



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE ECONOMIA AGRÍCOLA
CURSO DE PÓS-GRADUAÇÃO EM ECONOMIA RURAL**

LUCIANA DE OLIVEIRA RODRIGUES

**ENSAIOS SOBRE DIFERENCIAL DE DESEMPENHO ESCOLAR ENTRE
ALUNOS DE ESCOLAS RURAIS E URBANAS NO BRASIL**

FORTALEZA

2017

LUCIANA DE OLIVEIRA RODRIGUES

ENSAIOS SOBRE DIFERENCIAL DE DESEMPENHO ESCOLAR ENTRE
ALUNOS DE ESCOLAS RURAIS E URBANAS NO BRASIL

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Economia Rural do Departamento de Economia Agrícola da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Mestre. Área de concentração: Políticas Públicas e Desenvolvimento Rural.

Orientador: Prof. Dr. Edward Martins Costa

Coorientador: Prof. Dr. Vitor Hugo Miro Couto Silva.

Coorientadora: Prof. Dra. Francisca Zilania Mariano

FORTALEZA

2017

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- R614e Rodrigues, Luciana de Oliveira.
Ensaio sobre diferencial de desempenho escolar entre alunos de escolas rurais e urbanas no Brasil /
Luciana de Oliveira Rodrigues. – 2017.
109 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Programa de
Pós-Graduação em Economia Rural, Fortaleza, 2017.
Orientação: Prof. Dr. Edward Martins Costa .
1. Desempenho escolar. 2. Diferença de desempenho. 3. Rural-Urbano. I. Título.

CDD 338.1

LUCIANA DE OLIVEIRA RODRIGUES

ENSAIOS SOBRE DIFERENCIAL DE DESEMPENHO ESCOLAR ENTRE
ALUNOS DE ESCOLAS RURAIS E URBANAS NO BRASIL

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Economia Rural do Departamento de Economia Agrícola da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Mestre. Área de concentração: Políticas Públicas e Desenvolvimento Rural.

Aprovada em: ___/___/_____.

BANCA EXAMINADORA

Prof. Dr. Edward Martins Costa (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. José de Jesus Sousa Lemos
Universidade Federal do Ceará (UFC)

Prof. Dr. Vitor Hugo Miro Couto Silva
Universidade Federal do Ceará (UFC)

Dr. Miguel Nathan Foguel
Instituto de Pesquisa Econômica Aplicada (IPEA)

Ao meu soberano Deus.

A minha família, especialmente a minha
mãe.

AGRADECIMENTOS

Em primeiro lugar, agradeço ao meu Deus Altíssimo, pela inteligência, saúde, determinação e principalmente pela capacidade de lutar pelos meus objetivos e sonhos.

A toda a minha família, que sempre foi minha base de sustentação, em especial minha mãe, Marli, pelos cuidados em que teve durante todo este período que não tive tempo para quase nada. Obrigada pelo apoio, amor e paciência. Aos meus irmãos, Ana Maria, Alberto, Cristina e Paulo, pela torcida. Aos meus lindos sobrinhos: Murilo, Davi, Enzo, Pedro, Adrielly e Lara, por tornarem minha vida mais leve e feliz.

Agradeço muitíssimo ao professor Dr. Edward Martins Costa que foi mais que um orientador, pois foi também um amigo e um mestre sem igual. Obrigada por todas as dicas, broncas e, principalmente, pela confiança: elas foram e sempre serão importantes para a minha formação acadêmica e profissional.

Ao professor Dr. Vitor Miro, não apenas por aceitar participar da banca, mas também por ter contribuído com seus conhecimentos para a elaboração deste estudo. Ao professor Dr. J.J.S. Lemos por aceitar ser membro da banca e pelo conhecimento transmitido em várias etapas da minha carreira acadêmica, desde a graduação em Economia na FEAAC. E o professor Dr. Miguel Foguel por ter se disponibilizado em vir de tão longe para compartilhar seus conhecimentos e sua sabedoria.

Agradeço, também, à Prof^a Dr. Zilania Mariano, pelas contribuições e conhecimento transmitido durante a elaboração deste trabalho, que foram de enorme importância.

E a todos aqueles que contribuíram para a minha formação, especialmente os professores do Programa de Mestrado Acadêmico em Economia Rural – MAER da Universidade Federal do Ceará.

Agradeço a todos os meus colegas de curso que compartilharam, durante estes últimos dois anos momentos de felicidades, tristezas e angústias. Com vocês essa caminhada se tornou mais leve. Agradeço, especialmente, àqueles mais achegados: Camila, Cleidiane, Felipe, Gerrio, Patrícia, Poliana, Roberto e Wesley.

Ao meu amigo João Paulo, pelo apoio, pelos conselhos e a paciência nos momentos mais tensos. À Micheliana, que sempre esteve perto mesmo estando distante.

Obrigada pelo apoio de sempre e pelos anos de amizade. E a minha ex-colega do Instituto de Pesquisa Econômica do Ceará (IPECE) Raquel Sales, pela torcida e pelo incentivo para entrar no mestrado. E ao Leandro Costa, a quem devo muito. Se hoje tenho uma segurança profissional e conhecimento na área de Economia da Educação foi pelas oportunidades que ele me deu. Muito obrigada por todo o conhecimento compartilhado e a confiança de sempre.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro concedido durante todo o curso, que foi primordial para a dedicação exclusiva e conclusão deste trabalho.

À Universidade Federal do Ceará, por me proporcionar o título de Mestre.

Enfim, a todas as pessoas que contribuíram de algum modo com este momento tão importante na minha vida. Mais uma vez, obrigada a todos!

RESUMO

Essa dissertação é composta de dois artigos, o primeiro dos quais com o título “Fatores que determinam o diferencial de desempenho escolar entre estudantes das zonas rural e urbana no Brasil”, analisou os fatores que contribuem com o diferencial de rendimento de alunos que frequentavam escolas no meio rural e urbano. Para tal, foram utilizadas informações dos estudantes avaliados no 5º ano do Ensino Fundamental na Prova Brasil de 2015, nas disciplinas de Português e Matemática. O diferencial de rendimento escolar rural-urbano foi estimado pelo método de decomposição com base em regressões de função influência recentrada (FIR), proposto por Firpo, Fortin e Lemieux (2007), combinado com o procedimento de reponderação inspirada por Dinardo *et al.* (1996), que permitiu aplicar o método de decomposição de Oaxaca-Blinder (1973) tradicional nos *quantis* de distribuição. Os resultados apontaram para uma diferença entre alunos de escolas rurais e urbanas, mostrando uma vantagem para os estudantes de escolas urbanas, onde grande parte do diferencial entre os grupos se deve ao efeito características das escolas e da família do aluno, sendo crescente ao longo dos *quantis*. O segundo artigo recebeu o título de “Migração Educacional Rural-Urbana: análise do diferencial de desempenho escolar no Ensino Médio” e analisou como o desempenho de estudantes da 3ª série do Ensino Médio que residiam no meio rural era influenciado pela escola onde estudavam, seja na própria zona rural ou quando eram transferidos para escolas urbanas. Foi utilizada a combinação de duas metodologias: a primeira consistiu em um método de pareamento por meio das características observáveis entre alunos de escolas distintas, de modo que a única diferença entre um e o outro fosse somente o “tratamento” (a zona de residência ou localização da escola), o algoritmo *Coarsened Exact Matching* (CEM). E, para a decomposição, aplicou-se o método proposto por Oaxaca (1973) e Blinder (1973), que identificou as variáveis que mais contribuíam para explicar as diferenças de resultados educacionais entre os dois grupos. Dos resultados encontrados, observou-se que, mesmo controlando as características individuais e familiares do aluno, o estudante da zona rural que estudava em escola urbana apresentava maior desempenho do que seus pares que frequentavam escolas na própria zona rural. Tal ocorrência indica que é melhor para os alunos do meio rural estudar em escolas urbanas, pois essas denotavam melhor infraestrutura e professores mais qualificados. Quanto à decomposição dos diferenciais de alunos que moravam no meio rural com relação aos que viviam no meio urbano, estudando em escolas similares, observou-se que, mesmo estudando em escolas similares eles ainda registravam desvantagens em comparação aos seus colegas por conta de características individuais e familiares.

Palavras-chave: Desempenho escolar. Diferença de desempenho. Rural-Urbano.

ABSTRACT

This dissertation is composed of two articles, the first of which is entitled "Factors that determine the differential of school achievement among students from rural and urban areas in Brazil", analyzed the factors that contribute to the differential of income of students attending schools in rural and urban areas. For that, we used information from the students evaluated in the 5th year of Elementary School in the Brazil Test of 2015, in the Portuguese and Mathematics subjects. The rural-urban school achievement differential was estimated by the method of decomposition based on regressions of refocused influence function (FIR), proposed by Firpo, Fortin and Lemieux (2007), combined with the re-weighting procedure inspired by Dinardo *et al.* (1996), which allowed the application of the traditional Oaxaca-Blinder decomposition method (1973) in distribution quantiles. The results pointed to a difference between students from rural and urban schools, showing an advantage for urban school students, where a large part of the difference between groups is due to the characteristic effect of the schools and the student's family, increasing over the years. Quantiles The second article was entitled "Rural-Urban Educational Migration: Analysis of the School Achievement Differences in Secondary School" and analyzed how the achievement of high school students who lived in rural areas was influenced by the school where they studied, whether in the countryside itself or when they were transferred to urban schools. A combination of two methodologies was used: the first consisted of a pairing method through observable characteristics among students from different schools, so that the only difference between one and the other was only the "treatment" (the zone of residence or Location of the school), the Coarsened Exact Matching (CEM) algorithm. And, for the decomposition, the method proposed by Oaxaca (1973) and Blinder (1973) was applied, which identified the variables that contributed the most to explain the differences in educational outcomes between the two groups. From the results, it was observed that, even controlling the individual and family characteristics of the student, the rural student who studied in an urban school presented a higher achievement than his peers who attended schools in the rural area. Such an occurrence indicates that it is better for rural students to study in urban schools, since these denote better infrastructure and more qualified teachers. As for the breakdown of differentials of students living in rural areas compared to those living in the urban environment, studying in similar schools, it was observed that even studying in similar schools they still had disadvantages in comparison to their colleagues due to individual and family characteristics.

Keywords: School achievement. Student achievement gaps. Rural-Urban.

LISTA DE FIGURAS

Figura 1 - Densidades estimadas (ln da Proficiência em Português e Matemática).....	39
Figura 2 - Decomposição do diferencial de desempenho escolar (ln da proficiência em Português e Matemática).	46
Figura 3 - Efeito-Característica detalhado - Urbano-Rural - Português e Matemática. .	48
Figura 4 - Efeito Estrutural detalhado - Urbano-Rural - Português e Matemática.....	49

LISTA DE TABELAS

Tabela 1 - Descrição das variáveis incluídas no modelo	36
Tabela 2 - Média e desvio-padrão das variáveis utilizadas nas regressões – meio rural e urbano – Brasil, 2015.....	41
Tabela 3 - Fluxo escolar dos alunos matriculados no Ensino Fundamental, em 2011 ...	59
Tabela 4 - Informações dos alunos ingressos no ensino médio (EM), em 2012	60
Tabela 5 - Fluxo escolar dos alunos matriculados na 1ª série do EM, em 2012.....	60
Tabela 6 - Fluxo escolar dos alunos matriculados na 1ª e 2ª série EM, em 2013.....	61
Tabela 7 - Fluxo escolar dos alunos matriculados na 1ª, 2ª e 3ª série do EM, em 2014.	62
Tabela 8 - Descrição das variáveis	64
Tabela 9 - Estatística descritiva Média e Desvio-Padrão	74
Tabela 10 - Resultados do algoritmo CEM para o Grupo 1	78
Tabela 11 - Decomposição Oaxaca-Blinder do diferencial de rendimento escolar Grupo 1 para o Brasil - Sem ponderação	79
Tabela 12 - Decomposição Oaxaca-Blinder do diferencial de rendimento escolar Grupo 1 para o Brasil - Com ponderação	80
Tabela 13 - Resultados do algoritmo CEM para o Grupo 2	82
Tabela 14 - Decomposição Oaxaca-Blinder do diferencial de rendimento escolar Grupo 2 para o Brasil - Sem ponderação	83
Tabela 15 - Decomposição Oaxaca-Blinder do diferencial de rendimento escolar Grupo 2 para o Brasil - Com ponderação	85

LISTA DE QUADROS

Quadro 1 - Descrição dos itens utilizados para a elaboração do status econômico familiar.....	37
Quadro 2 - Mapa dos dados coletados no Censo Escolar de 2011 a 2015.	58
Quadro 3 - Descrição dos grupos de tratamento e controle.....	71

SUMÁRIO

INTRODUÇÃO GERAL	15
CAPÍTULO 1 - FATORES QUE DETERMINAM O DIFERENCIAL DE DESEMPENHO ESCOLAR ENTRE ESTUDANTES DAS ZONAS RURAL E URBANA NO BRASIL	17
1 INTRODUÇÃO	17
2 REFERENCIAL TEÓRICO.....	21
2.1 Fatores associados ao desempenho escolar	21
2.2 Diferenciais de rendimento entre escolas rurais e urbanas.....	24
3 ESTRATÉGIA EMPÍRICA.....	26
3.1 Funções de produção educacional e decomposição de Oaxaca-Blinder (1973)	26
3.2 Decomposição por Regressão Quantílica Incondicional (RQI).....	28
3.3 Bases de dados e tratamento das variáveis.....	33
4 RESULTADOS.....	39
4.1 Análises descritivas.....	39
4.2 Regressões FIR nos <i>quantis</i> incondicionais	42
4.3 Decomposições dos diferenciais de rendimento escolar	45
5 CONSIDERAÇÕES FINAIS	50
CAPÍTULO 2 - MIGRAÇÃO EDUCACIONAL RURAL-URBANA: ANÁLISE DO DIFERENCIAL DE DESEMPENHO ESCOLAR NO ENSINO MÉDIO	52
1 INTRODUÇÃO	52
2 TRAJETÓRIA DOS ESTUDANTES DA ZONA RURAL DIPLOMADOS NO ENSINO FUNDAMENTAL	57
3 BASE DE DADOS E TRATAMENTO DAS VARIÁVEIS	63
4 METODOLOGIA E SELEÇÃO DO GRUPO DE TRATAMENTO E CONTROLE.....	67
4.1 Algoritmo CEM.....	67
4.2 Funções de produção educacional e decomposição de Oaxaca-Blinder (1973)	69
4.3 Estratégia e seleção do grupo de tratamento e controle	70
5 RESULTADOS.....	73
5.1 Análises descritivas	73
5.2 Diferencial de desempenho escolar entre alunos do Rural-Rural e Rural-Urbano (Grupo 1).....	77
5.1.1 Pareamento pelo CEM.....	78

5.2.2 Resultados da Decomposição de Oaxaca e Blinder (1973) sem ponderação.....	78
5.2.3 Resultados da Decomposição de Oaxaca e Blinder (1973) com ponderação	80
5.3 Diferencial de desempenho escolar entre alunos do Rural-Urbano e Urbano-Urbano (Grupo 2).....	81
5.3.1 Pareamento pelo CEM	81
5.3.2 Resultados da Decomposição de Oaxaca e Blinder (1973) sem ponderação	82
5.3.3 Resultados da Decomposição de Oaxaca e Blinder (1973) com ponderação	83
6 CONSIDERAÇÕES FINAIS	86
CONCLUSÃO GERAL	88
REFERÊNCIAS	90
APÊNDICE 1	96
APÊNDICE 2	105

INTRODUÇÃO GERAL

A igualdade de acesso à Educação é um dos direitos humanos básicos e que deve beneficiar a todos. Neste sentido, a redução da pobreza e o acesso da população menos favorecida a serviços básicos, como à educação gratuita, está diretamente relacionada à melhoria da qualidade de vida das pessoas em diferentes contextos sociais e econômicos. Diante disso, nos últimos anos, o Brasil conseguiu alocar grande parte das crianças em idade escolar na escola. No entanto, a qualidade do ensino ofertado a essas crianças ainda é um dos principais entraves para o desenvolvimento do País, principalmente para aquelas que vivem na zona rural. A infraestrutura das escolas, a qualificação e a instabilidade dos docentes são fatores que influenciam o desempenho das crianças que estudam em escolas rurais, resultando em desempenho inferior aos daquelas frequentadoras de escolas urbanas.

