

CORRESPONDÊNCIA DE RAÇA PROFESSOR-ALUNO E O GAP EDUCACIONAL RACIAL

Janaína Rodrigues Feijó¹ Valdemar Rodrigues de Pinho Neto² João Mário Santos de França³

RESUMO

Estudos vem destacando cada vez mais a importância do professor no desempenho escolar dos alunos e tem sido apontado que estes tendem a se identificar com professores que possuem características similares (como raça). Isso ocorre principalmente no início da adolescência, fase em que os estudantes começam a reconhecer mais claramente essas associações. Esse artigo busca investigar os efeitos da congruência étnica do professor-aluno preto/pardo sobre o *gap* racial (pretos/pardos *versus* não pretos/pardos) de performance escolar. Analisamos esses efeitos na média e ao longo da distribuição de pontuações das provas de Matemática e Linguagens e Códigos do Enem. Os dados são provenientes do Censo Escolar e do Enem no ano de 2019, disponibilizados pelo INEP. Os resultados sugerem efeitos positivos do aumento da representatividade de negros/pardos no corpo docente das escolas sobre a redução do *gap* racial de desempenho. Em média, as diferenças de pontuação entre pretos/pardos e demais raças reduzem cerca de 31,74% em Matemática e 15,04% em Linguagens e Códigos quando alunos pretos/pardos são expostos a um corpo docente preto/pardo.

Palavras-Chaves: Desempenho Educacional. Disparidades raciais. Associação professor-aluno.

ABSTRACT

Some studies have already highlighted the importance of the teacher in students' school performance and note that students tend to identify with teachers who share the same race. This occurs mainly in early adolescence, when pupils begin to recognize these connections more clearly. This paper investigates the effects of ethnic congruence (black/brown) of the teacher and student on the racial gap (black/brown versus non-black/brown) in school performance. We analyze these effects on average and across the distribution/deciles of scores. We use microdata from the School Census and ENEM data in 2019, provided by INEP. The main results suggest positive effects of increasing the representation of blacks/browns in the teaching staff of schools to reduce the racial performance gap. On average, the differences in scores between students blacks/browns and other races reduced by around 31.74% in Mathematics and 15.04% in Languages and Codes, when these students were exposed to black/brown teachers instead of from other races.

Keywords: Educational Performance. Racial Differences. Teacher-Student Demographic Matches.
JEL CODES: I20, I21, I24

¹ Doutora em Economia - UFC/CAEN e Pesquisadora no Instituto Brasileiro de Economia (FGV IBRE).

² Professor na Graduação de Economia - FGV EPGE - Escola Brasileira de Economia e Finanças da Fundação Getúlio Vargas-RJ e Pesquisador no Centro de Estudos Empíricos em Economia (FGV CEEE).

³ Professor no Programa de Pós-graduação de Economia – CAEN/UFC e IPECE/SEPLAG.

1. INTRODUCAO

As análises sobre a performance escolar dos estudantes têm fomentado discussões acadêmicas e contribuído para a criação e direcionamento de políticas públicas educacionais. Já está bem documentado na literatura os efeitos benéficos da educação sobre a vida do indivíduo e na redução das desigualdades sociais, gerando externalidades positivas para a sociedade. Parte dessas pesquisas tem focado tanto nos determinantes quanto na investigação acerca das desigualdades de desempenho entre alunos.

Em relação a essas desigualdades, estudos brasileiros vêm mostrando que há diferenças de performance relevantes entre alunos pretos e brancos. A diferença de desempenho acontece mesmo quando são comparados apenas alunos brancos e pretos de alto nível socioeconômico. Por exemplo, no 5º ano, 74,8% dos alunos brancos de alto nível socioeconômico alcançaram bom desempenho em Língua Portuguesa enquanto para pretos esse percentual foi de 48,9% (IEDE, 2021). Vale mencionar que as evidências nacionais nesse campo em geral tratam-se de correlações, não controlando por outros fatores que potencialmente possam estar explicando as diferenças de performance entre raças.

As diferenças nos desfechos educacionais acontecem ao longo de toda a vida escolar e consolidam as desvantagens entre esses grupos étnicos na fase adulta. Vaz (2020) argumenta que os negros, mesmo quando avançam no nível educacional, têm maiores dificuldades de converter a escolaridade em rendimentos, não conseguindo eliminar as desigualdades raciais nos rendimentos do trabalho. Dentre as causas para esse quadro, a autora destaca as desvantagens dos negros em relação aos brancos no que se refere ao acesso à educação de qualidade, ao *background* cultural e ao *network* que media o processo de socialização, dificultando ascenderem às posições mais prestigiadas do mercado de trabalho.

Nesse sentido, a análise da relação professor-aluno vem trazendo novos *insights* na compreensão das disparidades educacionais entre grupos étnicos. Os professores assumem um papel importante na formação das crenças dos alunos sobre suas perspectivas escolares, principalmente entre alunos que raramente interagem com pessoas com ensino superior fora do ambiente escolar, influenciando consequentemente a performance estudantil (GERSHENSON, HOLT E PAPAGEORGE, 2016). De acordo com Dee (2005), a presença de um professor demograficamente semelhante aumenta a motivação e as expectativas acadêmicas de um aluno, podendo ocorrer situações em que os alunos percebem estereótipos (por exemplo, alunos negros com professores brancos) e experimentam uma apreensão que retarda sua identificação acadêmica e consequentemente seu desempenho. Essas identificações também ocorrem na adolescência. Alguns trabalhos têm evidenciado que os professores apresentam maiores expectativas sobre o desempenho dos alunos que possuem a sua mesma cor/raça (EHRENBERG, GOLDHABER E BREWER, 1995; SAFT E PIANTA, 2001; TENENBAUM E RUCK, 2007). Essas associações podem incentivar grupos que geralmente apresentam desvantagens educacionais, como os pretos/pardos, influenciando o *gap* racial de performance.

Egalite e Kisida (2018) salientam que, em média, os professores de grupos sociais minoritários são bem aceitos por estudantes de minorias e podem traduzir suas experiências e identidades para formar relações com alunos que até mesmo não compartilham a mesma raça/etnia. Esses professores

geralmente possuem experiências pessoais que os permitem compreender os estereótipos raciais e ajudam os estudantes a combatê-los. Essa relação construída com percepções positivas dos professores tem um alto potencial de contribuir para o bom desempenho escolar dos alunos.

