadas.

Dessa forma, os resultados indicam que o diferencial de desempenho entre os estudantes pesquisados é oriundo da desigualdade de contingência, ou seja, das diferenças nas habilidades inatas. Sendo assim, há razões para acreditar que mesmo em situações onde não haja desigualdade de oportunidade, ainda será observado diferenças no desempenho dos estudantes.

## Referências

ARAUJO, E. S. P.; ALMEIDA, A. T. C. Avaliação dos resultados educacionais dos alunos das escolas municipais de João Pessoa-PB. **MPGOA**, v. 2, p. 46-63, 2013.

BARROS, R. P.; MENDONÇA, R.; SANTOS, D. D.; QUINTAES, G. Determinantes do desempenho educacional no Brasil. **Pesquisa e Planejamento Econômico**, v. 31, n. 1, p. 1-42, abr. 2001.

BECKER, S. O.; ICHINO, A. Estimation of average treatment effects based on propensity scores. **The Stata Journal**, v. 2, n. 4, p. 358-377, 2002.

BLINDER, A. S. Wage discrimination: reduced form and structural estimates. **Journal of Human Resources**, v.8, n. 4, p. 436–455, 1973.

BRASIL. Ministério da Educação e Cultura / Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Microdados Prova Brasil 2011**: manual do usuário. Brasília-DF: INEP, 2012.

CURI, A. Z.; MENEZES FILHO, N. A. Os efeitos da pré-escola sobre os salários, a escolaridade e a proficiência escolar. In: XXXIV Encontro Nacional de Economia, 2006, Salvador. Anais do XXXIV Encontro Nacional de Economia. São Paulo: ANPEC, 2006. v. 1.

FIGUEIREDO, E.; SILVA, C. R. F. E. Desigualdade de Oportunidades no Brasil: Uma Decomposição Quantílica Contrafactual. **Pesquisa e Planejamento Econômico** (Rio de Janeiro), v. 42, p. 29-48, 2012.

GONDIM, E. John Rawls: a educação política. **Revista Thaumazein (Santa Maria)**, v. 5, p. 1-15, 2010.

GONDIM, E.; RODRIGUES, O. M. John Rawls: educação, cidadania e equilíbrio reflexivo. **Revista Filosofia Capital**, v. 6, p. 25-33, 2011.

HANUSHEK, E. A. **Education Production Functions**. Stanford: Stanford University, January 2007.

HANUSHEK, E. A. **Do Teachers Make a Difference?** In: U.S. Office of Education. The Production of Education, Teacher Quality, and Efficiency. Washington, D.C.: Government Printing Office, 1970, pp. 79-99.

HESHMATI, A. **Inequalities and their measurement**. Unpublished manuscript, 2004.

JANN, B. The Blinder–Oaxaca decomposition for linear regression models. **The Stata Journal**, v. 8, n. 4, p. 453–479, 2008.

KHANDKER, S. R.; KOOLWAL, G. B.; SAMAD, H. A. **Handbook on impact evaluation: Quantitative Methods and Practices**. Washington, DC: The World Bank, 2010.

KRANICH, L. Equitable opportunities: an axiomatic approach. **Journal of Economic Theory**, v. 71, p. 132-147, 1996.

MENEZES FILHO, N. (org.). **Avaliação econômica de projetos sociais**. 1. ed. São Paulo: Dinâmica Gráfica e Editora, 2012.

MENEZES FILHO, N. Os Determinantes do Desempenho Escolar no Brasil. Instituto Futuro Brasil, São Paulo, Ibmecc – SP, 2007.

OAXACA, R. Male–female wage differential in urban labor markets. **International Economic Review**, v. 14, n. 3, p. 693–709, 1973.

PATTANAIK, P. K.; XU, Y. On ranking opportunity sets in terms of freedom of choice. **Recherches Économiques de Louvain**, v. 56, n. 3/4, 1990.

RAWLS, J. **A theory of justice**. United States of America: Harvard University Press, 1971.

ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, v. 70, n. 1, p. 41–55, 1983.

SEN, A. **Commodities and Capabilities**. Amsterdam: North-Holland, 1985.

SOUZA, A. M. Determinantes da aprendizagem em escolas municipais. **Ensaio. Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 13, n.48, p. 271-292, 2005.

THOMSON, W. Fair allocation rules. In: ARROW, K.; SEN, A.; SUZUMURA, K. (Ed.). **Handbooks in economics: social choice and welfare**. Amsterdam: Elsevier, 2011.

## Apêndice

**Tabela 7 - Estatísticas descritivas da base de dados completa para os alunos do 5° e 9° ano**

	Variáveis	5° ano					9° ano				
		Obs.	Média	Desvio-padrão	Mínimo	Máximo	Obs.	Média	Desvio-padrão	Mínimo	Máximo
<b>Características do Aluno e do Domicílio</b>	Sexo_aluno	5.878	0,5090	0,5000	0	1	4.668	0,4653	0,4988	0	1
	Raça_aluno_branco	4.954	0,2814	0,4497	0	1	4.325	0,2460	0,4307	0	1
	Trabalho_aluno	5.780	0,1351	0,3419	0	1	4.651	0,1219	0,3272	0	1
	Incentivo_aluno	5.583	0,9645	0,1850	0	1	4.597	0,9852	0,1207	0	1
	Ler_livros_sempre	5.228	0,4216	0,4939	0	1	4.279	0,2720	0,4451	0	1
	Faz_tarefa_português_sempre	5.906	0,7345	0,4416	0	1	4.664	0,5375	0,4986	0	1
	Faz_tarefa_matemática_sempre	5.926	0,7373	0,4402	0	1	4.667	0,4958	0,5000	0	1
	Início_dos_estudos	5.807	0,7549	0,4301	0	1	4.628	0,8466	0,3604	0	1
	Reprovação_aluno	5.911	0,5945	0,4910	0	1	4.684	0,5645	0,4959	0	1
	Computador_com_internet	6.022	0,3678	0,4823	0	1	4.694	0,4902	0,5000	0	1
	Escol_mãe_ens_superior	3.630	0,1576	0,3644	0	1	3.905	0,0883	0,2838	0	1
	Prof_corrige_tarefa_port_sempre	5.886	0,7727	0,4191	0	1	4.686	0,7414	0,4379	0	1
	Prof_corrige_tarefa_mat_sempre	5.956	0,7915	0,4063	0	1	4.679	0,7476	0,4344	0	1
	Proficiência_português	6.172	185,1591	40,5794	83.4419	327.4591	4.753	239,3428	45,0441	106,0994	360,7970
	Proficiência_matemática	6.172	200,3675	39,4197	99.6468	337.1514	4.753	241,0977	44,7000	112,6782	396,0480
<b>Características da Escola</b>	Salas_escola_bom	5.991	0,7947	0,4040	0	1	6.448	0,7402	0,4385	0	1
	Banheiros_escola_bom	6.107	0,5800	0,4936	0	1	6.437	0,4513	0,4977	0	1
	Biblioteca_escola_bom	6.119	0,7395	0,4389	0	1	6.399	0,7862	0,4100	0	1
	Laboratório_escola_bom	5.957	0,4773	0,4995	0	1	6.159	0,4208	0,4937	0	1
<b>Características do Professor</b>	Escol_professor_superior	3.716	0,8539	0,3533	0	1	5.607	0,9941	0,0765	0	1
	Experiência_professor	3.759	0,7100	0,4538	0	1	5.540	0,6139	0,4869	0	1
<b>Características do Diretor</b>	Escolaridade_diretor_superior	6.027	0,947735	0,2226	0	1	6.468	0,9743	0,1581	0	1
	Experiência_diretor	6.027	0,963166	0,1884	0	1	6.468	0,9807	0,1377	0	1

Fonte: Microdados Prova Brasil 2011.