Outro fator que dificulta o acesso de adolescentes e jovens do meio rural à Educação é a falta de escolas que ofertam o Ensino Médio em localidades próximas as suas residências, de modo que muitos são obrigados a migrar para escolas do meio urbano ou são impossibilitados de dar continuidade aos estudos. Nesse contexto, esta dissertação está estruturada em dois capítulos, que abordam assuntos distintos, relacionados ao desempenho escolar de alunos dos meios rural e urbano.

No primeiro capítulo, realizou-se uma análise sobre as diferenças no desempenho escolar dos alunos do 5º ano do Ensino Fundamental, na Prova Brasil de 2015, que frequentavam escolas nos meios rural e urbano, tendo-se verificado quais são as características que mais contribuem com o diferencial entre os alunos (características do aluno, da família, dos professores e da escola). Foi utilizada a metodologia generalizada, de Oaxaca-Blinder (1973), proposta por Firpo, Fortin e Lemieux (2007), que decompõe o diferencial de rendimento escolar dos alunos além da média, ou seja, nos *quantis* de distribuição das notas.

No segundo capítulo, foi mensurado o diferencial de rendimento escolar entre alunos matriculados na 3ª série do Ensino Médio pelo Exame Nacional da Educação Básica (Enem) de 2014 para encontrar o **efeito-escola** e o **efeito-família** sobre os alunos residentes no meio rural, mas estudando em escolas urbanas. Para isso, foram empregados dois métodos: o primeiro consiste em um pareamento por meio das

características observáveis entre alunos de zonas distintas, conhecido na literatura por *Coarsened Exact Matching* (CEM). E o segundo ocorreu pelo método de decomposição proposto por Oaxaca (1973) e Blinder (1973), utilizando o peso gerado pelo método CEM para identificar quais as variáveis que mais contribuem para explicar as diferenças de resultados educacionais entre dois grupos, controlando as características observáveis dos alunos, da família e da escola, a fim de garantir um equilíbrio ou homogeneidade entre os dois grupos.

CAPÍTULO 1

FATORES QUE DETERMINAM O DIFERENCIAL DE DESEMPENHO ESCOLAR ENTRE ESTUDANTES DAS ZONAS RURAL E URBANA NO BRASIL

1 INTRODUÇÃO

Nos últimos anos, principalmente, desde o início dos anos de 1990, a quantidade de crianças que ingressaram no sistema educacional cresceu consideravelmente no Brasil, refletindo uma melhoria do acesso à Educação Básica da população, particularmente por parte do estrato menos favorecido em todas regiões do País. Se em 1992, 86,6% das crianças de sete a quatorze anos estavam frequentando escola ou creche. Em 2015, a universalização do Ensino Fundamental no Brasil foi praticamente atingida, quando se tem cerca de 98,8% das crianças e adolescentes nessa faixa de idade frequentando a escola (IPEA, 2010; IBGE, 2016).

Apesar dos esforços em garantir o acesso de toda a população em idade escolar ao ensino, a qualidade da Educação Básica no Brasil ainda está abaixo dos padrões internacionais. De acordo com dados divulgados pelo Programa Internacional de Avaliação de Estudantes – PISA (Siglas em inglês), em 2015, os estudantes brasileiros tiveram desempenho médio de 401 pontos na avaliação de Ciências. Este resultado é relativamente inferior à média alcançada pelos países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que foi de 493 pontos (OCDE, 2016; INEP 2016).

De igual modo, na avaliação de conhecimentos relacionados a Leitura, os estudantes brasileiros alcançaram a nota média de 407 pontos, sendo também significativamente inferior à nota média dos países da OCDE (493). E seu pior desempenho foi em Matemática, quando a nota média dos alunos avaliados foi de apenas 377 pontos, ficando entre os seis países que atingiram as menores notas no PISA em 2015 (OCDE, 2016).

A baixa qualidade da educação brasileira faz-se ainda mais evidente quando é analisada nos distintos contextos territoriais. Ainda considerando os dados do PISA (2015), o Espírito Santo foi a Unidade da Federação (UF) onde os estudantes de 15 anos de idade avaliados em Ciências (435) e Leitura (441) tiveram o melhor desempenho

acadêmico. Em Matemática, o Paraná registrou o melhor desempenho médio entre os estados brasileiros. Já Alagoas, aquele com pior desempenho em todas as avaliações (Ciências – 360; Leitura – 362 e Matemática – 339).

Além disso, os indicadores educacionais das escolas públicas também revelaram uma realidade bem diferente em relação ao contexto socioeconômico e zona de localidade das escolas brasileiras. De acordo com dados divulgados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), de 2015, a taxa média de aprovação do Ensino Fundamental (EF) das regiões Sudeste e Centro-Oeste foi de 92,3% e 90,9%, respectivamente. Enquanto que nas regiões Norte e Nordeste, essa mesma taxa foi de 85,4% e 84,1%, respectivamente.

Em relação aos indicadores de rendimento escolar entre alunos que frequentam escolas nas zonas rural e urbana, sobretudo, a taxa de reprovação e de abandono, existe uma variação considerável entre os ciclos, o que reflete sobre a distorção idade-série¹. A taxa de distorção idade-série² nos anos iniciais do EF, em 2015, era de 13,4% entre os alunos que frequentavam escolas urbanas, enquanto nas áreas rurais, a taxa de crianças que estavam fora da faixa de idade adequada ao ano escolar era superior a 21%. Este atraso escolar reflete nos demais anos, fazendo com que os alunos cheguem aos anos finais do EF e Ensino Médio (EM) com uma defasagem ainda maior, de 28,8% e 29,9% na zona urbana, respectivamente, e ainda mais elevada entre as crianças e adolescentes que estudam no meio rural, 41,7% nos anos finais do EF, e no EM a inadequação idade-série atinge 40,3% dos alunos.

As desigualdades entre as escolas do campo e das cidades também são significativas, pois, quando se analisa o desempenho escolar, os alunos pertencentes a escolas urbanas apresentam resultado superior àqueles de escolas rurais. Em 2015, na Prova Brasil, no País, considerando somente as escolas da rede pública de ensino e a avaliação do 5º ano, o desempenho médio dos estudantes de escolas urbanas foi de 207 pontos em Português e de 219 em Matemática. Enquanto isso nas escolas rurais, o rendimento médio foi de 182 e 196 pontos em Português e Matemática, respectivamente.

¹ Maiores informações podem ser consultadas no portal de Indicadores Educacionais do INEP no endereço: < <http://portal.inep.gov.br/indicadores-educacionais>>.

² Diferença entre idade do aluno e o ano ou série que ele frequenta.

Portanto, o baixo desempenho e o déficit de aprendizagem das crianças e adolescentes que frequentam o sistema de ensino brasileiro, principalmente das escolas públicas, são preocupantes e têm implicações potencialmente graves, não apenas para o desenvolvimento intelectual dessas pessoas, mas também sobre os níveis de equidade e desenvolvimento do Brasil. Embora a maioria dos estudos mostre que muitas das escolas não estão proporcionando um ambiente de aprendizagem adequado, a articulação de políticas públicas eficazes para esse problema continua sendo um dos principais desafios do sistema educacional da atualidade no País (RODRIGUES, 2009).

De acordo com resultados preliminares do IBGE, em 2015, a taxa de analfabetismo entre a população rural de 15 anos ou mais no Brasil era de 19,8%, enquanto nas áreas urbanas era de 5,9%. Pelo nível de instrução mais alto alcançado da população adulta (pessoas com 25 anos ou mais), na zona rural do Brasil, 56,5% não haviam nem ao menos completado o Ensino Fundamental, escolaridade mínima obrigatória no Brasil; e apenas 13,5% tinham concluído o Ensino Médio. Em contrapartida, na zona urbana, 38,6% tinham Ensino Médio completo e 29,8% não haviam completado o Ensino Fundamental.

De fato, as áreas urbanas estão mais propensas a se diferenciar das zonas rurais por apresentarem maiores benefícios em relação ao investimento em educação. Isso ocorre porque os residentes no meio rural tendem a enfrentar desvantagens geográficas e institucionais, que influenciam relativamente sobre suas decisões em investir em capital humano. Ou seja, nas áreas rurais, a eficiência de empregos e a prestação de serviços sociais são muitas vezes ineficazes, em decorrência da baixa densidade populacional e da distância geográfica. Os residentes urbanos são tipicamente mais bem posicionados para alcançar as oportunidades de emprego e serviços sociais e tendem a ter maior mobilidade geográfica e profissional (DONAHUE, 2002 *apud* AMINI; NIVOROZHKIN, 2015).

Nesse sentido, como grande parte dos estudos sobre desempenho escolar não buscam analisar as diferenças relacionadas à localidade da escola, este ensaio pretende contribuir para a literatura, considerando as diferenças de rendimento escolar dos alunos que frequentam escolas nos meios rural e urbano. As principais perguntas que se busca responder são as seguintes: o que explica as diferenças dos resultados de aprendizagem das crianças no campo e nas cidades? Seriam as características

individuais ou o tipo de infraestrutura das escolas? Ou o acesso aos recursos didáticos e a professores qualificados? Ou as características socioeconômicas das famílias é o fator mais relevante para explicar a diferença entre os dois grupos?

Portanto, para encontrar essas respostas e as possíveis diferenças nos resultados educacionais para alunos nos meios rural e urbano, utilizam-se informações dos alunos avaliados no 5º ano do EF na Prova Brasil de 2015, de todo o Território brasileiro, nas disciplinas de Matemática e Português. A principal justificativa para a escolha deste grupo decorre da formação da capacidade cognitiva das crianças, que são desenvolvidas durante os primeiros anos de vida, isto é, nos primeiros anos escolares.

A decomposição do diferencial de desempenho escolar rural-urbano é estimada por uma função de produção educacional que inclui variáveis explicativas relacionadas às características dos alunos, de suas famílias, dos professores e da escola que frequentam. Por meio da metodologia generalizada de Oaxaca-Blinder (1973), proposta por Firpo, Fortin e Lemieux (2007), que decompõe o diferencial de rendimento escolar dos alunos além da média, ou seja, nos *quantis* de distribuição das notas.

A metodologia de Firpo, Fortin e Lemieux (2007) é baseada no conceito de Função de Influência Recentrada (FIR) combinada com o procedimento de reponderação desenvolvido por Dinardo *et al.* (1996). Assim como os coeficientes de regressão, é utilizado no caso da média, o modelo de regressão proposto por Firpo, Fortin e Lemieux (2007) pode ser usado para inferir o impacto de várias variáveis explicativas sobre o *quantil* incondicional.

Nesse sentido, o trabalho é constituído por mais quatro seções, além dessa introdução. A segunda seção dedica-se ao levantamento da literatura sobre o tema. Na terceira, demonstra-se a estratégia empírica da pesquisa e as informações sobre a base de dados e tratamento das variáveis. Na parte seguinte, estão delineados resultados dos modelos estimados e, por fim, são inclusas as considerações finais do estudo.

2 REFERENCIAL TEÓRICO

Nesta etapa do estudo, realizou-se o levantamento bibliográfico a respeito do tema, sobre fatores relacionados ao desempenho escolar e a diferença das condições socioeconômicas dos alunos que frequentam escolas rurais e urbanas.

2.1 Fatores associados ao desempenho escolar

De acordo com estudos à disposição na literatura específica, são inúmeras causas que podem influenciar o desempenho escolar dos alunos, podendo ser classificados em três grupos: características individuais, o *background* familiar e as características da escola (como o contexto socioeconômico e as práticas e insumo pedagógico) (NIETO; RAMOS, 2014).

O Relatório Coleman (COLEMAN *et al.*, 1966) foi um dos primeiros estudos a mostrar que os fatores podem influenciar no desempenho do aluno. Dos resultados encontrados, os elementos escolares exerciam menor influência sobre o desempenho do estudante do que os aspectos relacionados às características familiares. O estudo encontrou evidências de que estudantes cujos pais tinham alta escolaridade obtinham melhores resultados do que aqueles com pais portadores de educação baixa.

Embora os fatores escolares não tenham demonstrado uma parcela explicativa relativamente alta sobre o desempenho alcançado pelos alunos no estudo desenvolvido por Coleman *et al.* (1966), eles são amplamente examinados pela literatura, especialmente, por denotarem uma importância significativa do ponto de vista de implementação de políticas públicas, sendo suficientemente relevantes para alterar a trajetória escolar dos alunos (WOBMANN, 2003).

Rivkin, Hanushek e Kain (2005), ao analisarem os fatores determinantes do desempenho escolar dos estudantes do Texas, chegaram à conclusão de que, embora as variáveis relacionadas à escola e aos professores tivessem baixa influência sobre os resultados dos alunos, elas são importantes para aumentar o rendimento de estudantes de baixa renda; e que bons professores poderiam compensar substancialmente a defasagem das crianças com *background* socioeconômico desfavorável.

Ao mensurar os fatores determinantes do desempenho escolar dos estudantes em alguns países, Nieto e Ramos (2014) encontraram evidências de que a qualidade da escola e do professor e as melhores práticas de ensino são importantes,

mesmo em diferentes contextos institucionais. Os recursos educacionais disponíveis contribuem para melhorar o desempenho escolar dos estudantes, como também ajudam a reduzir as diferenças entre alunos mais e menos favorecidos em termos de características socioeconômicas.

No âmbito nacional, nas últimas décadas, o INEP reuniu importantes informações sobre o contexto educacional nas diferentes etapas de ensino do País (Educação Básica e Superior). Com suportes nesses dados, diversas pesquisas são desenvolvidas com o objetivo de analisar a qualidade e a evolução dos indicadores educacionais em todo o Território Brasileiro. Com a criação do Sistema de Avaliação da Educação Básica (Saeb) em 1990, muitos estudos já foram desenvolvidos com os mais variados métodos de análise para compreender os fatores determinantes do desempenho escolar dos estudantes no final de cada ciclo de ensino³, ou seja, no 5º ano (ou antiga 4ª série) e 9º ano (ou antiga 8ª série) do EF e 3ª série do EM.

Portanto, de acordo com a análise de Soares (2006), o fato de a criança ser de origem pobre dificulta o processo de aprendizagem, mas não justifica o fato da criança não aprender a ler e a escrever, visto que, em outros países, crianças com características similares às brasileiras, conseguem ter êxito no processo de aprendizagem. Desse modo, às defasagens relacionadas as características familiares e individuais podem ser compensadas (ou minimizadas) mediante intervenções de políticas de incentivos e pela escola que frequentam (tempo durante o qual permanece na escola e a qualidade do ensino ofertado).