Embora existam evidências de que os estudantes que são da mesma raça que seus professores tendem a ter um desempenho acadêmico melhor, é necessário considerar o tempo dessa interação. Analisando estudantes do Tennessee, Penney (2017) verificou que exposições antecipadas a professores de mesma raça geram benefícios que persistem no médio prazo. Esse efeito de correspondência de mesma raça pode explicar uma parte não trivial do *gap* de pontuação do teste entre pretos e brancos. Por outro lado, Banerjee (2018), analisando também a correspondência étnica, concluiu que a designação para professores de mesma raça não afeta o crescimento das notas em leitura e matemática dos alunos estrangeiros entre o jardim de infância e o terceiro ano. Yarnell e Bohrstedt (2018) também examinaram a relação racial aluno-professor com o desempenho dos alunos negros. Utilizando dados de estudantes da 4ª série que realizaram a prova de leitura nos EUA e equações estruturais multiníveis constataram uma associação positiva entre professores e alunos negros, já sugeridas por Clotfelter et al., 2007; Dee, 2004 e Egalite et al., 2015.

No Brasil, embora o tema já tenha sido levantado, praticamente não há estudos analisando o efeito da correspondência aluno-professor nos desfechos educacionais. Dessa forma, esse artigo tem como objetivo investigar o efeito da associação do professor-aluno via raça sobre o *gap* racial de desempenho (negros/pardos versus demais raças) nas provas de Matemática (MT) e Linguagens e Códigos (LC) do Enem. Os dados para o desenvolvimento dessa análise são provenientes dos microdados do Enem e do Censo Escolar do ano de 2019, disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). A amostra é composta por estudantes brasileiros que estavam finalizando o 3º do ensino médio.

A estratégia empírica adotada consiste na estimação de um modelo econométrico visando explicar o desempenho dos alunos e entender o diferencial de desempenho entre alunos pretos/pardos e demais raças. A variável dependente (pontuações em MT e LC) foi regredida sobre uma variável que indica “1” se o aluno for preto/pardo, a fração de professores pretos/pardos no ensino médio e a interação dessas duas variáveis, seguindo uma abordagem similar a um modelo de diferenças-em-diferenças. Com isso, podemos mensurar o quanto que a exposição aos professores com determinada característica étnica (pretos e pardos) similar a dos alunos pode reduzir o *gap* racial de desempenho. Para controlar pelo maior número de características observáveis possível, e assim minimizar os efeitos de uma possível endogeneidade na relação de interesse, o modelo inclui um conjunto de efeitos fixos de escola e um vetor de variáveis observáveis (incluindo renda familiar, educação dos pais, infraestrutura domiciliar e tamanho da família). Ainda, visando contornar a endogeneidade na relação estimada, nesse trabalho também utilizaremos o algoritmo Coarsened Exact Matching-CEM, desenvolvido por Iacus, King e Porro (2008), para fazer o pareamento entre professores-alunos, de modo a controlar por possíveis diferenças na distribuição/alocação de professores/alunos com base em, por exemplo, características étnicas.

Devido a estrutura dos dados, não é possível obter informações anteriores sobre os alunos, como por exemplo, saber quais e a quantos professores negros/pardos os alunos tiveram contato ao longo da vida escolar. Isso também impossibilita o *match* perfeito entre professor e aluno. Buscando contornar, em parte, essas limitações, consideramos a exposição do aluno ao corpo docente que lecionava no ensino médio presente em cada escola nas disciplinas de matemática e de português.

Dessa forma, a análise considera a exposição/contato dos alunos a um corpo docente preto/pardo e analisamos o que ocorre com o *gap* racial de desempenho nesse contexto. Também analisamos o que ocorre com esses *gaps* ao longo da distribuição de notas (decil 1 ao 9).

Os resultados mostram que, em média, alunos pretos/pardos tiram pontuações inferiores as demais raças em 12,49 pontos em MT e em 7,91 pontos em LC quando são expostos a um corpo docente formado somente por pessoas brancas. Essas diferenças caem, respectivamente, para 8,52 e 6,72 pontos quando eles passam a ter contato com um corpo docente preto/pardo, representando uma redução de 31,74% e de 15,04% nos diferenciais raciais de desempenho em MT e LC.

O artigo está estruturado em quatro seções incluindo esta Introdução. Na segunda seção descreve-se os dados e a estratégia empírica. Os resultados das estimações são apresentados na terceira seção e na última seção tem-se as considerações finais.

2. DADOS E ESTRATEGIA EMPÍRICA

3.1 Dados

A análise desenvolvida nessa pesquisa está centrada no desempenho dos estudantes ao final do ensino médio regular nas provas de Linguagens e Códigos (LC) e Matemática (MT), período de finalização do ensino básico. Os dados utilizados são provenientes do Enem e do Censo Escolar 2019, ambos disponibilizados pelo Inep. Além das pontuações nas competências, a base do Enem fornece um conjunto de informações socioeconômicas dos participantes, mas não sobre suas escolas e professores, sendo necessário recorrer ao Censo Escolar para obter tais informações.

Em 2019 mais de 5 milhões de pessoas realizaram o Enem, contudo apenas 1,5 milhões destes estavam finalizando o 3º ano do ensino médio, grupo de interesse do estudo. Dessa amostra, 73% dos estudantes estavam na rede de ensino regular⁴ e possuíam o código da escola, sendo este último fundamental para agregar as informações com os dados do censo escolar. Sendo assim, o grupo de interesse foi composto por 1.066.872 estudantes, dos quais 947.586 fizeram a prova de linguagens e códigos e 908.327 a prova de matemática⁵. Em relação as informações dos professores, cerca de 20% dos docentes que lecionavam no 3º ano do ensino regular não reportaram como se autodeclaravam. A falta dessa informação, dado que é uma variável estratégica para a análise, impossibilitou realizar o *merge* das duas bases de dados para alguns estudantes. A amostra final foi constituída de 723.057 que realizaram a prova de matemática e 762.945 que estavam presentes na prova de Linguagens e Códigos.

A descrição das variáveis pode ser visualizada na Tabela 1, abaixo. Nessa tabela, as variáveis de controle foram agrupadas em três grupos. O primeiro diz respeito as características intrínsecas dos estudantes, tais como raça e gênero. No segundo enquadram-se as características socioeconômicas,

⁴ Esse filtro foi realizado para não viesar as análises, uma vez que estudantes das escolas profissionalizantes, integrados e EJA possuem escolas e grades curriculares distintas.

⁵ Optou-se por não analisar as demais provas que compõem o Enem (Redação, Ciências da Natureza e Ciências Humanas) por não ser possível obter uma relação “um-para-um” entre prova do Enem e professor que leciona na competência da prova. Por exemplo, na prova Ciências da Natureza mede-se o desempenho do aluno com relação a Biologia, Física e Química, não sendo trivial saber o efeito de um professor específico de uma dessas matérias sobre o desempenho do aluno na prova Ciências da Natureza.

como *background* familiar (renda familiar, escolaridade do pai e da mãe), infraestrutura do domicílio - IFD em que vivem e o tamanho da sua família (quantidade de pessoas que moram na mesma residência do estudante). O IFD foi criado a partir da análise de componentes principais, técnica de análise multivariada comumente utilizada para o desenvolvimento de indicadores sintéticos que permitem condensar as informações contidas em várias variáveis.