Neste sentido, como a escola pode influenciar no desempenho do aluno, então, alunos com características individuais e familiares semelhantes podem atingir níveis diferentes de aproveitamento simplesmente por estudarem em escolas distintas (RODRIGUES, 2009). Isto é, alunos de escolas com melhores infraestrutura e professores com escolaridade elevada podem demonstrar uma *performance* melhor do que aqueles que estudam em escolas que não possuem tais características.

Após o controle do nível socioeconômico dos alunos, Barbosa e Fernandes (2001) encontraram evidências de que as variáveis relacionadas à *infraestrutura e equipamentos escolares* têm forte impacto sobre a proficiência dos alunos, explicando

³ O ciclo de ensino no Brasil é classificado em três: Anos iniciais do ensino fundamental (1º ao 5º ano), Anos finais do Ensino Fundamental (6º ao 9º ano) e Ensino Médio.

em 54% a variabilidade da proficiência entre escolas. E, controlando os fatores relacionados a infraestrutura e equipamentos escolares para encontrar a influência das características dos professores nos resultados obtidos pelos alunos, verificaram que as características associadas aos docentes e à sua interação com a turma apareceram com grande poder explicativo no desempenho escolar. Além disso, *a motivação* do aluno e a *relação casa-escola* também são fatores explicativos da proficiência do estudante. De acordo com os resultados do estudo, aqueles que “gostam de estudar” obtêm melhores resultados, e os estudantes que mais conversam sobre a escola em casa, em média, são melhores alunos.

Albernaz, Ferreira e Franco (2002), após controlarem as características socioeconômicas dos alunos, constataram que o nível de escolaridade do professor influencia positivamente no desempenho médio nas escolas, isto é, quanto maior for a escolaridade dos professores, melhores serão as notas dos alunos. Os autores também encontraram evidências estatisticamente significantes de que salas de aulas bem arejadas e sem ruídos externos também afetam positivamente nos resultados alcançados pelos alunos.

Resultados similares foram alcançados por Biondi e Felício (2007). Para as autoras, características como a formação superior dos professores, a experiência em sala de aula por mais de dois anos e a utilização de computadores como meios pedagógicos são positivamente significantes no desempenho dos alunos. Em relação à infraestrutura da escola, ressaltam que estudantes de escolas que possuíam laboratórios de ciência e/ou informática apresentavam desempenho melhor do que aqueles que estudavam em escolas que não tinham tais recursos.

Ao analisar os determinantes do desempenho escolar dos estudantes do Rio de Janeiro, Palermo, Silva e Novellino (2014) verificaram que os elementos relacionados ao *background* do aluno tiveram maior influxo na nota da prova, sejam aqueles relativos a características intrínsecas ou as relacionadas às condições familiares, embora, no entanto, os fatores relacionados à escola tenham denotado poder limitado em relação ao rendimento escolar do aluno. As autoras ressaltam que as características da escola mostram artifícios característicos que influenciam nos resultados alcançados pelos estudantes, direta ou indiretamente.

2.2 Diferenciais de rendimento entre escolas rurais e urbanas

A literatura sobre a eficácia escolar e os fatores que contribuem para o aprendizado dos alunos é extensa, entretanto, são poucos os estudos que investigam a qualidade da Educação levando em consideração a localização da escola e na residência do aluno, isto é, entre o meio rural e urbano. Isto deixa à margem características geográficas e estruturais de cada zona, ignorando, por sua vez, os problemas relacionados ao baixo desempenho dos alunos que fazem parte exclusivamente do contexto social onde vivem.

Ao analisarem o diferencial de rendimento entre as escolas rurais e urbanas na Rússia, Amini e Nivorozhkin (2015) constataram que o desempenho dos alunos variava substancialmente segundo a localidade das escolas em todas as disciplinas avaliadas (Leitura, Matemática e Ciências), sendo significativamente maior entre os estudantes das áreas urbanas. Além disso, os autores verificaram que existem diferenças expressivas nos *backgrounds* dos alunos e nas características das escolas. As escolas urbanas apresentam uma proporção maior de pais de alunos com ensino superior completo e os estudantes têm mais livros disponíveis em casa, ou seja, as famílias dos alunos de escolas urbanas tendem a ter um *status* socioeconômico maior, se comparados com as famílias de alunos que vivem no meio rural.

Em relação às características das escolas e de seus insumos, segundo o estudo supracitado, a proporção de professores certificados e o tamanho da escola são significativamente maiores nas escolas urbanas, enquanto a proporção de computadores e a razão de professores e alunos são menores na zona rural. Este fato pode refletir não só uma disparidade na maneira como os recursos são distribuídos, mas também um número menor de alunos em escolas rurais. Com efeito, pela análise de decomposição, o estudo revelou que as características individuais e familiares dos estudantes foram as que mais contribuíram para o hiato educacional entre áreas urbanas e rurais.

Lounkaew (2013) analisando o diferencial de rendimento dos alunos da Tailândia, verificou que a influência das características individuais, da família e da escola variam ao longo da distribuição das notas alcançadas pelos estudantes. Notou, ainda, que grande parte das diferenças entre estudantes do meio urbano e rural decorre das características não mensuráveis das escolas, que apresentam variações ao longo do percentil de desempenho dos alunos, representando cerca de 12 a 15 por cento entre

estudantes de baixo desempenho e aumentando para patamares de 61 a 69 por cento entre escolares de alto desempenho.

Soares-Neto *et al.* (2013), ao analisarem a infraestrutura das escolas no Brasil, que eles classificaram em quatro categorias - Elementar, Básica, Adequada e Avançada⁴ - identificaram o fato de que as escolas do meio rural oferecem infraestrutura muito precária, em comparação com as urbanas. Segundo o estudo, enquanto mais de 85% das escolas da zona urbana se estavam na categoria Elementar, apenas 18% das escolas rurais estavam nesta categoria.

Como informam Soares, Razo e Fariñas (2006), as crianças que vivem em áreas rurais no Brasil, além de terem origem socioeconômica familiar que as situam em desvantagem em relação às crianças da cidade, elas também estudam em escolas que são mal equipadas e com professores com baixa qualificação. O que explica uma parte importante das diferenças de desempenho entre as escolas rurais e urbanas.

Visando a analisar a qualidade do ensino no Ceará, Lavor e Arraes (2014) chamam atenção para as grandes diferenças observadas entre as escolas localizadas nos meios rural e urbano, sobretudo no que se refere à disponibilidade dos diversos recursos escolares, como acesso à internet e a biblioteca. Para os autores, os alunos no meio rural, além de registrarem maior incidência de trabalho infantil, também estudam em escolas mais carentes em infraestrutura, recursos didáticos e com professores menos qualificados.

Bezerra e Kassouf (2006), ao analisarem os determinantes do desempenho escolar de crianças do meio rural e urbano no Brasil, concluíram que dentre os principais fatores que afetam positivamente o aprendizado das crianças no meio rural e urbano são as relacionadas a escolaridade da mãe, a renda familiar, a infraestrutura escolar de laboratórios e materiais audiovisuais, a escolaridade e o nível de salários dos professores, além da variável que representa o comprometimento dos alunos com os estudos.

⁴ De acordo com a definição de Soares-Neto *et al.* (2013), as escolas consideradas no nível *Elementar* possuíam apenas água, sanitários, energia, esgoto e cozinha. Na categoria *Básica* estavam presentes, além dos itens já citados, sala de diretoria e equipamentos como TV, DVD, computadores e impressores. *Adequada*, além dos itens das categorias anteriores, possuíam salas disponíveis para os professores, bibliotecas, laboratórios de informática, acesso à internet e espaços para o convívio social das crianças, como quadras esportivas. No nível *Avançado*, além dos itens presentes nos níveis anteriores, as escolas possuíam uma infraestrutura escolar mais robusta e mais próxima do ideal, com a presença de laboratório de ciências e dependências adequadas para atender estudantes com necessidades especiais.

3 ESTRATÉGIA EMPÍRICA

Nessa seção apresenta-se o método empírico aplicado no estudo. Inicialmente, o modelo de Oaxaca e Blinder (1973) e em seguida o método generalizado proposto por Firpo, Fortin e Lemieux (2007).

3.1 Funções de produção educacional e decomposição de Oaxaca-Blinder (1973)

Tendo como referência os estudos consolidados na literatura sobre os fatores associados ao rendimento escolar dos estudantes, as diferenças no desempenho escolar entre os que estudam em escolas no meio rural e em áreas urbanas é estimado por uma Função de Produção Educacional (FPE) que relaciona uma serie de “inputs”, incluído as características observáveis e não observáveis dos alunos, de seus familiares e da escola que frequentam. Cada característica considerada possui uma determinada taxa de retorno ou efeito marginal sobre o resultado de interesse, isto é, nos níveis de aprendizado alcançado pelos estudantes (RODRIGUES, 2009).

A FPE pode ser representada pela seguinte equação:

$$D = f(X, F, E, P) + \varepsilon \quad (1)$$

Onde, D representa uma variável de resultado, no presente caso, o desempenho ou nível de habilidade acadêmica adquirida pelo aluno, que está em função de um conjunto de fatores agrupados em quatro categorias: características individuais e motivação aos estudos (X), características familiares (F), características da escola que estuda (E), características de seus professores (P) e o termo do erro (ε), que representa todas as variáveis ou características não observáveis que influenciam no desempenho escolar do estudante que não foram capturadas pelo conjunto de informações disponíveis⁵. Desse modo a equação (1) pode ser expressa na forma linear como:

$$D_{ijk} = \beta_0 + \beta_1 X_{ij} + \beta_2 F_{ij} + \beta_3 E_{ij} + \beta_4 P_{ij} + \varepsilon_i \quad (2)$$

onde, $i = 1, \dots, n$, $j = 1 \dots J$ $k = r, u$

em que D_{ijk} é a nota do aluno i , na escola j e na zona k . E β_0 a β_4 são os parâmetros a serem estimados.

⁵ Mais informações sobre a Função de Produção Educacional podem ser obtidas nos trabalhos de Hanushek e Woessmann (2011, 2012) e Wobmann (2003)

Portanto, como foi apresentado inicialmente, o modelo a ser aplicado neste estudo é uma decomposição do tipo Oaxaca-Blinder (OB) (OAXACA, 1973; BLINDER, 1973) para os *quantis* de distribuição das notas dos estudantes. Por isso, aqui é expresso o modelo tradicional de OB para analisar a estimação para a média.

Com base na FPE de cada grupo é aplicada a metodologia de decomposição de OB, com o objetivo de explorar os fatores associados às diferenças nos resultados educacionais entre os grupos de alunos nas zonas rural e urbana. Este método é originalmente utilizado para análises discriminatórias dos salários entre distintos grupos, como por sexo, raça, setores (público-privado) ou outras características dos trabalhadores. Embora essa metodologia já venha sendo utilizada com bastante frequência no contexto internacional⁶ para explicar os diferenciais de rendimento escolar entre estudantes, no Brasil, poucos estudos aplicam essa técnica, com algumas exceções, como os estudos de Rodrigues (2009) e Lombardi Filho, Cruz e Irffi (2014).

Portanto, a técnica de decomposição de OB permite identificar as causas que dão origem às diferenças entre dois grupos (no caso, rural e urbano), por meio de dois componentes: uma parte explicada pelas características observadas, que são obtidas com base nas características dos alunos de cada grupo, e outro componente inexplicável ou resíduos, onde são incluídos todos os fatores não diretamente mensuráveis ou que foram omitidas do modelo. Portanto, a técnica de decomposição de OB, que tem como base a FPE, pode ser expressa para cada grupo como:

$$D_{ik} = X_i\beta_i + \varepsilon_i \quad (3)$$

Em que D representa o logaritmo do rendimento escolar do aluno i ; X é o vetor que aponta as características observáveis dos estudantes (incluindo as características individuais, da família e da escola); k é definida pela zona rural (r) e urbana (u); e o termo do erro (ε) que admite a independência das covariáveis, isto é $E[\varepsilon_i/X] = 0$.

Assim, a decomposição do desempenho acadêmico entre os estudantes das zonas urbana e rural é definida por:

$$D_u - D_r = [\hat{\beta}_u(\bar{X}_u - \bar{X}_r)] + [\bar{X}_r(\hat{\beta}_u - \hat{\beta}_r)] \quad (4)$$

em que $D_u - D_r$ representa a diferença do desempenho dos alunos avaliados

⁶ Essa metodologia foi usada por Aristizabal *et al* (2016), Ramos *et al* (2016), Amini; Nivorozhkin (2015), Aristizabal *et al* (2014), Badr *et al* (2012) Osorio *et al* (2011), Valenzuela *et al* (2009) e outros.

na Prova Brasil nas avaliações de Português e Matemática em escolas urbanas e rurais; $\hat{\beta}_u - \hat{\beta}_r$ são os coeficientes estimados a partir da FPE e; $\bar{X}_u - \bar{X}_r$ são os vetores dos valores médios das características observadas para os alunos do meio urbano e rural, respectivamente. A equação (4) pode ser decomposta em dois componentes. O primeiro termo à direita da equação $[\hat{\beta}_u(\bar{X}_u - \bar{X}_r)]$ representa o componente “explicado” que mede a parcela do diferencial do rendimento médio dos alunos explicada pela diferença nas características observáveis. Será chamado neste estudo de **efeito-característica**.

O segundo termo $[\bar{X}_r(\hat{\beta}_u - \hat{\beta}_r)]$, aqui denominado de **efeito-estrutural**, representa o componente “não explicado” e mede o efeito da disparidade do desempenho dos alunos que reflete as diferenças no retorno das características dos indivíduos. Este efeito surge quando um grupo apresenta maior vantagem em relação ao outro, dadas as mesmas características observadas. O efeito estrutural capta todos os potenciais efeitos das distintas características não observáveis dos alunos⁷.

Como o método de decomposição de OB decompõe apenas a diferença no desempenho escolar dos alunos na média, e como há o interesse em saber como se comporta o efeito composição em diferentes pontos da distribuição do rendimento, neste estudo, utiliza-se uma generalização do método de OB, proposto por Firpo, Fortin e Lemieux (2007) que tem como base na regressão de *quantis* não condicionadas em diferentes pontos da distribuição da variável dependente.

3.2 Decomposição por Regressão Quantílica Incondicional (RQI)

Quando se pretende realizar uma decomposição do tipo Oaxaca-Blinder em diferentes *quantis* ao longo da distribuição, não se pode simplesmente executar RQ nos *quantis* de interesse e usar os coeficientes estimados no método OB, pois, diferentemente dos obtidos via Mínimos Quadrados (MQ), que possuem interpretação condicional e incondicional, os coeficientes estimados pelas regressões quantílicas têm apenas a interpretação condicional.

Sabe-se que o método MQ estima os coeficientes (β) de uma média condicional, $E[Y|X] = X\beta$, da qual se interpreta β como $\partial E[Y|X]/\partial X$. Aplicando a lei

⁷ Em uma versão estendida, Jann (2008) incorpora um terceiro termo a equação original desenvolvida por Oaxaca-Blinder (1973) que representa a interação dos coeficientes e as dotações.

das expectativas iteradas, $E_X[E[Y|X]] = E[Y] = E[X]\beta$. β também pode ser explicado como o efeito das alterações na média de X na média incondicional de Y, porém essa lei não é válida para *quantis*, ou seja, se aplicar a esperança condicional em *quantil* τ específico, $Q_\tau(X)$, não se encontra o *quantil* incondicional, $E_X[Q_\tau(X)] \neq Q_\tau$. Consequentemente, os coeficientes da regressão linear quantílica $Q_\tau(X) = X\beta$, tem somente interpretação condicional, de modo que não se pode interpretá-lo como o efeito das alterações na média de X no *quantil* τ th marginal (incondicional).

Dessa forma, RQ fornece uma abordagem dos impactos diferenciados de covariáveis ao longo da distribuição de um resultado, assim, esta é utilizada para avaliar o efeito de uma variável em um *quantil* do resultado condicional a valores específicos de outras variáveis. Porém, segundo Borah e Basu (2013), na maioria dos casos, RQ pode gerar resultados que muitas vezes não são generalizáveis ou interpretável.

Na média, é suficiente estimar somente o valor médio condicional nos quantis de interesse, embora todo o contrafactual da distribuição acumulada incondicional, $F_{Y_B^A}(Y)$, deve ser estimada, e um contrafactual *quantil* incondicional de interesse pode ser recuperado por inversão, $Q_{\tau,B}^A = F_{Y_B^A}^{-1}(\tau)$, no qual os subscritos A e B, representam os dois grupos em comparação.

Vários procedimentos para a obtenção de distribuição contrafactual $F_{Y_B^A}(Y)$ tem sido sugerido na literatura. Juhn, Murphy e Pierce (1993), Machado e Mata (2005) e Melly (2005) visam substituir cada resultado em B com um contrafactual Y_B^A . Juhn, Murphy e Pierce (1993) utilizam uma abordagem de imputação residual, enquanto Machado e Mata (2005) se baseia em simulações e regressões *quantil* condicional. DiNardo *et al.* (1996) propuseram uma abordagem semiparamétrica reponderada, que usa um fator de reponderação estimado de "impor" as características dos indivíduos do grupo B sobre os indivíduos do grupo A. Existem também métodos que, primeiro estimam a distribuição condicional $F_{Y_A|X_A}(Y|X)$ e depois integrá-lo sobre a distribuição dos indivíduos do grupo B, $F_{X_B}(X)$ para obter a distribuição contrafactual $F_{Y_B^A}(Y)$. Estes incluem os métodos paramétricos de Donald, Green e Paarsch (2000) e Fortin e Lemieux (1998), bem como uma abordagem menos restritiva de Chernozhukov, Fernandez-Val e Melly (2013).

Se o objetivo for apenas realizar uma decomposição agregada do *quantil*, ou

seja, estimar os efeitos explicáveis e não explicáveis da variável de resultado sem o interesse de se fazer uma decomposição mais detalhada, qualquer um desses métodos citados anteriormente poderia ser utilizado, caso contrário seriam limitados e inapropriados. Apesar de alguns permitirem, os procedimentos variam em grau de complexidade ou a decomposição deve ser feita sequencialmente, assim, os resultados irão depender da ordem que as características foram introduzidas. Firpo, Fortim e Lemieux (2007) propuseram um procedimento simples que permite fazer uma decomposição detalhada no quantil de interesse, cujos resultados independem da sequência. O método baseia-se na função de regressão de influência (FI) ou RQI de Firpo, Fortim e Lemieux (2009).

A RQI tem sido aplicado em diversos trabalhos na literatura, tais como: Fortin (2008), Edoka (2012), Le e Booth (2013), Borah e Basu (2013), Rubil (2013), Fisher e Marchand (2014), Kassenboehmer e Sinning (2014), Morin (2015) e Nguyen (2015). Segundo Firpo, Fortim e Lemieux (2007), existem duas formas de se obter o efeito de uma covariável no *quantil* incondicional. Primeira, usar as estimativas dos coeficientes da RQ para recuperar o impacto incondicional, porém, apesar de intuitivamente simples, muitas vezes torna-se intratável. Especificamente, eles mostram que o efeito parcial de uma covariável em um *quantil* incondicional de Y pode ser escrito como uma média ponderada (através da distribuição de X) do efeito parcial sobre um *quantil* condicional específico de Y, que corresponde ao *quantil* incondicional de interesse. Portanto, se conseguir mapear todos os *quantis* incondicionais de Y para os *quantis* condicionais correspondentes sob diferentes argumentos condicionado, então tal abordagem ponderada pode ser facilmente implementada. No entanto, é evidente que isso é uma tarefa árdua, requer técnicas não paramétricas, e é muitas vezes intratável.

Uma alternativa para a abordagem acima mencionada foi proposta por Machado e Mata (2005), que assumiu uma mudança na distribuição incondicional ao longo do tempo e a decompôs em componentes que são atribuíveis a mudanças na distribuição marginal de diferentes Xs. No entanto, esse método captura o efeito total de uma alteração na distribuição marginal de X sobre todos os *quantis* incondicionais de Y, mas não apenas em um *quantil* específico.

A segunda abordagem foi proposta por Firpo, Fortim e Lemieux (2007), o qual soluciona o problema de intratabilidade e supera a limitação do modelo

condicional. Eles sugerem um modelo RQI baseadas nos conceitos de Função de Influência (FI) e na Função de Influência Recentrada (FIR), tal como utilizado na robusta literatura estatística (HAMPEL *et al.*, 1986). Uma FI é uma ferramenta analítica que pode ser usada para avaliar o efeito (influência) de uma remoção/adição de uma observação sobre o valor de uma estatística, $v(F)$, sem ter que recalculá-la, e é definida por:

$$FI(y, v(F)) = \lim_{\varepsilon \rightarrow 0} \left\{ \frac{v[(1 - \varepsilon) \cdot F + \varepsilon \cdot \delta_y] - v(F)}{\varepsilon} \right\}, 0 \leq \varepsilon \leq 1 \quad (5)$$

Onde F representa a função de distribuição acumulada de Y e δ_y é uma distribuição que agregada no valor y .

Uma FIR é obtida adicionando a estatística $v(F)$ a FI:

$$FIR(y, v) = v(F) + FI(y, v) \quad (6)$$

Uma característica dessa abordagem é que a expectativa é igual a $v(F)$, ou seja, se a estatística de interesse for a média, então, a FI será o residual avaliado em y e a FIR será o próprio valor de y :

$$FI(y, \mu) = \lim_{\varepsilon \rightarrow 0} \left\{ \frac{(1 - \varepsilon) \cdot \mu + \varepsilon \cdot y - \mu}{\varepsilon} \right\} = y - \mu, \text{ e } FIR(y, \mu) = \mu + y - \mu = y \quad (7)$$

Por conseguinte, a regressão da FIR para a média, em X , produz os mesmos coeficientes da regressão MQ padrão.

Quando a estatística de interesse é um quantil τ específico da distribuição:

$$FI(y, q_\tau) = \frac{\tau - I\{Y \leq q_\tau\}}{f_Y(q_\tau)} \quad (8)$$

Onde q_τ se refere ao quantil τ th da distribuição incondicional de Y , $f_Y(q_\tau)$ é a função densidade de probabilidade de Y avaliado no τ -ésimo quantil (q_τ) e $I\{Y \leq q_\tau\}$ é uma variável indicadora que delimita a ocorrência de Y até o limite do quantil. Por definição,

$$FIR(y, q_\tau) = q_\tau + FI(y, q_\tau) \quad (9)$$

Quando a esperança condicional da $FIR(y, q_\tau)$ é modelada como uma função das variáveis explicativas, X , $E[FIR(y, q_\tau)|X = x] = m_\tau(x)$, uma regressão FIR pode ser vista como uma RQI (FIRPO; FORTIM; LEMIEUX, 2009). Por definição da FIR, $E_X E[FIR(y, q_\tau)|X] = q_\tau$, então $E_X(dm_\tau(x)/dX)$ pode ser interpretado como efeito marginal de uma pequena variação na distribuição das covariadas no τ quantil

incondicional de Y, mantendo tudo o mais constante.

Firpo, Fortim e Lemieux (2007) afirmam que os procedimentos para estimação da RQI são simples e semelhantes à regressão MQ e se dá da seguinte forma: Primeiro, para um quantil τ específico, estima-se a RIF do *quantil* τ de Y seguindo (3.2.4) e (3.2.5). q_τ é estimado usando a estimativa da amostra do quantil τ incondicional, assim como, a densidade $f_y(q_\tau)$ no ponto q_τ é estimada usando método de kernel. Segundo, aplica-se MQ na $FIR(y, q_\tau)$ sobre as covariadas observadas, X. Os autores também destacam as medidas para calcular o efeito quantílico incondicional, que mede o impacto de uma covariável sobre Y num *quantil* específico, onde, na regressão FIR-MQO, pressupõe-se que, os efeitos parciais quantílicos incondicionais são os coeficientes estimados⁸. Dessa forma, enquanto a regressão quantílica condicional provê estimativas do retorno de características individuais, *coeteris paribus*, onde, este retorno é variável entre os indivíduos de acordo com o *quantil* incondicional ao qual ele pertence, a regressão quantílica incondicional estima, *coeteris paribus*, o efeito de pequenas mudanças em uma característica dos indivíduos em cada *quantil* da distribuição, permitindo assim, avaliar o efeito sobre diversas estatísticas da distribuição (FOURNIER; KOSKE, 2012).

Uma característica do método RQI é que os resultados podem ser diretamente aplicados ao modelo de decomposição de Oaxaca-Blinder⁹, o qual será utilizado por este trabalho para verificar os fatores que contribuem para os diferenciais de notas da Prova Brasil entre os alunos das escolas públicas municipais da zona rural e urbana do Brasil ao longo de toda a distribuição, definidos genericamente como A e B, aplica-se um método de decomposição proposto por Firpo, Fortim e Lemieux (2007). Generalizando a decomposição de Oaxaca-Blinder, pode-se representar a diferença nas distribuições de notas entre os dois tipos de escolas por meio de alguma estatística dessas distribuições. Assim, o diferencial total entre as distribuições é escrito como:

$$\Delta^v = v(F_{yA}) - v(F_{yB}) \quad (10)$$

Onde $v(F_{yr})$ representa uma estatística da distribuição de notas para a

⁸Podem ser implementados pelo software STATA a partir do comando *rifreg*. Disponível no site: <<http://faculty.arts.ubc.ca/nfortin/datahead.html>>.

⁹ A regressão FIR quantílica pode ser usada na decomposição de Oaxaca por meio do comando *oaxaca* no software STATA.

escola $r = A, B$. A decomposição irá dividir Δ^v em um componente associado a diferenças de características observáveis dos indivíduos e um componente associado a diferenças na estrutura de notas entre os dois tipos de escolas. Para tanto, faz-se necessário a construção de um contrafactual (F_{yC}) para que se possa simular uma distribuição de notas compondo a estrutura de notas dos alunos A e a distribuição de características (observáveis e não observáveis) dos alunos da escola B. Construindo esse contrafactual e algumas manipulações algébricas pode-se obter:

$$\Delta^v = [v(F_{yB}) - v(F_{yC}) + [v(F_{yC}) - v(F_{yA})]] \quad (11)$$

$$\Delta^v = \Delta_S^v + \Delta_X^v \quad (12)$$

O primeiro termo constitui a *parte não-explicada* da decomposição – efeito coeficiente (retorno) ou discriminação quando se refere a mercado de trabalho, definido como uma medida da diferença entre os coeficientes estimados das covariadas para o grupo B em relação ao A. O último termo deste diferencial total é o efeito quantidade, considerado a *parte explicada* da decomposição – efeito dotação ou característico, o qual refere-se à diferença nas dotações dos dois grupos, onde a diferença de nota entre os alunos das escolas no meio rural frente aos das zona urbana pode ser parcialmente atribuída ao fato de possuírem características mais favoráveis para atingir notas mais elevadas.

Aplicando o valor esperado das funções de influência recentrada e assumindo uma especificação linear, têm-se as equações (13) e (14) para as distribuições observadas e para a distribuição contrafactual, respectivamente.

3.3 Bases de dados e tratamento das variáveis

Grande parte das informações utilizadas neste estudo faz parte dos microdados da Prova Brasil/Saeb de 2015, disponibilizada pelo INEP¹⁰. Aplicado pela primeira vez em 1990, o Sistema de Avaliação da Educação Básica (Saeb) desempenha um importante papel de monitoramento e acompanhamento da evolução da qualidade no ensino em todo do Território Brasileiro. Desde sua criação, passou por diversas mudanças metodológicas. Sua última reestruturação aconteceu em 2005, quando passou a ser composto por duas avaliações: a Avaliação Nacional da Educação Básica (ANEB)

¹⁰ Os microdados estão disponíveis em: <<http://portal.inep.gov.br/basica-levantamentos-acessar>>

e Avaliação Nacional do Rendimento Escolar (ANRESC), conhecida como Prova Brasil.

A base metodológica das duas avaliações é a mesma, a diferença está na população de estudantes que abrange e, conseqüentemente, nos resultados que cada uma fornece. A Prova Brasil avalia as turmas do 5º e 9º anos do Ensino Fundamental da rede pública (estadual, municipal e federal) nas áreas rurais e urbanas, desde que a escola tenha mais de vinte alunos matriculados no ano avaliado. Por sua vez, a ANEB faz uma seleção aleatória das escolas da rede pública e privada com base nos seguintes critérios: escolas das zonas rural e urbana com turmas que tenham mais de dez alunos matriculados nos anos finais de cada ciclo do Ensino Fundamental e na última série do Ensino Médio.

Além das notas dos testes padronizados de Língua Portuguesa e Matemática, por aluno, escola, município e estado, a Prova Brasil também disponibiliza questionários socioeconômicos respondidos pelos alunos, professores e diretores, que têm por objetivo entender a situação social e econômica dos estudantes e da escola que eles frequentam. Com o apoio dessas bases de dados, é possível obter informações sobre o ambiente familiar do estudante, seus hábitos de estudos e leitura, motivação e trajetória, que podem ser usados para explicar seu desempenho na prova, e também, identificar as principais dificuldades encontradas no contexto escolar. Dados complementares referentes a infraestrutura, professores e matrículas foram obtidos pelo Censo Escolar de 2015 e Indicadores Educacionais¹¹ retirados diretamente do sítio do INEP.

Portanto, sendo a Prova Brasil um importante instrumento ou ferramenta para medir o nível de qualidade e desempenho escolar dos estudantes, este estudo tem como base as notas dos alunos avaliados no 5º ano do EF de 2015 da rede pública de ensino estadual e municipal¹². Uma das principais justificativas para a escolha deste grupo decorre de considerar que essa é uma idade em que as crianças ainda são muito dependentes dos pais ou responsáveis e que a qualidade da educação ofertada a elas pode ser decisiva em sua trajetória escolar.

¹¹ Mais informações em: < <http://portal.inep.gov.br/indicadores-educacionais> >.

¹² Como o objetivo do estudo é decompor o desempenho escolar dos estudantes do meio rural em relação aos alunos de escolas urbanas, e como a existência de escolas da rede federal e privada na zona rural é menor que 1%, achou-se mais conveniente excluí-las da base de dados.

Sendo o objetivo desse estudo decompor o diferencial de desempenho escolar entre alunos nos meios rural e urbano, as variáveis dependentes consideradas nesta pesquisa correspondem ao desempenho do aluno nas avaliações de Português e Matemática e, para efeito de estimação da FPE, será considerado o *logaritmo* natural da respectiva variável. As variáveis explicativas relacionadas às características do aluno, da família, dos professores e da escola foram selecionadas com base no referencial teórico sobre o tema e nas possibilidades dos questionários contextuais da Prova Brasil/Saeb e Censo Escolar de 2015. Na Tabela 1, estão as variáveis utilizadas no estudo.