Tabela 1 – Descrição das variáveis

Variáveis	Descrição
Dependentes	
Pontuação_LC	Pontuação na prova de Linguagens e Códigos
Pntuação_MT	Pontuação na prova de Matemática
Explicativas	
<i>Controles - Características individuais</i>	
Genero	"1" se o participante for do gênero feminino e "0" se o participante for do gênero masculino
Preto/pardo	"1" se o participante se autodeclarar preto ou pardo e "0" caso contrário
<i>Demais controles - Características socioeconômicas</i>	
Escolaridade do pai	Dummies indicadoras da escolaridade do pai ou do homem responsável: Educpai1 (Categoria base): Nunca estudou ou possui escolaridade inferior ao ensino médio completo. Educpai2: Completou o Ensino Médio, mas não completou a Faculdade. Educpai3: Possui Ensino Superior Completo ou Pós-Graduação. Educpai4: participantes que não sabiam a escolaridade do pai.
Escolaridade da mãe	Dummies indicadoras da escolaridade da mãe ou da mulher responsável: Educmae1(Categoria base): Nunca estudou ou possui escolaridade inferior ao ensino médio completo. Também foi inserido nessa categoria os participantes que não sabiam a escolaridade da mãe. Educmae2: Completou o Ensino Médio, mas não completou a Faculdade. Educmae3: Possui Ensino Superior Completo ou Pós-Graduação. Educmae4: participantes que não sabiam a escolaridade da mãe.
Renda	Dummies indicadoras das faixas de renda familiar mensal. Renda1 (Categoria base): R\$0,00 a R\$998,00; Renda2: R\$998,01 a R\$1.497,00; Renda3: R\$1.497,01 a R\$ 1996,00; Renda4: R\$1996,01 a R\$2.495,00; Renda5: R\$2.495,01 a R\$3.992,00; Renda6: R\$3.992,01 a R\$6.986,00; Renda 7: mais de R\$6.986,01.
Índice de infraestrutura domiciliar (IFD)	Leva em consideração se na residência dos inscritos existiam: pelo menos 1 máquina de lavar, pelo menos 1 forno micro-ondas, pelo menos 1 carro, pelo menos 1 computador, pelo menos 1 geladeira, aspirador de Pó, TV por assinatura, aparelho DVD, telefone Fixo, Internet, mais de um banheiro, mais de dois quartos, mais de uma TV e quantidade de celulares (0 a 4)
Tamanho da família (Tam_fam)	Tamanho da família. Representa a quantidade de pessoas que moram na residência do participante. A quantidade varia de 1 a 20.
<i>Características dos Professores</i>	
preto/pardo_mt	"1" se o professor que leciona matemática no 3º ano do EM é autodeclarado preto ou pardo e "0" caso contrário. Para cada escola calcula-se a proporção de professores que se autodeclararam preto/pardo e que lecionam matemática.
preto/pardo_pt	"1" se o professor que leciona português no 3º ano do EM é autodeclarado preto ou pardo e "0" caso contrário. Para cada escola calcula-se a proporção de professores que se autodeclararam preto/pardo e que lecionam matemática

Fonte: própria dos autores com base nos microdados do Censo Escolar e do Enem 2019.

No último grupo tem-se as informações de raça dos docentes. A partir delas calculamos a proporção de professores autodeclarados pretos ou pardos no ensino médio regular por disciplina e por escola. Essa variável funciona como uma *proxy* para captar o efeito da diversidade de raça do corpo docente

do ensino médio no *gap* racial de desempenho, dado que, devido a estrutura dos dados, não é possível sabermos quais foram exatamente os professores que cada aluno teve contato. Logo, essa *proxy* também pode ser entendida como uma referência que o corpo docente passa para os discentes. Além disso, pressupõe-se que grande parte dos alunos tendem a cursar os três anos do ensino médio na mesma escola, aumentando a probabilidade dos alunos já terem tido, de fato, contato com os professores que lecionam no ensino médio. Vale salientar que embora as variáveis de infraestrutura escolar não tenham sido incorporadas diretamente no modelo, fez-se uso de efeitos fixos de escola, nos quais consideram em seu computo essas características, permitindo distinguir as escolas ao máximo possível e fazer comparações intraescola.

3.2 Estratégia Empírica

O modelo econométrico básico a ser estimado nesse artigo tem a seguinte especificação:

$$Y_i = \alpha_e + \gamma A_i + \theta P_i + \varphi A_i P_i + Z_{ie}'\Omega + \varepsilon_i \quad (1)$$

Onde Y_i representa a nota do aluno “i” em uma determinada prova do Enem, no caso, em Matemática e Português. A variável A_i indica valor 1 se o aluno “i” for preto/pardo. Analogamente, a variável P_i representa a fração de professores pretos/pardos no ensino médio. O parâmetro de interesse do estudo, φ , mede a diferença marginal nas notas que um aluno preto/pardo obtém, em média, caso esteja exposto a uma maior/menor proporção de professores que também sejam pretos/pardos. Adicionalmente, o vetor de variáveis observáveis, Z_{ie} , corresponde a um conjunto de informações individuais, tais como, renda familiar, educação dos pais, infraestrutura domiciliar e tamanho da família. O modelo inclui ainda um conjunto efeitos fixos de escola, α_e , que funcionam como *proxies* para todas as características da escola onde o indivíduo “i” estuda⁶. Por fim, o termo ε_i representa o termo de erro idiossincrático que assumimos atender as hipóteses convencionais.

Nota-se que, a partir desse modelo econométrico, é possível entender como a similaridade (em termos de raça) do professor-aluno afeta o *gap* no desempenho dos alunos (também em termos de raça) ao final do ensino médio. Em suma, o principal parâmetro de interesse, φ , tem sua interpretação parecida com o de um modelo de diferença-em-diferenças. Nesse sentido, ele medirá o quanto que a similaridade de raça entre o corpo docente e o aluno contribui para reduzir/aumentar o *gap* de raça entre os estudantes.

Para ver isto, observe que a nota esperada de um aluno preto/pardo (i.e., $A_i = 1$), com dadas características e em uma dada escola, seria:

$$E[Y_i|A_i = 1, Z_{ie}, \alpha_e, P_i] = \alpha_e + \gamma + \theta P_i + \varphi P_i + Z_{ie}'\Omega \quad (2)$$

Analogamente, a nota esperada de um aluno que NÃO seja preto/pardo (i.e., $A_i = 0$), com dadas características e em uma dada escola, seria:

$$E[Y_i|A_i = 0, Z_{ie}, \alpha_e, P_i] = \alpha_e + \theta P_i + Z_{ie}'\Omega \quad (3)$$

⁶ Note que os efeitos fixos de escola englobam inclusive a característica do professor, P_i , que poderá ser omitida nessa versão do modelo sem perda de generalidade.

Note que a diferença entre (2) e (3) fica:

$$E[Y_i|A_i = 1, Z_{ie}, \alpha_e, P_i] - E[Y_i|A_i = 0, Z_{ie}, \alpha_e, P_i] = \gamma + \varphi P_i \quad (4)$$

Portanto o *gap* de desempenho de um aluno preto/pardo depende, de acordo com (3.4), da proporção de professores do ensino médio que também sejam pretos/pardos. Em particular, se não houver professores pretos/pardos (i.e., $P_i = 0$) na escola, o *gap* de desempenho seria dado pelo parâmetro γ , enquanto no caso de todos os professores serem pretos/pardos (i.e., $P_i = 1$) teríamos um *gap* de $\gamma + \varphi$. Ou seja, na prática o parâmetro φ mede justamente o diferencial no *gap* de desempenho que um aluno preto/pardo pode, em média, obter caso ele esteja exposto a professores da mesma raça na escola.

3.2.1 *Coarsened Exact Matching-CEM*

Possivelmente existe alguma endogeneidade na relação entre as notas dos alunos e a possibilidade de eles terem um professor com características étnicas similares. Na prática, embora a distribuição de raças, tanto dos professores quanto dos alunos, possa a princípio ser determinada aleatoriamente na população, pode haver, por exemplo, um maior agrupamento de professores/alunos de uma determinada característica étnica (bem como *background* social, etc.) em certas escolas. Isso, por sua vez, pode afetar a estimação do parâmetro de interesse por introduzir um componente de *assortative matching* entre escolas que também pode influenciar as notas dos alunos. Outra possível fonte de endogeneidade ocorreria se os pais de filhos negros/pardos preferem colocar seus filhos em escolas aonde a proporção de pretos/pardos no corpo docente é maior, por acreditar que isso impactaria nas notas. Ou ainda se professores de uma dada etnia escolhem as escolas onde pretendem trabalhar com base nas características étnicas dos potenciais alunos. Ao menos em parte, os efeitos dessas fontes de endogeneidade podem ser mitigados quando incluímos efeitos fixos de escolas e outros controles de *background* social dos alunos. Ou seja, em última instância estamos comparando alunos dentro da mesma escola e com características socioeconômicas similares.

Alternativamente, com o objetivo de mitigar os efeitos da endogeneidade na relação que buscamos estimar, nesse trabalho também utilizaremos o algoritmo *Coarsened Exact Matching-CEM* para fazer o pareamento entre professores-alunos de modo a controlar por alguma diferença/desequilíbrio na distribuição/alocação de professores e alunos a partir de suas características étnicas. O pareamento realizado por meio do *CEM*, desenvolvido por Iacus, King e Porro (2008), reduz o desequilíbrio na distribuição empírica entre “tratados” e “controles” com base em características pré-escolhidas pelo próprio pesquisador. As características escolhidas nesse trabalho visando tal balanceamento são medidas do *background* socioeconômico dos alunos e de suas famílias, tais como: tamanho da família, educação dos pais, renda, infraestrutura domiciliar etc. Uma das vantagens do algoritmo é que ele não faz suposições sobre o processo gerador dos dados, com exceção de ignorabilidade, e assegura que os desequilíbrios entre os grupos após o pareamento não sejam maiores que um limite tolerável. Em suma, o método permite a construção de “pesos” (baseados na estratificação das características selecionadas) que servem para ponderar os potenciais indivíduos no grupo de controle (aqueles com percentual de professores negros abaixo de 50%) de modo deixá-los o mais parecidos possível com os tratados (alunos com exposição a professores pretos/pardos no ensino médio superior a 50%). Após a aplicação do algoritmo *CEM*, o modelo descrito em (1), sem os controles em Z_{ie} , pode ser estimado utilizando os pesos na regressão.

3.2.2 Outras características da distribuição de notas

A princípio, o efeito das características dos professores sobre o desempenho dos alunos que compartilham essas mesmas características pode não ser homogêneo ao longo da distribuição de notas, como apontado na literatura (ver Joshi, Doan e Springer, 2018). Tendo em vista essa possibilidade, além da análise acima, que se baseia no estudo das diferenças de desempenho na média, estimou-se também os parâmetros de interesse focando nos decis da distribuição de notas.

Ou seja, estimou-se também versões do modelo econométrico representado na equação 1 a partir do método de regressão quantílica combinado com o método de efeitos fixos. Vale destacar que, a implementação de tal análise é dificultada computacionalmente pela grande dimensão da base de dados bem como a elevada quantidade de efeitos fixos. Para contornar esse problema, fez-se uso da abordagem desenvolvida em Firpo, Fortin e Lemieux (2018), em que a estimação se torna bem mais viável computacionalmente, partindo-se da estimação da Função de Influência Recentralizada para os quantis de interesse e, a partir disso, estima-se o modelo linear para o modelo com a variável dependente transformada.

3. RESULTADOS

3.1 Análise descritiva

A Tabela A1, Anexo A, apresenta as estatísticas descritivas das variáveis utilizadas para implementar a estratégia empírica. Verifica-se que embora a pontuação média em LC (516,16) tenha sido inferior à de MT (521,80), a mediana de MT (499,90) foi superior à de LC (428,3), trazendo indícios de que as duas provas são assimétricas e com padrões de assimetrias diferentes. Enquanto a distribuição de notas em LC tem traços de assimetria a esquerda, a de MT é assimétrica a direita. Além disso, os alunos pretos pardos tendem a estar mais concentrados nos estratos inferiores da distribuição de notas e isso ocorre mais acentuadamente em MT (Figura 1).

Outro fato que chama atenção é alta incidência de alunos pretos/pardos com baixas pontuações. Por exemplo, dos estudantes localizados no primeiro decil da distribuição, 70% eram pretos/pardos. Essa proporção vai caindo, à medida que se avança para os decis mais altos. No topo da distribuição essa proporção cai pela metade (cerca de 35% em Linguagens e Códigos e 30% em matemática). Mesmo os pretos/pardos sendo a maioria na amostra, não conseguem estar bem representados nos decis mais altos (Figura 1).