Tabela 1 - Descrição das variáveis incluídas no modelo

Variáveis	Descrição
Variável Dependente	
nota_lp	Nota de Língua Portuguesa (escala do SAEB/97)
nota_mat	Nota de Matemática (escala do SAEB/97)
ln_lp	ln da nota de Português
ln_mat	ln da nota em Matemática
Variáveis Explicativas	
<i>Características do aluno</i>	
idade	Idade do Aluno
Sexo	1-Masculino; 0- Feminino
branco	1 - Branco; 0 - c.c.
nunca_reprov	1- Nunca reprovou; 0-c.c
nunca_aband	1- Nunca abandonou a escola; 0-c.c
idade_certa ^(a)	Se está na idade certa para frequentar o 5º ano (1-sim; 0-c.c.)
atitu_leitura ^(b)	Atitude em relação a leitura (1-se ler com frequência; 0-c.c.)
possui_pc	Tem computador em casa (1 -sim; 0 - c.c.)
faz_dever1	Faz sempre dever de casa de português (1-sim; 0-c.c)
faz_dever2	Faz sempre dever de casa de matemática (1-sim; 0-c.c)
trab_fora	Trabalha fora de casa (remunerado ou não) (1-sim; 0-c.c)
<i>Característica da família</i>	
SEF	Status Econômico Familiar (SEF)
Edu_pai1	Pai é analfabeto ou tem EF incompleto (1- sim; 0 -c.c)
Edu_pai2	Pai tem ensino fundamental completo (1 - sim; 0 -c.c)
Edu_pai3	Pai tem ensino médio completo (1 - sim; 0 -c.c)
Edu_pai4	Pai tem ensino superior completo (1 - sim; 0 -c.c)
Edu_mae1	Mãe é analfabeta ou tem EF incompleto (1-sim; 0-c.c)
Edu_mae2	Mãe tem ensino fundamental completo (1 - sim; 0 -c.c)
Edu_mae3	Mãe tem ensino médio completo (1 - sim; 0 -c.c)
Edu_mae4	Mãe tem ensino superior completo (1 - sim; 0 -c.c)
mora_pai_mae	Reside com o pai e a mãe (1 - sim; 0- c.c)
pais_incentiva	Pais incentiva ir à escola e/ou a não faltar às aulas (1-sim; 0-c.c)
<i>Características dos professores</i>	
prof_superior ¹	Proporção de docentes com formação superior
experiencia	Prop. de docentes do 5º ano com mais de 6 anos de experiência
salário	Prop. de doc. do 5º ano que ganham + de R\$ 3.152,01 (ou 4 salários-mínimos de 2015)
<i>Características da escola</i>	
urbano	<i>Dummy</i> de localização da escola (1 - Urbano; 0 - Rural)
rede_munic	Rede de ensino (1 - Municipal; 0 - Estadual)
tam_escola ¹	Tamanho da Escola (nº total de matriculas em todas as turmas)
aluno_prof ¹	Proporção de alunos por professor (1º a 5º ano)
m_tempo_aula ¹	Média de duração das turmas (em minutos)
escass_material ^(c)	Escassez de material de aprendizagem (1 - sim; 0 c.c)
sel_diretor	Forma de escolha do diretor (1-selecionado e/ou eleito; 0-c.c.)
falta_de_prof ^(d)	Falta de professores (1 - sim; 0 c.c)
IRD ¹	Indicador de Regularidade Docente
IIE ¹	Índice de infraestrutura da escola
IMCE	Indicador de má conservação da escola

Fonte: RODRIGUES (2017). Microdados do Censo da Educação Básica-INEP/MEC e Prova Brasil 2015. Nota: ¹refere-se a informações coletadas no Censo Escolar e dos Indicadores Educacionais de 2015. ^(a) A variável capta se o aluno está na idade correta para frequentar o 5º ano, ou seja, se tem dez anos de idade, com margem para um ano a mais ou um a menos; ^(b) é relacionada à frequência em que o aluno lê jornais, livros e/ou revistas em geral, - sendo considerada a resposta de que o aluno sempre lê. ^(c) Variável extraída do questionário do diretor (Questão 71), sendo 1 quando a resposta é sim (pouco, moderado e muito). ^(d) Variável extraída do questionário do diretor (Questão 73), sendo 1 quando a resposta é sim (pouco, moderado e muito).

O **Status Econômico Familiar (SEF)** – refere-se a um indicador que mede os bens duráveis e prestação de serviços por terceiros no domicílio do aluno e grau de instrução do chefe do domicílio. É baseado no Critério de Classificação Econômica Brasil (CCEB) da Associação Brasileira de Empresas de Pesquisa (ABEP), de 1º de janeiro de 2014. Seu principal objetivo é estimar o poder de compra das famílias e classificá-las em oito classes, que são elas: A1 (42 – 46 pontos); A2 (35 a 41 pontos), B1 (29 a 34 pontos), B2 (23 a 28 pontos), C1 (18 a 22 pontos), C2 (14 a 17 pontos), D (8 a 13 pontos) e E (0 a 7 pontos) (ALMEIDA, 2014).

No Quadro 1 são apresentados os principais itens utilizados para elaboração do SEF.

Quadro 1 - Descrição dos itens utilizados para a elaboração do status econômico familiar

Escolaridade		Pontos				
Escolaridade do chefe de família ¹	Analfabeto ou primeira fase do ensino fundamental incompleta	0				
	Primeira fase do ensino fundamental completa ou segunda fase incompleta	1				
	Segunda fase do ensino fundamental completa ou ensino médio incompleto	2				
	Ensino médio completo ou superior incompleto	4				
	Superior completo	8				
Itens	Quantidade de Itens e pontos					
	0	1	2	3	4 ou +	
Televisão em cores	0	1	2	3	4	
Rádio	0	1	2	3	4	
Banheiro	0	4	5	6	7	
Automóvel	0	4	7	9	9	
Empregada mensalista	0	3	4	4	4	
Máquina de lavar	0	2	2	2	2	
Videocassete e/ou DVD	0	2	2	2	2	
Geladeira	0	4	4	4	4	
Freezer (aparelho independente ou parte da geladeira duplex)	0	2	2	2	2	

Fonte: RODRIGUES (2017) com base do Índice ABEP (2014)

Nota: ¹ Assim como definido por Almeida (2014), chefe da família é a pessoa (pai, mãe ou responsável) com maior nível de escolaridade.

A variável **prof_superior** representa o percentual de professores com formação superior na escola do 1º ao 5º ano do Ensino Fundamental. Pela impossibilidade de identificar corretamente os professores que lecionavam Português e Matemática em turmas do 5º ano, haja vista do fato que, no questionário do professor, a resposta poderia ser mais de uma disciplina, as informações das variáveis de experiência e salários foram calculadas em relação à proporção de professores que responderam ao questionário. Optou-se por sintetizar as respectivas variáveis em professores que tinham mais de seis anos de experiência como professor (1 – se sim; 0 - caso contrário (c.c)) e

com salário bruto como professor superior a quatro salários-mínimos de 2015 (1 – se sim; 0 - c.c).

O **tamanho da escola** é calculado pelo número total de matrículas em todos os níveis de ensino ofertado pela escola. A **Proporção de alunos por professor** foi calculada pelo total de alunos matriculados do 1º ao 5º ano e o total de professores no respectivo nível.

Já o **Indicador de Regularidade Docente** é calculado pelo INEP e avalia a regularidade do corpo docente nas escolas de Educação Básica a partir da observação da permanência dos professores nas escolas nos últimos cinco anos¹³.

O **Indicador de Má Conservação da Escola (IMCE)**¹⁴ foi calculado com amparo em 13 perguntas feitas sobre o estado de conservação dos itens e equipamentos do prédio ou a inexistência deles. Os itens considerados são os seguintes: telhado, paredes, piso, entrada do prédio, pátio, corredores, salas de aula, portas, janelas, banheiros, cozinha, instalações hidráulicas e instalações elétricas. Assim, o índice está de 0 (nenhum problema de conservação relatado) a 1 (a escola apresenta em todas as 13 questões problemas na conservação do prédio) (ALMEIDA, 2014).

O **Índice de Infraestrutura da Escola (IIE)** foi gerado pela análise de componentes principais, que engloba a existência das seguintes características das escolas obtidas no Censo Escolar de 2015: abastecimento de água por via rede pública, de esgoto sanitário por meio rede pública, coleta periódica de lixo, sala de diretoria, sala de professores, laboratório de informática e ciências, quadra de esportes, cozinha, biblioteca, parque infantil, sanitário adequado a aluno portador de necessidades especiais (PNEs), dependências e vias adequadas a PNEs, televisão, de videocassete, de DVD, de parabólica, de copiadora, de retroprojeto, de impressora, de computadores, de acesso à internet, e de alimentação escolar para os alunos.

¹³ Mais informações podem ser obtidas na Nota Técnica CGCQTI/DEED/INEP nº 11/2015, disponível em:

<http://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_regularidade_vinculo/nota_tecnica_indicador_regularidade_2015.pdf>.

¹⁴ O cálculo do índice de má conservação da escola (IMCE) é dada por: $IMCE = (\sum_{i=1}^{13} X_i / 13)$, onde X_i é igual a 1 se a escola apresenta estado de conservação ruim ou inexistente o item ou equipamento i .

4 RESULTADOS

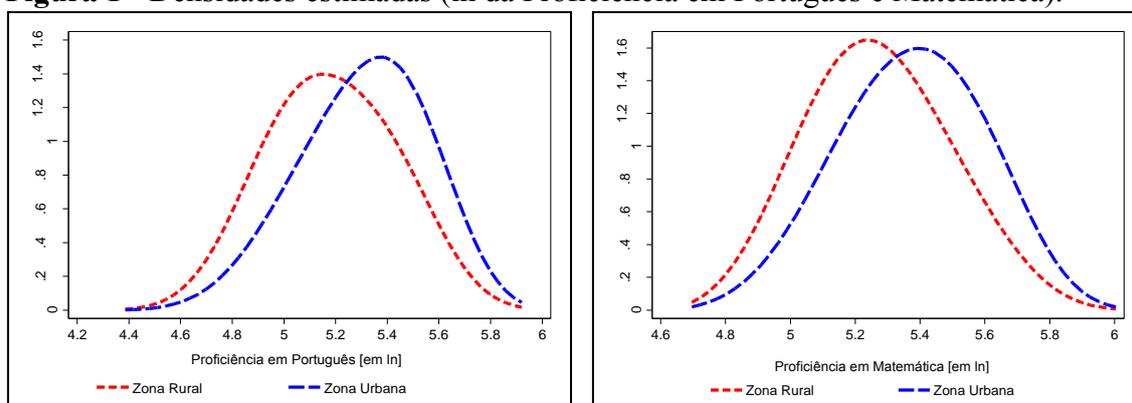
Aqui, apresentam-se os resultados da pesquisa de acordo como os dados coletados da Prova Brasil. Inicialmente com base nos dados descritivos e em seguida com nas regressões estimadas.

4.1 Análises descritivas

Na Tabela 1.2, são expostas as estatísticas descritivas das variáveis utilizadas neste estudo. Foi selecionada uma amostra de 784.120 alunos do 5º ano do Ensino Fundamental de escolas das redes municipal e estadual, sendo que 710.680 alunos estudavam em escolas urbanas e 73.440 frequentavam escolas na zona rural.

Analisando a nota dos alunos, segundo a zona da escola que frequenta, observa-se que na zona urbana, em Português, a nota média é de 207,84 pontos e, em Matemática, 220,72 pontos. No meio rural, a média é bastante inferior à zona urbana: em Português, a nota média é de 183,77 pontos, uma diferença de mais de 24 pontos a favor das escolas localizadas no meio urbano. Em Matemática, a nota média alcançada pelos estudantes é de 198,74 pontos o que representa uma diferença de aproximadamente 22 pontos inferior à média atingida pelos estudantes de escolas urbanas. Nos gráficos da Figura 1, mostra-se a densidade estimada do rendimento escolar dos estudantes nas duas avaliações para o meio rural e urbano, medido pelo logaritmo (natural) do rendimento escolar dos alunos.

Figura 1 - Densidades estimadas (ln da Proficiência em Português e Matemática).



Fonte: RODRIGUES (2017) com base no microdados da Prova Brasil/Saeb, 2015.

Em relação ao perfil dos alunos, na Tabela 1.2, nota-se que, entre as escolas

rurais, a idade média apresentada pelos estudantes é ligeiramente superior à idade expressa no meio urbano e predominantemente são do sexo masculino. Em relação aos que nunca foram reprovados, no meio rural é a proporção, é de 62%, enquanto, na média das escolas urbanas, 72% nunca reprovaram. A taxa de abandono também é maior entre os estudantes das escolas rurais, em média, 10% já abandonaram a escola pelo menos uma vez. Em contrapartida, na zona urbana, apenas 6%. Outro ponto importante a ser observado é que a proporção de alunos no meio rural trabalhando fora de casa, representa cerca de oito pontos percentuais superiores à média dos alunos do meio urbano.

Quando a comparação é feita em relação ao *Status* Econômico Familiar (SEF), percebe-se que, em média, as crianças que estudam em escolas urbanas vivem em domicílios com melhores condições do que aquelas que estudam no meio rural. Além disso, a escolaridade dos pais também é pior entre os alunos no meio rural; ou seja, as crianças rurais apresentam condições socioeconômicas desfavoráveis em relação às que estudam em escolas urbanas, tanto do ponto de vista de bens no domicílio, conforme o SEF, como sob o prisma de capital humano disponível no lar.

Quanto à escolarização e o salário dos professores, ainda na Tabela 2, nota-se que no meio rural a proporção de professores com Ensino Superior é menor do que no meio urbano, 61% e 83%, respectivamente. Em relação ao salário bruto recebido pelos professores, somente 23% dos professores das escolas rurais declararam receber salários com valores superiores a quatro salários-mínimos. Enquanto isso, na média das escolas urbanas, 38% recebia salários maiores do que R\$ 3.152,00. Além disso, nas escolas urbanas, a proporção de professores que trabalham há mais de seis anos como docentes é superior às do meio rural (79% e 85%, respectivamente).