Mais especificamente, em termos de amostra, mais de 56% dos alunos se autodeclararam pretos/pardos e 58% eram do gênero feminino. Alunos das demais raças representaram 43% da amostra, sendo compostos por brancos (41%), amarelos (2%) e indígenas (1%). Em relação a educação do pai, 48% eram filhos de pais ou que nunca estudou ou que possuía a escolaridade inferior ao ensino médio completo. Já em relação as mães, esse percentual foi de 42%. Também se observa que o tamanho médio das famílias dos estudantes era de 4 integrantes e que 50% deles declararam uma renda familiar mensal inferior a R\$ 1.497,00. A amostra dos professores é composta por 40% deles sendo autodeclarados preto/pardo (Tabela A1).

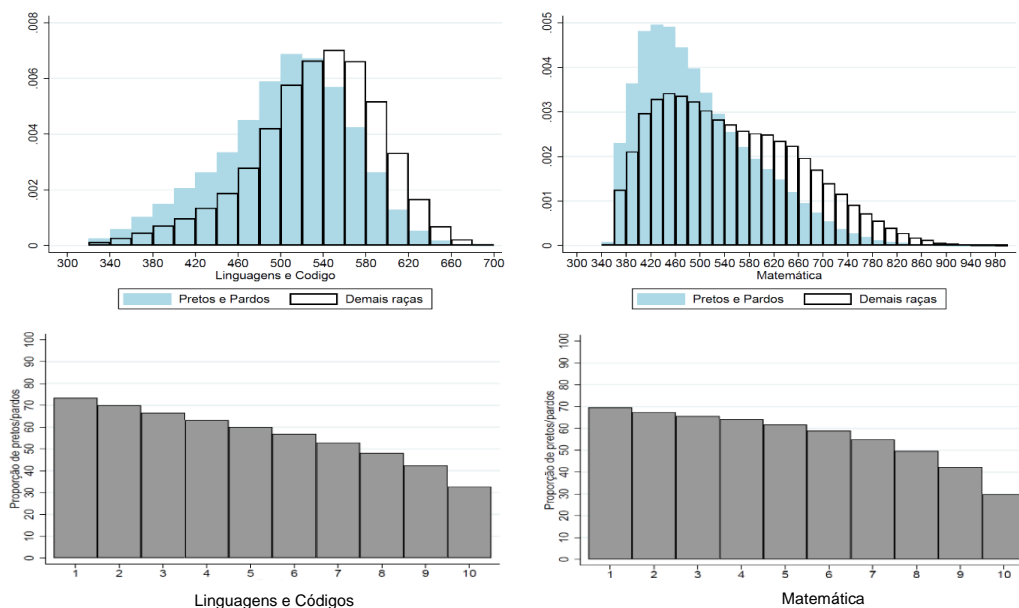


Figura 1 – Distribuição das notas de linguagens e códigos e matemática e proporção de pretos/pardos por decis. Enem 2019.

Fonte: elaboração própria dos autores com base nos microdados do Censo Escolar e do Enem 2019.

3.2 Discussão dos resultados

Os resultados das estimações dos *gaps* de notas existentes entre alunos preto/pardos e demais raças nas provas de Linguagens e Códigos e Matemática estão compilados na Tabela 2, como também os efeitos da redução desse *gap* quando os alunos pretos/pardos estão expostos a um corpo docente que também se autodeclara preto/pardo.

As colunas 1 a 3 são reportadas a fim de transparência, revelando que existe uma correlação entre as notas dos alunos e raça dos professores e alunos. Interessante frisar que não se tratam de uma relação causal, uma vez que tais características estão correlacionadas com outras características do aluno, das escolas e de *background* socioeconômico. Dessa forma, consideramos que a coluna 4 seja a que mais se aproxima de um efeito causal, uma vez que utilizamos o *CEM-matching*, com pesos relacionados ao *background* dos alunos, e os efeitos fixos de escola. Os efeitos fixos de escola ajudam a controlar pelo fato das escolas serem diferentes umas das outras.

Mesmo considerando alunos com características escolares e familiares semelhantes, os alunos pretos/pardos que são expostos apenas a professores não pretos/pardos pontuam, em média, 12,49 pontos a menos do que os das demais raças em MT (equação 4). Por outro lado, quando os alunos preto/pardo são expostos a um corpo docente do ensino médio com apenas professores pretos e pardos, esse *gap* reduz para -8,52 pontos ($-12,49+3,96$, ver coluna 4 da Tabela 2).

Analisando esses casos extremos, percebe-se que as diferenças entre negros/pardos e demais raças cai cerca de 31,74% quando estes passam a estudar e ter contato com professores negros. Dessa forma, os resultados sugerem que pode haver efeitos positivos em promover uma maior

representatividade do corpo docente sobre o *gap* de desempenho racial, pois os alunos negros se beneficiam quando estão em contato com um corpo docente que se assemelha a sua raça. No presente contexto, em que identificamos que 40% dos professores da amostra são pretos/pardos, podemos ajustar esse *gap* para essa proporção, conforme já explicitado na equação 4. Logo, o *gap* de desempenho racial passa a ser de -10,90 pontos, em média.

Em relação aos resultados para a prova de Linguagens e Códigos, observa-se um padrão semelhante ao de matemática, contudo em magnitudes menores. Por exemplo, o *gap* de desempenho é de -7,91 pontos quando os alunos pretos/pardos são expostos apenas a professores não pretos/pardos e esse *gap* cai para -6,72 pontos em um contexto de corpo docente preto/pardo, apresentando uma queda de 15,04% no *gap*. Realizando exercício análogo ao descrito no parágrafo anterior, como 40% do corpo docente era preto/pardo, temos que o *gap* racial, considerando essa proporção, é de -7,43%.

Tabela 2 – Resultado das estimações: pontuação em matemática e linguagens e códigos. Enem 2019. Brasil

Matemática				
pretpard_aluno (i)	-50.7828*** (0.3522)	-13.8493*** (0.3289)	-9.8596*** (0.3201)	-12.4908*** (0.3291)
prof_pretpard_mt (ii)	-35.8809*** (0.5166)			
same_pretpard_mt (iii)	13.1343*** (0.6658)	5.3044*** (0.6117)	3.7711*** (0.5924)	3.9646*** (0.6039)
Constante	561.1733*** (0.2329)			513.7637*** (0.1729)
Observações	723,057	722,291	722,291	722,963
R-quadrado	0.0686	0.4036	0.4410	0.3622
Gap (i)/(iii)				-31,74%
Linguagens e Códigos				
pretpard_aluno (i)	-26.4066*** (0.2072)	-7.9822*** (0.2060)	-5.3289*** (0.2040)	-7.9139*** (0.2168)
prof_pretpard_pt (ii)	-22.6258*** (0.3110)			
same_pretpard_pt (iii)	2.2524*** (0.3992)	1.2520*** (0.3889)	0.4326 (0.3833)	1.1905*** (0.4039)
Constante	539.4881*** (0.1382)			511.8570*** (0.1137)
Observações	762,945	762,234	762,234	762,802
R-quadrado	0.0684	0.3226	0.3422	0.2912
Gap (i)/(iii)				15,04%
Escola FE	Não	Sim	Sim	Sim
Controles	Não	Não	Sim	Não
CEM-Matching	Não	Não	Não	Sim

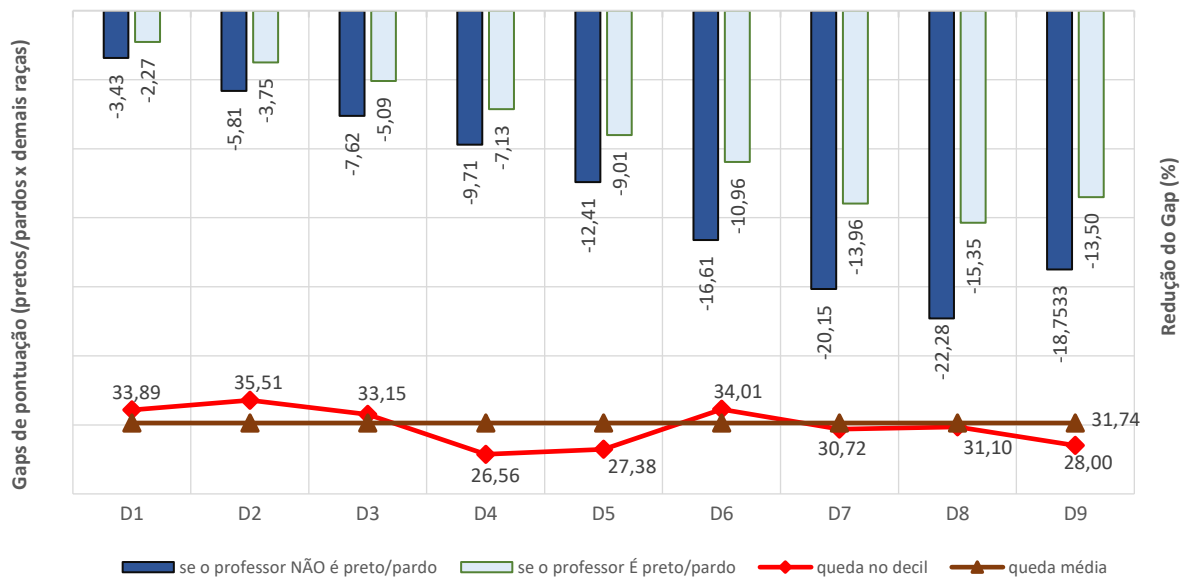
Fonte: elaboração própria dos autores com base nos microdados do Censo Escolar e do Enem 2019.

Nota: Erros padrões entre parênteses *** p<0.01, ** p<0.05, * p<0.1.

Uma vez que as distribuições das notas são assimétricas, torna-se interessante também investigar se os resultados apresentados na coluna 4 da Tabela 2 podem ser extrapolados para toda a distribuição de notas. Portanto, as Figuras 2 e 3 complementam a análise anterior ao detalhar os resultados das estimações por decis.

De acordo com a Figura 2, nota-se que as magnitudes dos *gaps* são diferentes ao longo da distribuição⁷. Os menores *gaps* foram observados nos Decis 1 e 2 enquanto os maiores estão nos decis 7 e 8. Por exemplo, no decil 8 os alunos pretos/pardos que tiveram contato com um corpo docente composto por nenhum professor preto/pardo tiveram uma pontuação inferior as demais raças em 22,28 pontos, já quando estão expostos a somente professores preto/pardo o *gap* é de 15,35. A exposição a um corpo docente mais representativo em Matemática contribui para o melhor desempenho dos estudantes pretos/pardos, consequentemente contribuindo para redução do *gap* de desempenho racial. Os resultados revelam que essa redução ocorre ao longo de toda a distribuição de notas, variando entre 26,6% e 36%.

Figura 2 - Resultado das estimações: decis de pontuação em Matemática. Enem 2019. Brasil.



Fonte: elaboração própria dos autores com base nos microdados do Censo Escolar e do Enem 2019.

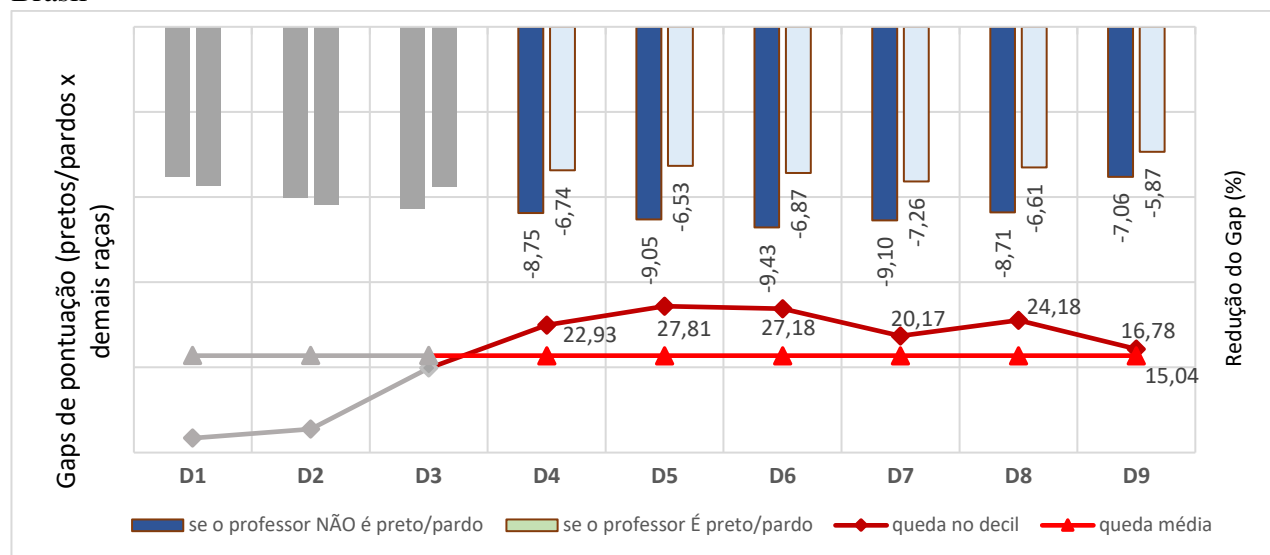
Nota: os valores das barras coloridas foram significantes ao nível de pelo menos 10%. Para mais detalhes visualizar Tabela B2, Anexo B.

Em Linguagens e Códigos, os *gaps* estimados não apresentaram grandes diferenças ao longo da distribuição de notas. No decil 6, a título de exemplo, os estudantes negros tiram nota inferior as demais raças em 9,43 pontos quando estão expostos a professores de etnias diferentes das suas, e esse *gap* cai para 6,87% quando tem contato com um corpo docente negro/pardo.