Tabela 2 - Média e Desvio-Padrão das variáveis utilizadas nas regressões – meio rural e urbano¹⁵ – Brasil, 2015

Variáveis		Rural		Urbano	
		Média	D.P.	Média	D.P.
Variável Dependente					
Variáveis de Interesse	nota_lp	183,77	45,25	207,84	47,78
	nota_mat	198,74	42,81	220,72	46,34
	ln_lp	5,18	0,24	5,31	0,24
	ln_mat	5,27	0,21	5,37	0,21
Variáveis Explicativas					
<i>Características do aluno e da família</i>					
Perfil do Aluno	idade	11,23	1,31	10,91	1,03
	sexo	0,52	0,50	0,52	0,50
	branco	0,25	0,43	0,29	0,45
	nunca_reprov	0,62	0,49	0,74	0,44
	nunca_aband	0,90	0,30	0,94	0,23
	idade_certa	0,70	0,46	0,82	0,38
	atitu_leitura	0,61	0,49	0,56	0,50
	possui_pc	0,34	0,47	0,65	0,48
	faz_dever1	0,69	0,46	0,72	0,45
	faz_dever2	0,76	0,43	0,79	0,41
	trab_fora	0,22	0,41	0,14	0,35
Características da família	SEF	19,69	6,72	24,27	6,57
	Edu_pai1	0,64	0,48	0,40	0,49
	Edu_pai2	0,15	0,36	0,16	0,37
	Edu_pai3	0,11	0,31	0,22	0,41
	Edu_pai4	0,10	0,30	0,21	0,41
	Edu_mae1	0,61	0,49	0,38	0,49
	Edu_mae2	0,14	0,35	0,15	0,36
	Edu_mae3	0,14	0,34	0,25	0,43
	Edu_mae4	0,11	0,32	0,21	0,41
	mora_mae_pai	0,68	0,47	0,63	0,48
	pais_incentiva	0,92	0,27	0,95	0,21
<i>Características dos professores e da escola</i>					
Características dos docentes	prof_superior	0,61	0,31	0,83	0,20
	experiencia	0,79	0,31	0,85	0,22
	salario	0,23	0,35	0,38	0,36
Características da escola	rede_munic	0,94	0,24	0,77	0,42
	tam_escola	443,53	292,10	725,18	425,49
	aluno_prof	20,57	7,10	21,15	7,00
	m_duracao_aula	254,44	55,20	254,65	53,10
	escass_material	0,57	0,49	0,51	0,50
	falta_de_prof	0,32	0,47	0,42	0,49
	sel_diretor	0,16	0,37	0,32	0,47
	IRD	3,14	0,64	3,10	0,56
	IMCE	0,17	0,23	0,12	0,11
	IIE	0,54	0,17	0,73	0,12

Fonte: RODRIGUES (2017) com base nos microdados do Censo Escolar e Prova Brasil/Saeb, 2015.

Ao analisar as variáveis relacionadas ao contexto escolar, observa-se que,

¹⁵ O INEP não especifica o que define uma escola rural ou urbana, portanto, a definição utilizada é que: aluno rural é quem estuda em escola que se define como rural e estudantes que estudam em áreas urbanas são os que estão em escolas que se definem localizam em área urbana.

em média, as escolas urbanas têm mais alunos matriculados e que a proporção de alunos por professor é menor nas escolas rurais. E as escolas rurais também apresentam características físicas bastante diferenciadas das escolas localizadas no meio urbano. Isso pode ser observado pelo Índice de Má Conservação de Escola (IMCE) e pelo Índice de Infraestrutura da Escola (IIE).

4.2 Regressões FIR nos *quantis* incondicionais

O diferencial de rendimento dos estudantes pode esconder informações importantes ao longo da distribuição das notas, ou seja, entre os alunos que atingiram níveis de proficiência baixo e aqueles que alcançaram nota mais elevada. Os resultados estão expostos nas Tabelas A.1.1 e A.1.2 no Apêndice 1 para três *quantis* (q10, q50 e q90). Para efeito de comparação, também foi estimada uma função de produção educacional por MQO para as notas médias.

De modo geral, nota-se que os efeitos não são constante ao longo da distribuição, o que justifica a análise realizada para os *quantis*¹⁶, e apresentam variação entre as zonas de localização de cada escola. Para a disciplina de Português, na Tabela A.1.1, os coeficientes estimados para cada zona, não apresentou um padrão bem definido ao longo da distribuição.

No caso da idade do aluno, na zona urbana, essa variável foi estatisticamente significativa e negativa no q10, porém, passa a ser positivo no q50 e q90. No meio rural, não foi estatisticamente significativa no q10 nem no q50, no entanto é positivo e estatisticamente significativa no q90. Isso diverge de resultados encontrados na literatura, pois a distorção idade série é uma das variáveis responsáveis por baixo desempenho dos alunos. Este resultado, contudo, pode ser em decorrência do ano escolar da amostra analisada, já que no 5º ano do EF a distorção idade-série ainda é relativamente baixa.

Já os coeficientes estimados para sexo (tendo como referência pessoas do sexo masculino), na avaliação de Português, foram negativos e estatisticamente significantes em todos os *quantis* analisados, nas zonas rural e urbana. Quando, porém,

¹⁶ Visto que existem alunos com nota muito baixa e outros com notas muito altas, mensurar a decomposição na média pode não representar o verdadeiro diferencial entre os alunos do meio rural e urbano.

se analisa a avaliação de matemática, o sinal passa a ser positivo em todos os *quantis* da distribuição, fato sugestivo, então, de que os meninos apresentam melhor desempenho que as meninas na avaliação de Matemática, por outro lado, apresentam rendimento relativamente inferior em Português.

Com relação a raça ou cor (tendo como referência a cor branca), o primeiro *quantil* e na mediana da zona rural é estatisticamente significativo e negativo, no entanto, passa a ser positivo no q90. Da mesma forma, na zona urbana, é negativo e estatisticamente significativo no primeiro *quantil*, mas passa a ser positivo na mediana e no último quantil na avaliação de Português. Já na avaliação de Matemática, é expressa uma relação negativa apenas no q10, tanto na zona rural como urbana, sendo positivo nos demais *quantis*. Assim, alunos brancos, em geral, predominam entre as notas mais altas.

Quanto aos coeficientes nunca ter reprovação, nem abandono a escola indicaram diferenciais positivos e estatisticamente significantes em relação aos que já foram reprovados e abandonaram, tanto na média, como na maioria dos *quantis* de distribuição para as duas avaliações e nas zonas de localização das escolas, com exceção do último *quantil* (q90), que apresentou sinal inverso ao esperado (negativo e estatisticamente significativo). Da mesma maneira, observa-se uma relação positiva e significativa quando o aluno se encontra na idade certa para frequentar o 5º ano do Ensino Fundamental em todos os *quantis* de distribuição nas zonas rural e urbana e, nas duas avaliações.

Ainda, considerando as variáveis do perfil do aluno, destacam-se também as variáveis relacionadas a fazer o dever de casa (de Matemática e /ou Português) e trabalhar fora. A primeira variável exibiu sinal positivo (contra os que não fazem) e estatisticamente significativo em todos os *quantis* de distribuição e na média, tanto na zona urbana como na zona rural, nas duas avaliações. Em contraste, observa-se a relação negativa quando o estudante trabalha fora de casa (em relação aos que não trabalham). O maior impacto dessa variável sobre a nota, tanto de Português como de Matemática, é observado entre os alunos de escolas urbanas.

O SEF dos estudantes desempenha um papel importante na determinação do desempenho escolar dos alunos em todos os *quantis* e na média, sendo estatisticamente significativo e positivo nas zonas rural e urbana, na avaliação de Português e

Matemática. Quanto as *dummies* de escolaridade dos pais ou responsáveis (tendo como base se é analfabeto ou não concluiu o Ensino Fundamental), não se observou um padrão bem definido na escolaridade do pai que tem Ensino Fundamental completo e Ensino Superior completo, em alguns *quantis* o impacto foi o inverso do esperado (estatisticamente significativa e negativo). Para os pais com Ensino Médio completo o sinal foi o esperado, positiva e estatisticamente significativa em todos os *quantis* da distribuição.

Resultado similar foi encontrado por Bezerra e Kassouf (2006). Uma das justificativas usadas pelos autores é que o questionário socioeconômico do aluno é respondido pelas próprias crianças e, por terem pouca idade, pode faltar conhecimento sobre o verdadeiro nível de escolaridade dos pais e ocorrer incoerência nas respostas.

Em relação à escolaridade da mãe, contudo, nas duas avaliações e nas zonas rural e urbana apresentou o sinal esperado (positivo e estatisticamente significativa) para os alunos localizados na mediana, e em todos os *quantis* da distribuição, exceto na Edu_mae4 (mãe com Ensino Superior completo) que foi insignificante no q10 na avaliação de Português e na mediana de Matemática. Morar com o pai e a mãe e ter o incentivo deles para os estudos também apresentaram uma relação positiva e estatisticamente significativa no desempenho escolar dos estudantes.

Com relação ao nível de escolaridade dos professores (mensurado pela proporção de professores com Ensino Superior), verifica-se uma relação positiva em todos os *quantis* nas duas avaliações e para as zonas rural e urbana. O mesmo padrão não foi observado em relação a experiência e salário dos professores. A experiência (mais de seis anos que trabalha como professor) só foi estatisticamente significativa no *quantil* mediano e no q90 da zona urbana, tanto na avaliação de Português como de Matemática. Enquanto isso, o salário dos docentes exibiu sinal positivo e significativo em todos os efeitos para a zona urbana nas duas avaliações, exceto na mediana para a prova de Matemática que foi insignificante.

Por fim, considerando as variáveis relacionadas à escola, ao analisar a rede de ensino onde o aluno estuda (tendo como referência a rede municipal), na avaliação de Português, observa-se que é estatisticamente significativa apenas na zona urbana, porém, não expressou o mesmo sinal em todos os *quantis*, sendo negativo na média e na mediana e no último *quantil*, mas foi positivo no primeiro. Já na avaliação de

Matemática foi estatisticamente significativa e negativo na média, no q50 e q90, mas insignificante no q10. Para a zona rural, nesta avaliação foi estatisticamente significativa e positivo para todos os efeitos analisados.

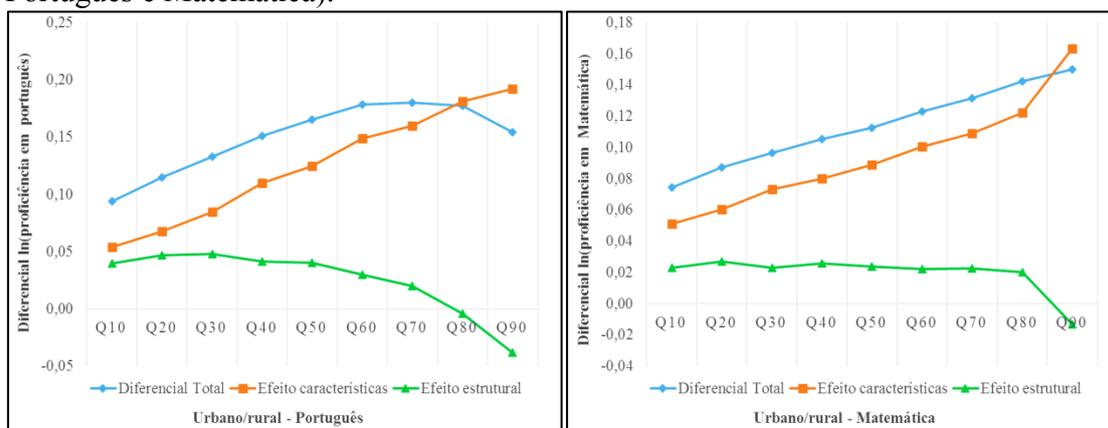
O tamanho da escola e a escassez de material pedagógico, a falta de professores e o índice de má conservação da escola (IMCE) mostraram negativo e estatisticamente significativa em todos os *quantis* da distribuição em toda a análise. Influi, portanto, negativamente, na nota do aluno de baixo e alto desempenho, ao passo que o índice de infraestrutura escolar (IIE) foi positivo e estatisticamente significativa em todos os *quantis* analisados, isto é, quanto melhor for a infraestrutura da escola, melhor será o desempenho dos estudantes.

4.3 Decomposições dos diferenciais de rendimento escolar

De acordo com os resultados da seção anterior, alguns insumos influenciam de modo diferente a nota dos alunos ao longo da distribuição em escolas rurais e urbanas, o que justifica a análise de decomposição por via de regressões quantílicas. Portanto, nessa seção é mensurada a decomposição sugerida por Firpo, Fortin e Lemieux (2007) com base em regressões de FIR que é identificar à quais são as características (explicáveis) ou estrutural (coeficientes ou inexplicáveis) que determinam os diferenciais no rendimento escolar dos estudantes entre as zonas (rural e urbana).

Pelas estimativas apresentadas nos gráficos da Figura 2, que mostram o diferencial de rendimento escolar nas avaliações de Português e Matemática, é possível observar que grande parte do diferencial do desempenho entre estudantes de escolas rurais e urbanas decorre do efeito-característica dos dois grupos e que o diferencial é positivo para alunos de escolas urbanas.

Figura 2 - Decomposição do diferencial de desempenho escolar (*ln* da proficiência em Português e Matemática).



Fonte: RODRIGUES (2017) com base nos resultados da pesquisa.

Nota: significativa a 1%.

Além disso, os gráficos mostram que o diferencial total de desempenho escolar varia ao longo da distribuição, no caso da avaliação de Português, é crescente até o *decil* 70, onde atinge seu valor máximo, e passa a decrescer desde os *decis* 80 e 90, quando o efeito estrutural passa a ser negativo. Já em Matemática, nota-se que o diferencial total entre as notas dos alunos de escolas urbanas e rurais é monotonamente crescente nos *decis*, e o aumento da diferença de desempenho na parte superior (q90) indica o efeito *glassceiling*¹⁷, conhecido na literatura brasileira como “teto de vidro”.

A Tabela A.1.3 no Apêndice 1 mostra a participação de cada efeito no diferencial total entre alunos das duas zonas. Em todos os *quantis*, a diferença entre alunos de escolas urbanas e rurais foi explicada de maneira significativa ao nível de 1%, tanto pelo efeito características como pelo efeito estrutural. De modo que, em todos os *quantis* da distribuição, os alunos de escolas urbanas têm melhor desempenho que os que estudam em escolas rurais. Sendo o efeito característica e estrutural positivos do q10 ao q70 nas duas avaliações, mas o efeito característica é predominante nestes *quantis*.

No caso da avaliação de Português, o diferencial entre os grupos é explicado pelo efeito característica e estrutural que agem positivamente sobre o desempenho dos alunos, sendo sempre crescente. Nos q80 e q90, contudo, o efeito estrutural age no sentido de reduzir as diferenças entre os grupos. Isto significa que as mudanças na

¹⁷ O efeito característica aumenta à medida que cresce a nota dos alunos.

composição do alunado, no que tange às características não mensuráveis (em relação às características individuais, dos familiares, professores e escola) operam no sentido de reduzir as diferenças entre os grupos, principalmente entre aqueles estudantes da zona rural e urbana com maiores habilidades cognitivas.

O mesmo comportamento é observado para a avaliação de Matemática, sendo que o diferencial total também é positivo e crescente em todos os *quantis*, já que tanto o efeito característica como o efeito estrutural são positivos, pois, contribuem para o aumento do rendimento escolar entre alunos dos meios rural e urbano. Exceção é feita ao *quantil* q90, em que o efeito estrutural atua no sentido oposto, reduzindo o diferencial total entre os grupos.

Portanto, dado o comportamento apresentado pelo efeito-característica, é possível afirmar, para a base dados, que as características observáveis dos estudantes contribuem significativamente com o diferencial de rendimento escolar, sendo favorável aos alunos de escolas urbanas, uma vez que o efeito é positivo ao longo de toda a distribuição.