Com base na Figura 3, observa-se que a queda percentual do *gap* de desempenho racial se mostrou mais heterogênea, com a maior redução ocorrendo no decil 5 (27,81%) e o menor estando no D9 (15,04%). Vale salientar que os *gaps* estimados nos três primeiros decis não foram estatisticamente diferentes do valor nulo ao nível de 10% de significância.

⁷ Os resultados por decis de pontuação estão na Tabela B2, anexo B.

Figura 3 - Resultado das estimações: decis de pontuação. Linguagens e Códigos. Enem 2019. Brasil



Fonte: elaboração própria dos autores com base nos microdados do Censo Escolar e do Enem 2019.

Nota: os valores das barras e linhas cinzas não foram significantes ao nível de pelo menos 10%. Para mais detalhes visualizar Tabela B2, Anexo B.

Além da diferença de desempenho entre raças ter sido maior em Matemática, também foi observado que nessa disciplina a redução do *gap* foi maior. Essas evidências podem estar relacionadas ao fato de que a prova de Matemática exige do aluno um conhecimento mais específico no qual a principal forma de obtê-lo é na escola, por meio dos professores. Nesse contexto, a figura do professor preto/pardo se torna ainda mais fundamental para que os estudantes pretos/pardos sejam motivados e consigam bons resultados, diminuindo as diferenças de desempenho entre esses grupos étnicos.

4. CONSIDERAÇÕES FINAIS

Internacionalmente, diversos estudos têm analisado os efeitos da similaridade de raça entre professor e alunos sobre o desempenho dos estudantes. Contudo, essa discussão ainda permanece pouco explorada a nível nacional. Dessa forma, esse artigo buscou investigar se a exposição dos alunos pretos/pardos do 3º ano do ensino médio a um corpo docente preto/pardo tem efeitos sobre o *gap* racial de desempenho nas provas de Linguagens e Códigos e Matemática do Enem 2019. Para tanto, foram utilizados os microdados do censo escolar e do Enem.

As estimações dos *gaps* foram realizadas considerando o algoritmo CEM *matching* que constrói “pesos” que servem para ponderar os potenciais indivíduos no grupo de controle de modo a deixá-los o mais parecidos possível com os tratados (alunos com exposição a professores pretos/pardos no ensino médio superior a 50%). Os resultados mostraram que os alunos pretos e pardos obtiveram pontuações inferiores às demais raças, independente da raça do professor, por outro lado esse *gap* racial reduz cerca de 31,74% em Matemática e 15,07% em Linguagens e Códigos quando os alunos pretos/pardos saem de um contexto em que estão expostos apenas a professores brancos e passam a ter contato apenas com professores pretos/pardos. Efeitos positivos como esses são encontrados ao longo de toda a distribuição de pontuações em Matemática e, a partir do decil 4, em Linguagens e

Códigos. Embora nossa análise considere cenários extremos, é possível também criar cenários sobre o que ocorre com os *gaps* à medida que se aumenta a proporção de professores preto/pardos no corpo docente, evidenciando a importância da diversificação do corpo docente.

Nesse sentido, Gershenson, Hansen e Lindsay (2021) argumentam que a diversificação do corpo docente pode ser um importante instrumento de política social voltado para combater as disparidades raciais de desempenho. Os autores salientam que a diversidade de professores deve ser vista como um elemento da qualidade do professor, e as políticas voltadas para a melhoria da qualidade do professor devem levar a raça explicitamente em consideração. Evidências mais robustas poderiam ser obtidas a partir de trabalhos futuros que buscassem analisar a associação professor-aluno ao longo de toda a vida escolar dos estudantes pretos/pardos, ou seja, as interações que os alunos pretos/pardos têm com o professor desde o início do ensino fundamental até o ensino médio.

REFERÊNCIAS

- Banerjee, N. (2018). Effects of teacher-student ethnoracial matching and overall teacher diversity in elementary schools on educational outcomes. *Journal of Research in Childhood Education*, 32(1), 94-118.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). *How and why do teacher credentials matter for student achievement?* (No. w12828). National Bureau of Economic Research
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter?. *American Economic Review*, 95(2), 158-165.
- Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, 86(1), 195-210.
- Egalite, A. J., & Kisida, B. (2018). The effects of teacher match on students' academic perceptions and attitudes. *Educational Evaluation and Policy Analysis*, 40(1), 59-81.
- Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, 45, 44-52.
- Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *ILR Review*, 48(3), 547-561
- Firpo, S. P., Fortin, N. M., & Lemieux, T. (2018). Decomposing wage distributions using recentered influence function regressions. *Econometrics*, 6(2), 28.
- Gershenson, S., Hansen, M., & Lindsay, C. A. (2021). *Teacher Diversity and Student Success: Why Racial Representation Matters in the Classroom*. Harvard Education Press. 8 Story Street First Floor, Cambridge, MA 02138.
- Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student-teacher demographic match on teacher expectations. *Economics of education review*, 52, 209-224.

Iacus, S. M., King, G., & Porro, G. (2008). Matching for causal inference without balance checking. *Available at SSRN 1152391*.

Interdisciplinaridade e Evidências no Debate Educacional – Iede (2020). Estudo aponta desigualdade racial na educação em todas as faixas de renda. Reportagem na CNN. Acesso em 12 de julho de 2021. <https://www.portaliiede.com.br/cnn-estudo-aponta-desigualdade-racial-na-educacao-em-todas-as-faixas-de-renda/>

Joshi, E., Doan, S., & Springer, M. G. (2018). Student-teacher race congruence: New evidence and insight from Tennessee. *AERA Open*, 4(4), 2332858418817528.

Penney, J. (2017). Racial interaction effects and student achievement. *Education Finance and Policy*, 12(4), 447-467

Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly*, 16(2), 125.

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of educational psychology*, 99(2), 253

Vaz, D. V. (2020). Background familiar, retornos da educação e desigualdade racial no Brasil. *Cadernos de Pesquisa*, 50, 845-864.

Yarnell, L. M., & Bohrnstedt, G. W. (2018). Student-teacher racial match and its association with Black student achievement: An exploration using multilevel structural equation modeling. *American Educational Research Journal*, 55(2), 287-324.