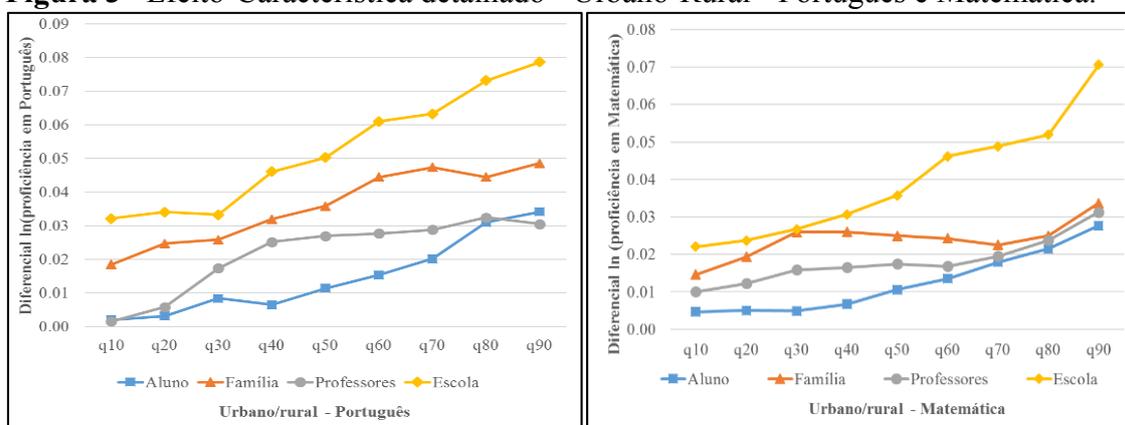
Para análise mais detalhada sobre os fatores que mais contribuem no diferencial entre alunos do meio rural e urbano, que é a principal contribuição deste estudo, a Figura 3 contém gráficos que evidenciam o efeito-característica agrupado em quatro grupos: perfil do aluno, características familiares, característica dos professores e da escola.

Examinando os resultados dos grupos mediante a distribuição, nas duas avaliações, vê-se que as características da escola e da família foram os fatores com maior importância dentro do **efeito característica**, sendo mais expressivos para estudantes de alto desempenho. Conforme avança, porém, nos *quantis*, as características individuais e as características dos professores passam a ter maior peso, e a participação dos fatores relacionados à escola e à família diminuem. Por exemplo, para alunos no q10 de pontuação, as características da escola explicam cerca de 43% e 59% do efeito-característica, nas avaliações de Matemática e Português, respectivamente (ver Tabelas A.1.4 no Apêndice 1). Já as características familiares, explica neste *quantil* o efeito-característica cerca de 28% e 34%, em ambas as avaliações. Com menor participação, as características individuais (4% e 9%, em Português e Matemática, respectivamente) e dos professores (3% e 20%, em Português e Matemática, respectivamente) (ver Tabelas

A.1.4 no Apêndice 1 para detalhes).

Para o aluno localizado no *quantil* mediano cerca de 9% e 12% do efeito-característica são explicados pelas características individuais, 29% e 28%, pelas características familiares, 22% e 20% pelas características dos professores e 40% pela escola, nas avaliações de Português e Matemática, respectivamente. Para pontuação mais elevada, no q90, as características do aluno explicam cerca de 18% e 17%, as características da família 25% e 21%, dos professores 16% e 19% e as características da escola 41% e 43%, na avaliação de Português e Matemática, respectivamente.

Figura 3– Efeito-Característica detalhado - Urbano-Rural - Português e Matemática.



Fonte: RODRIGUES (2017) com base nos Microdados do Censo Escolar e Prova Brasil/Saeb, 2015.

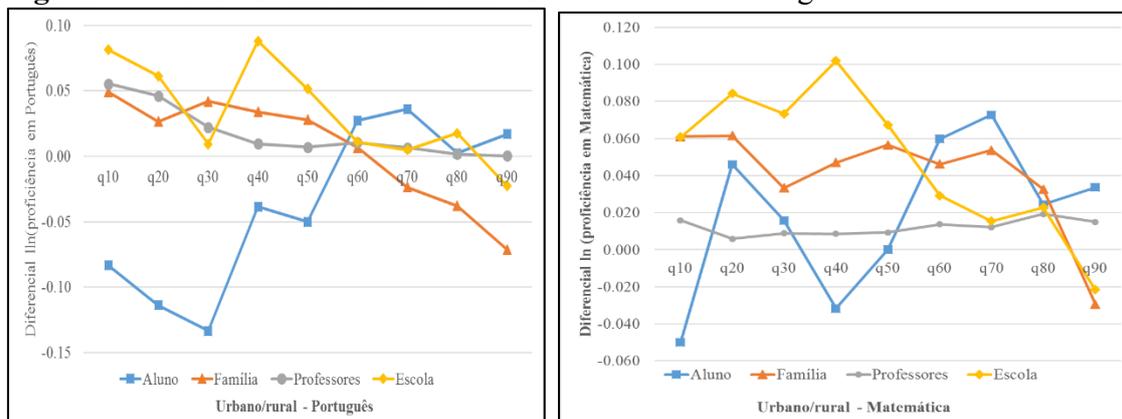
Portanto, dentro do efeito-característica, a escola e a família agem de maneira a contribuir com o diferencial de rendimento entre alunos de escolas urbanas e rurais, porém as características dos professores e dos estudantes são importantes, sobretudo para explicar o diferencial de desempenho nos *quantis* mais elevados.

O detalhamento do **efeito estrutural** é exibido na Figura 4, o qual não mostra um padrão bem definido entre os *decis*, além de ser insignificante em alguns pontos da distribuição (Ver Tabela A.1.5 no Apêndice 1). No caso das variáveis associadas ao perfil do aluno, variou consideravelmente entre os *decis*, sendo negativo entre os primeiros *decis* em ambas as avaliações e positivo a partir do *decil* 60, no entanto, não foi estatisticamente significativo nos q60, q80 e q90 em Português e nos q80 e q90 em Matemática.

Na parte inferior da distribuição, o **efeito estrutural** seria menor e contribuiria para a redução do hiato entre as notas da avaliação de Português, se os estudantes de escolas rurais tivessem uma melhoria nas características individuais; e um

maior retorno, nos *quantis* superiores, se houvesse um aumento das características familiares, principalmente para aqueles que se encontram no q80 e q90 de pontuação, que atuariam de forma a reduzir o efeito estrutural em aproximadamente 933% e 188%, respectivamente.

Figura 4 - Efeito Estrutural detalhado - Urbano-Rural - Português e Matemática



Fonte: RODRIGUES (2017) com base nos Microdados do Censo Escolar e Prova Brasil/Saeb, 2015.

Considerando a nota de Matemática, também na Figura 4, observa-se que o **efeito estrutural** das notas poderia ser reduzido, e, assim, contribuir para a queda da desigualdade de notas entre alunos de escolas urbanas e rurais, se no q10 os estudantes de escolas rurais melhorassem as características pessoais. Além disso, o efeito estrutural reduziria no q90 com o aumento do efeito pelas características da escola.

5 CONSIDERAÇÕES FINAIS

Este estudo se propôs a contribuir para o debate sobre a qualidade e a equidade da Educação Básica ao analisar as diferenças no rendimento escolar dos alunos de escolas das áreas rurais e urbanas do Brasil. Tendo como referência os estudos consolidados na literatura, que apontam para os fatores associados ao rendimento escolar dos estudantes e a riquíssima base de dados disponibilizada pelo INEP, tanto da Prova Brasil como no Censo Escolar de 2015, foi estimado o método de decomposição com base em regressões de Função de Influência Recentrada (FIR) proposto por Firpo, Fortim e Lemieux (2007), combinado com o procedimento de reponderação inspirado por Dinardo *et al.* (1996), que permite aplicar o método de decomposição de Oaxaca-Blinder tradicional.

As estimativas pelo método de Regressão Quantílica Incondicional (RQI) da função de produção educacional ao nível de estudantes apontam que as contribuições das características dos alunos, da família, dos professores e da escola não são simétricas ao longo da distribuição das notas nas duas avaliações (Português e Matemática). Também é estatisticamente significativa o hiato entre as duas zonas, apresentando vantagens os alunos que estudam em escolas urbanas em relação aos que frequentam escolas rurais. Os resultados da decomposição também mostram que, nas duas avaliações, grande parte do diferencial de rendimento entre alunos de escolas urbanas e rurais decorre do efeito-característica.

Além disso, os exercícios de decomposição por *quantis* de realização revelaram o papel crescente do efeito-característica, ou seja, quanto mais elevado é o *quantil* de desempenho, mais importantes são as características na explicação do hiato educacional entre os grupos. E o efeito estrutural (fatores não observáveis), embora tenha revelado peso relativamente menor, também contribui para o aumento do diferencial e, não pode ser ignorado, excerto nos *quantis* mais elevados, pois apresentou sinal negativo.

Efetivamente, em sentido de implementação de política públicas que tenham como objetivo diminuir as disparidades entre estudantes de escolas rurais e urbanas, os fazedores de políticas públicas devem considerar que efeitos assimétricos das características do aluno, da família, dos professores e da escola pelos *quantis* de desempenho exigem uma abordagem diferenciada entre os alunos. Onde as iniciativas

de melhorias educacionais devem ser formadas nos níveis escolares, mas que levem em conta também as diferenças na composição socioeconômica dos alunos.

Ademais, por exercer uma parcela importante no diferencial de rendimento, os professores precisam ser qualificados e bem remunerados, e as escolas devem ter boa infraestrutura de apoio, principalmente quando o nível de escolaridade dos pais e dos responsáveis dos alunos for insuficiente para mantê-los na escola e engajados no aprendizado.

Em segundo lugar, investimentos financeiros, por si, não garantem que sejam o suficiente para melhorar a qualidade e equidade educacional no Brasil. Iniciativas para melhorar aspectos não mensuráveis das escolas (como a participação dos pais, incentivo para os alunos frequentarem a biblioteca e outros) merecem atenção e são igualmente importantes. E, por fim, o possível sucesso de uma boa política educacional, para lidar com a desigualdade e melhorar a qualidade da escola pública, depende de encontrar o equilíbrio certo entre o investimento financeiro e desenvolver de um ambiente escolar propício para o aprendizado de crianças com vulnerabilidade social.

CAPÍTULO 2

MIGRAÇÃO EDUCACIONAL RURAL-URBANA: ANÁLISE DO DIFERENCIAL DE DESEMPENHO ESCOLAR NO ENSINO MÉDIO

1 INTRODUÇÃO

No Brasil, nos últimos anos, observam-se avanços significativos na expansão do ensino em todo o Território, sobretudo, a partir da década de 1990. Naquele período, o Ensino Fundamental foi praticamente universalizado, havendo aumento expressivo no acesso ao Ensino Médio. Essa conquista se deve muito à Constituição Federal de 1988, que expõe explicitamente o dever do Estado em ofertar gratuitamente a Educação Elementar, sendo responsabilidade dos estados e municípios a oferta gratuita do Ensino Fundamental e Médio a todas as crianças e adolescentes em idade escolar.

Mesmo, todavia, que o acesso à Educação seja garantido para a maioria dos brasileiros, problemas relacionados a repetência, evasão, abandono, bem como a baixa qualidade do ensino, mostram que, para a maioria da população, a educação ofertada não tem proporcionado uma formação adequada para as novas gerações, seja para a participação ativa e crítica na vida social do País ou na aquisição de habilidades e conhecimentos em uma economia cada vez mais globalizada e tecnologicamente avançada. Assim, como concluem alguns estudos, o maior obstáculo para a expansão da escolaridade entre a população mais pobre é a qualidade do ensino ofertado (NEY; SOUZA; PONCIANO, 2010; PESSADOR; LOPES, 2014).

O Ensino Médio, objeto de estudo deste artigo, por ser o ciclo final da Educação Básica no Brasil, representa uma das etapas de ensino mais importante para a formação da pessoa. Afinal, a conclusão do Ensino Médio potencializa a obtenção de melhores colocações no mercado de trabalho, além de ser condição necessária para o ingresso no Ensino Superior. (SOUZA *et al.* 2012).

Apesar das várias mudanças ocorridas no Brasil, principalmente em termos da oferta de vagas em escolas públicas, o Ensino Médio não tem evoluído em relação à qualidade do ensino ofertado. De acordo com dados divulgados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), de 2015, o desempenho educacional dos alunos nessa etapa de ensino mantém-se estagnado desde 2011. Pelos

resultados do Índice de Desenvolvimento da Educação Básica (Ideb)¹⁸, na média das escolas do País, o índice foi de 3,7 pontos, sendo abaixo da meta estabelecida para aquele ano, que era de 4,3 pontos¹⁹.

O fracasso escolar nessa etapa de ensino também é confirmado pela taxa de distorção idade-série, abandono e evasão escolar. Em 2015, a distorção idade-série atingia cerca de 27,4% dos alunos matriculados, a taxa de abandono e/ou evasão escolar era de 6,8%, sendo que a situação era mais grave entre as escolas públicas, onde a taxa de distorção era de 30,4% e a de abandono e/ou evasão era de 7,8%.

Vale ressaltar, todavia, o fato de que estes problemas não decorrem exclusivamente do Ensino Médio, pois a distorção idade-série²⁰ é reflexo de reprovação e/ou abandono do aluno ao longo da vida escolar, mas que parecem se intensificar no Ensino Médio. O abandono escolar acontece, principalmente, a partir do 2º ciclo do Ensino Fundamental (do 5º ao 9º ano), sobretudo entre os alunos de famílias pobres.

Segundo Ney *et al.* (2010), enquanto aproximadamente 96% dos filhos de famílias ricas conseguem concluir o Ensino Fundamental, entre os mais pobres, a proporção de formandos é de apenas 61,9%. Ainda segundo os mesmos autores, ao analisarem a desigualdade de acesso à escola e a evasão escolar entre ricos e pobres no Brasil rural e urbano, verificaram que o acesso ao Ensino Médio e Superior é restrito em grande medida a filhos de famílias ricas. A evasão escolar dos filhos de famílias de baixa renda é alta durante todo o Ensino Fundamental, e muitos chegam à maioridade sem sequer ter concluído essa etapa de ensino.

Pelo fato de no meio rural, a evasão escolar dos mais pobres ser mais alta do que em áreas urbanas, também se observa uma proporção maior de jovens brasileiros com baixa escolaridade (NEY; SOUZA; PONCIANO, 2010). Segundo a Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2015, dos adolescentes com idade de

¹⁸ O Ideb é formulado para medir a qualidade do aprendizado nacional e estabelecer metas para a melhoria do ensino. Funciona como um indicador nacional que possibilita o monitoramento da qualidade da Educação e é calculado com base em dois componentes: a taxa de rendimento escolar (aprovação) e as médias de desempenho nos exames aplicados pelo Inep. Os índices de aprovação são obtidos do Censo Escolar, realizado anualmente.

¹⁹ O Ministério da Educação (MEC) projetou metas de qualidade da Educação que devem ser atingidas até 2021. O objetivo é tornar a Educação de fato, eficaz alcançado o nível educacional dos países desenvolvidos.

²⁰ É a proporção de alunos com mais de dois anos de atraso escolar. No Brasil, a criança deve ingressar no 1º ano do Ensino Fundamental e com seis anos de idade, permanecendo no Ensino Fundamental até o 9º ano, com a expectativa de que conclua os estudos nesta modalidade até os 14 anos.

15 a 17 anos (idade equivalente ao Ensino Médio), apenas 46,64%, estavam matriculados no Ensino Médio ou já haviam concluído essa etapa. No meio urbano, a proporção era de 62,84%. Daqueles que não estavam matriculados e não haviam concluído o Ensino Médio, nessa faixa etária e na zona rural, 30,3% estavam fora da escola e 69,7% frequentavam o Ensino Fundamental, ou seja, apresentavam distorção idade-série. Dos adolescentes que se encontravam fora da escola, apenas 5,2% haviam completado o Ensino Fundamental.