ANEXO A

Tabela A.1 – Estatística descritiva das variáveis

Variáveis	LINGUAGENS E CÓDIGOS		MATEMÁTICA	
	(1)	(2)	(3)	(4)
	média	desvio padrão	média	desvio padrão
Nota_LC	516,16	64,19	521,81	106,84
fem_aluno	0,58	0,49	0,58	0,49
nao_declarado_aluno	0,00	0,00	0,00	0,00
branco_aluno	0,41	0,49	0,41	0,49
preto_aluno	0,11	0,31	0,11	0,31
pardo_aluno	0,46	0,50	0,46	0,50
amarelo_aluno	0,02	0,15	0,02	0,15
indigena_aluno	0,01	0,08	0,01	0,08
pretpard_aluno	0,57	0,50	0,56	0,50
educpai1	0,48	0,50	0,48	0,50
educpai2	0,29	0,45	0,29	0,45
educpai3	0,13	0,34	0,13	0,34
educpai4	0,10	0,30	0,10	0,29
educmae1	0,42	0,49	0,41	0,49
educmae2	0,36	0,48	0,36	0,48
educmae3	0,19	0,39	0,20	0,40
educmae4	0,04	0,19	0,04	0,19
renda1	0,29	0,45	0,29	0,45
renda2	0,25	0,43	0,24	0,43
renda3	0,09	0,28	0,09	0,28
renda4	0,10	0,30	0,10	0,30
renda5	0,11	0,32	0,12	0,32
renda6	0,09	0,28	0,09	0,28
renda7	0,08	0,27	0,08	0,27
tam_fam	4,09	1,36	4,08	1,35
dom_MAQUINA_LAVAR	0,64	0,48	0,65	0,48
dom_MICROONDAS	0,53	0,50	0,53	0,50
dom_ASPIRADOR	0,23	0,42	0,23	0,42
dom_DVD	0,34	0,47	0,34	0,47
dom_TV_ASSINATURA	0,25	0,43	0,25	0,44
dom_FONEFIXO	0,26	0,44	0,27	0,44
dom_INTERNET	0,78	0,41	0,78	0,41
dom_CARRO	0,49	0,50	0,50	0,50
dom_GELAD	0,98	0,12	0,98	0,12
dom_BANHEIRO	0,31	0,46	0,31	0,46
dom_QUARTO	0,39	0,49	0,39	0,49
dom_CELULAR	2,51	1,10	2,52	1,09
dom_COMPUTADOR	0,52	0,50	0,53	0,50
(mean) prof_pretpard	0,40	0,38	0,40	0,39
dif_pretpard	0,28	0,37	0,28	0,38
treat_raca	0,40	0,49	0,41	0,49
número de observações	762.945		723057,00	

Fonte: elaboração própria dos autores com base nos microdados do Censo Escolar e do Enem 2019.

ANEXO B

Tabela B.1 – Resultados das Estimações Normalizados pelo Desvio Padrão

	(1)	(2)	(3)	(4)
Matemática				
pretpard_aluno (i)	-0.4753*** (0.0033)	-0.1296*** (0.0031)	-0.0923*** (0.0030)	-0.1169*** (0.0031)
prof_pretpard_mt (ii)	-0.3358*** (0.0048)			
same_pretpard_mt (iii)	0.1229*** (0.0062)	0.0496*** (0.0057)	0.0353*** (0.0055)	0.0371*** (0.0057)
Constante	0.3685*** (0.0022)			-0.0753*** (0.0016)
Observações	723,057	722,291	722,291	722,963
R-quadrado	0.0686	0.4036	0.4410	0.3622
Linguagens e Códigos				
pretpard_aluno (i)	-0.4114*** (0.0032)	-0.1244*** (0.0032)	-0.0830*** (0.0032)	-0.1233*** (0.0034)
prof_pretpard_pt (ii)	-0.3525*** (0.0048)			
same_pretpard_pt (iii)	0.0351*** (0.0062)	0.0195*** (0.0061)	0.0067 (0.0060)	0.0185*** (0.0063)
Constante	0.3634*** (0.0022)			-0.0671*** (0.0018)
Observações	762,945	762,234	762,234	762,802
R-quadrado	0.0684	0.3226	0.3422	0.2912
Escola FE	Não	Sim	Sim	Sim
Controles	Não	Não	Sim	Não
CEM-Matching	Não	Não	Não	Sim

Fonte: elaboração própria dos autores com base nos microdados do Censo Escolar e do Enem 2019.

Tabela B.2 – Resultados das Estimações por decil de renda em desvios padrão

	D10	D20	D30	D40	D50	D60	Decil 70	Decil 80	Decil 90
MATEMÁTICA									
pretpard_aluno (i)	-3.4286*** (0.2945)	-5.8086*** (0.3408)	-7.6154*** (0.3719)	-9.7108*** (0.4185)	-12.4064*** (0.4724)	-16.6121*** (0.5468)	-20.1505*** (0.6321)	-22.2812*** (0.7051)	-18.7533*** (0.7582)
same_pretpard_mt	1.1620** (0.5405)	2.0627*** (0.6255)	2.5243*** (0.6825)	2.5791*** (0.7681)	3.3965*** (0.8670)	5.6497*** (1.0037)	6.1897*** (1.1601)	6.9296*** (1.2941)	5.2517*** (1.3917)
Constante	398.7653** (0.1547)	421.9748** (0.1791)	444.0776** (0.1954)	466.9036** (0.2199)	492.0856** (0.2482)	521.6086** (0.2873)	558.0086** (0.3321)	603.5384** (0.3705)	661.5745** (0.3984)
Observações	722,963	722,963	722,963	722,963	722,963	722,963	722,963	722,963	722,963
R-quadrado	0.0652	0.0921	0.1195	0.1473	0.1759	0.2083	0.2447	0.2855	0.3295
LINGUAGENS E CÓDIGOS									
pretpard_aluno (i)	-7.0333*** (0.5359)	-8.0472*** (0.4357)	-8.5401*** (0.3381)	-8.7452*** (0.2846)	-9.0483*** (0.2622)	-9.4283*** (0.2588)	-9.1002*** (0.2597)	-8.7126*** (0.2669)	-7.0554*** (0.2956)
same_pretpard_pt	-0.4408 (0.9986)	-0.3165 (0.8119)	1.0091 (0.6300)	2.0055*** (0.5304)	2.5166*** (0.4886)	2.5624*** (0.4822)	1.8356*** (0.4840)	2.1065*** (0.4974)	1.1841** (0.5508)
Constante	423.9137** (0.2810)	461.3012** (0.2285)	485.3073** (0.1773)	502.9496** (0.1492)	518.1712** (0.1375)	532.9514** (0.1357)	548.3202** (0.1362)	565.4836** (0.1400)	587.4840** (0.1550)
Observações	762,802	762,802	762,802	762,802	762,802	762,802	762,802	762,802	762,802
R-quadrado	0.0994	0.1427	0.1725	0.1937	0.2097	0.2252	0.2421	0.2593	0.2702
Escola FE: Sim CE-Matching: Sim									

Fonte: elaboração própria dos autores com base nos microdados do Censo Escolar e do Enem 2019.