Embora a Educação seja reconhecida como importante mecanismo de desenvolvimento econômico e na formação do capital humano, o sistema educacional brasileiro ainda tem produzido disparidades entre a população residente nas zonas rural e urbana. A oferta e a qualidade do ensino no meio rural são geralmente inferiores ao que é oferecido nas áreas urbanas (HELFAND; PEREREIRA, 2012).

De acordo com informações disponibilizadas pelo Inep, em geral, alunos de escolas urbanas apresentam desempenho superior aos que frequentam escolas no meio rural. Pelos dados divulgados pelo Exame Nacional do Ensino Médio (Enem) de 2014, os alunos que frequentavam escolas em áreas urbanas tiveram nota média de 503,89 pontos, enquanto os estudantes de escolas rurais exibiram desempenho médio de 479,78 pontos.

Além disso, são poucas as escolas que ofertam o Ensino Médio em áreas rurais no Brasil. Segundo informações do Censo Escolar de 2015, apenas 10% das escolas que ofertavam Ensino Médio estavam localizadas no meio rural, sendo comum o deslocamento de alunos da área rural para escolas urbanas, obrigando-os a percorrer diariamente grandes distâncias para chegar até a escola.

Perante o quadro exposto e a escassez na literatura brasileira sobre estudos que analisam o efeito da localização da escola e a origem dos alunos (rural e urbano) sobre o desempenho escolar, este estudo tem por objetivo investigar com detalhes como o desempenho de estudantes que residem no meio rural pode ser influenciado pela escola onde estudam, seja na própria zona rural ou quando são transferidos para instituições na zona urbana. Portanto, o estudo tem dois objetivos principais. O primeiro é decompor o diferencial de desempenho escolar entre alunos da zona rural, um estudando em escolas rurais e o outro em escolas urbanas, que será chamado de **efeito-escola**. O segundo é encontrar o **efeito-família**, nesse caso, a decomposição é realizada

entre alunos das zonas rural e urbana, ambos estudando em escolas urbanas.

Os objetivos específicos consistem em verificar se a diferença entre o desempenho dos alunos que estudam em escolas rurais e urbanas decorre das características dos alunos e de seus pais (denominado aqui de **efeito-família**) ou se é devido a características das escolas como melhor infraestrutura e gestão escolar (**efeito-escola**). A ideia é analisar se, após o controle do *background* familiar, o diferencial de desempenho entre estudantes dos meios rural e urbano se altera, passando a ser explicado por diferenças entre as escolas, ou seja, pelo **efeito-escola** tendo como comparação alunos com características similares. Assim também, é possível averiguar se, após o controle das características da escola, as diferenças entre os grupos diminuem.

Para isso, a análise empírica do presente trabalho combina dois métodos: o primeiro consiste em um método de pareamento por meio das características observáveis entre alunos de diferentes escolas. Na literatura, existem diversas metodologias de pareamento para o controle de características observáveis das pessoas. Nessa pesquisa é utilizado o algoritmo *Coarsened Exact Matching* (CEM). O segundo método realiza uma análise de decomposição do diferencial de desempenho entre alunos de escolas rurais e urbanas, por meio da aplicação do método proposto por Oaxaca (1973) e Blinder (1973). O método de decomposição de Oaxaca-Blinder permite identificar as variáveis que mais contribuem para explicar as diferenças de resultados educacionais entre dois grupos. As informações utilizadas são dos alunos avaliados no Enem de 2014²¹, de todo o Brasil, considerando o desempenho em quatro áreas de conhecimento e habilidades e na média geral do exame.

Assim, o estudo tem duas principais contribuições para a literatura. A primeira é a própria abordagem do tema, pois não foram encontrados estudos mensurando a interferência da escola sobre o desempenho escolar dos alunos do meio rural que estudam em escolas urbanas. Em geral, os estudos analisam apenas o desempenho escolar dos alunos de escolas rurais e urbanas isoladamente, sem considerar a origem ou o local de moradia do aluno. E a segunda, é pela utilização do método CEM combinado com o método de decomposição de Oaxaca-Blinder, para

²¹ O questionário do Enem de 2015 não disponibiliza informações sobre a zona de residência da família do aluno.

controlar as características observáveis dos alunos, da família e da escola, a fim de garantir um equilíbrio ou homogeneidade entre os dois grupos.

Este capítulo da dissertação divide-se em seis seções, incluindo esta introdução. Na seção seguinte, é analisado o fluxo dos estudantes do meio rural aprovados no 9º ano ou 8ª série do Ensino Fundamental de 2011, ao longo do Ensino Médio. Na seção três são mostradas a base de dados e as variáveis utilizadas nas regressões. Em seguida, descreve-se a estratégia empírica e a seleção da amostra. Posteriormente, são feitas as análises descritivas dos dados e a discussão dos resultados estimados. Logo em seguida, são feitas as considerações finais do estudo.

2 TRAJETÓRIA DOS ESTUDANTES DA ZONA RURAL DIPLOMADOS NO ENSINO FUNDAMENTAL

A realidade vivida pela maioria dos alunos do meio rural que iniciam seus estudos em escolas rurais é, depois da conclusão do Ensino Fundamental, ter como opção se deslocar para escolas na própria zona rural que ofertam Ensino Médio ou serem transferidos para escolas urbanas. Nem sempre eles terão essas duas escolhas, pois, além de as escolas rurais que ofertam o Ensino Médio serem poucas para suprir a demanda, geralmente, são espacialmente dispersas e ficam distantes das suas residências. Além do mais, para o Poder Público, pode ser mais econômico, em termos de transporte escolar, levá-los à sede dos respectivos municípios.

Essa transição implica uma nova escola, outros professores, novos colegas e disciplinas diferentes, considerando que, no Ensino Médio, o estudante depara com conteúdo mais complexo, de modo que essa mudança pode implicar um ambiente nada parecido com a realidade à qual o aluno estava acostumado: antes na zona rural, geralmente na própria comunidade de residência ou na vizinha, agora em um contexto totalmente distante, a zona urbana. E a adaptação nem sempre é fácil, o que resulta em reprovações, abandono e evasão escolar.

De acordo com o estudo de Silva e Kichow (2013), que analisou a transição de crianças do 5º ano do Ensino Fundamental da zona rural para escolas urbanas, observaram que a mudança de escola resulta em muitas reprovações no 6º ano. E, além disso, entre os alunos do meio rural transferidos para escolas urbanas, se encontravam os maiores índices de evasão. Eles destacam que o rendimento dos estudantes residentes na zona rural era afetado pela rotina cansativa no deslocamento até a escola urbana. Aqueles alunos que estavam tomados pelo entusiasmo e expectativas no início do ano letivo, no transcorrer dos dias, eram tomados por uma sensação de desânimo e cansaço, que culminava com a reprovação. Isso fazia com que, no ano seguinte muitos deles não voltassem mais à escola.

Outro agravante observado pelos autores era a rotina diária a que eles eram submetidos. Muitos precisavam trabalhar nas fazendas onde moravam e, assim, mostravam sempre mais cansados em comparação com os que residiam na zona urbana. O terceiro fator de desgaste diz respeito ao transporte escolar, pois, como as distâncias eram relativamente grandes entre a escola e as fazendas, o aluno saía em horário muito

cedo de sua casa e retornava, geralmente, tarde da noite (SILVA E KICHOW, 2013).

Outro problema relacionado ao descolamento dos alunos para escolas urbanas é a qualidade do transporte escolar. Como destaca um estudo realizado pela UNICEF (2009), nem sempre o transporte escolar dos alunos residentes no meio rural é feito em condições adequadas de segurança ou atende completamente a demanda. As estradas rurais são precárias e não oferecem condições de acesso a todas as comunidades, excluindo aqueles estudantes que vivem em localidades de acesso difícil.

Portanto, para melhor compreensão do fluxo escolar dos alunos que concluem o Ensino Fundamental na zona rural, é analisada nesta seção a trajetória dos estudantes que concluíram a 8ª série ou 9ª ano desse nível em 2011, o ingresso no Ensino Médio e quais foram os resultados alcançados por eles nos próximos três anos. No Quadro 2 há informações obtidas no Censo Escolar de 2011 a 2015. É importante informar que a base de dados foi constituída com suporte no código do aluno, fornecido no Censo Escolar de 2011. Pelo código do aluno, é possível conhecer a trajetória escolar dos estudantes matriculados no Ensino Médio, desde que ele tenha continuado a frequentar a escola.

Quadro 2 - Mapa dos dados coletados no Censo Escolar de 2011 a 2015.

2011	Alunos matriculados na 8ª série ou 9º ano do EF residentes na zona rural e estudando em escolas rurais.
	<i>Objetivo:</i> identificar e analisar o perfil dos alunos que estavam no último ano do ensino fundamental no meio rural.
2012	Cruzamento de informações pelo código do aluno capturado no Censo Escolar de 2011 com o Censo Escolar de 2012.
	<i>Objetivo:</i> identificar os alunos que foram aprovados e deram continuidade aos estudos no EM ou etapa equivalente, os que reprovaram e aqueles que evadiram da escola de 2011 para 2012.
2013	Cruzamento de informações pelo código do aluno capturado no Censo Escolar de 2011 com o Censo Escolar de 2013.
	<i>Objetivo:</i> identificar os alunos que foram aprovados e reprovados na 1ª série do EM, bem como os que abandonaram a escola de 2012 para 2013.
2014	Cruzamento de informações pelo código do aluno capturado no Censo Escolar de 2011 com o Censo Escolar de 2014.
	<i>Objetivo:</i> identificar os alunos que foram aprovados para a 3ª série e os reprovados na 1ª série ou 2ª série do EM, além daqueles que abandonaram a escola de 2013 para 2014.
2015	Cruzamento de informações pelo código do aluno capturado no Censo Escolar de 2011 com o Censo Escolar de 2015.
	<i>Objetivo:</i> identificar os alunos que reprovaram na 1ª série, 2ª série ou 3ª série do EM e aqueles que abandonaram a escola de 2014 para 2015 que estavam matriculados na 1ª e 2ª série do EM.

Fonte: RODRIGUES (2017).

A primeira análise é feita para os alunos matriculados na 8ª série ou 9º ano do EF, de 2011. A base de dados, após o filtro para alunos residentes e que estudavam

em escolas do meio rural foi, de 264.922 estudantes. A idade média desses alunos era de 15,79 anos, com desvio-padrão de 3,23 anos. A cada 100 alunos, 52 eram meninas e a maioria, estudava em escolas da rede municipal (79,13%) e 89,2% em turmas no período diurno (ver Tabela A.2.1 no Apêndice 2).

A Tabela 3 mostra o fluxo escolar dos alunos matriculados em 2011, a partir do cruzamento com informações do Censo Escolar de 2012. Observa-se que 72,58% foram aprovados e se matricularam em turmas equivalentes ao Ensino Médio, 7% foram reprovados e, repetiram a mesma série/ano em 2012 ou se matricularam em turmas de Educação de Jovens e Adultos (EJA) do Ensino Fundamental, enquanto 20,4% se evadiram da escola. Dos alunos evadidos, pelas informações disponibilizadas pelo Inep, não é possível saber se eles foram alvos de reprovação em 2011 e não retornaram à escola em 2012 ou foram aprovados, concluído o Ensino Fundamental, mas não deram continuidade aos estudos no Ensino Médio no ano seguinte.

Tabela 3 - Fluxo escolar dos alunos matriculados no Ensino Fundamental, em 2011

Fluxo escolar			
Total	Aprovados ¹	Reprovados ²	Evasão
	192.278	18.589	54.055
%	72,58%	7,02%	20,40%

Fonte: RODRIGUES (2017) com base nos dados do Censo Escolar de 2011 e 2012.

Nota: ¹ alunos matriculados em 2012 em turmas equivalentes ao Ensino Médio. ² alunos matriculados em 2012 na 8ª série ou 9º ano do Ensino Fundamental ou em turmas equivalentes

Dos alunos que ingressaram no Ensino Médio em 2012, 77,53% trocaram de escola, sendo que, destes 81,60% foram para escolas localizadas no meio urbano. Outra informação interessante é que 29,46% dos estudantes do meio rural do Ensino Médio frequentavam aulas noturnas, enquanto que em 2011, quando estavam no Ensino Fundamental, essa proporção era de 10,8%. Essa condição precisa ser mais bem avaliada, pois, como observam Gonçalves, Passos e Passos (2005), de modo geral, o Ensino Médio noturno apresenta características singulares, pois tem por clientela alunos esgotados, que, na sua maioria, chegam à escola depois de uma longa jornada de trabalho. É um aluno que já vem à escola reprovado pelo cansaço e que apresenta maior

probabilidade de abandonar ou evadir²².

Tabela 4 - Informações dos alunos ingressos no ensino médio (EM), em 2012

Total	Trocaram de escola	Zona da Escola		Turno	
		Rural	Urbana	Diurno	Noite
	163.481	30.073	133.408	115.317	48.164
%	77,53%	18,40%	81,60%	70,54%	29,46%

Fonte: RODRIGUES (2017) com base nos dados do Censo Escolar de 2011 e 2012

Feitas as primeiras análises para encontrar o fluxo escolar dos alunos no Ensino Médio, é necessário fazer novos filtros para a base de dados de 2012, isso porque nem todos os alunos aprovados em 2011, que deram continuidade aos estudos, se matricularam no Ensino Médio regular ou seriado. Do total, aproximadamente 3,8% da amostra frequentavam turmas com ensino equivalente ao EM como: o Ensino Médio não-seriado, EJA e Educação Profissional (Concomitante e Subsequente). Os demais estavam matriculados em turmas do EM regular (91,8%), ensino médio integrado (3,5%) e ensino médio - Normal/Magistério (1,0%), resultando em uma amostra de 185.032 alunos. A tabela a seguir mostra o fluxo escolar desses alunos em 2012.

Tabela 5 - Fluxo escolar dos alunos matriculados na 1ª série do EM, em 2012.

Fluxo escolar			
Total	Aprovados ¹	Reprovados ²	Evasão ³
	141.234	20.059	23.739
%	76,33%	10,84%	12,83%

Fonte: RODRIGUES (2017) com base nos dados do Censo Escolar de 2011 e 2012.

Nota: ¹ alunos matriculados em 2013 na 2ª série do EM. ² alunos matriculados em 2012 e 2013 na 1ª série do EM. ³ Aluno não matriculado em 2013.

Observa-se que houve uma taxa de reprovação superior à de 2011. De cada 100 alunos que se matricularam em 2012, cerca de 11 estudantes ficaram reprovados na primeira série do Ensino Médio. E a taxa de evasão foi de 12,83%. Fazendo uma análise mais minuciosa sobre os alunos reprovados e evadidos, notou-se que a evasão escolar entre os alunos do período noturno é superior à dos alunos matriculados no turno diurno, enquanto a taxa de abandono entre os escolares de turmas da manhã e tarde foi de 11,55%. No período noturno, essa taxa foi de 15,95%. Por outro lado, a taxa de reprovação é maior entre alunos que estudavam no período diurno.

²² Embora essa variável não seja utilizada nas estimativas dos modelos utilizados neste estudo, pois o questionário do Enem não informa o turno no qual o aluno estuda, ela pode influenciar diretamente nos resultados obtidos pelos alunos que estudam no período noturno.

Em relação ao gênero do aluno, a proporção de alunos evadidos e reprovados em 2011 foi maior entre os meninos, uma vez que 13,62% dos garotos não passaram de ano, contra 8,45% de reprovação entre as meninas. E a evasão escolar entre os meninos foi de 14,10% e entre as meninas de 11,73%. Os alunos da zona rural que estudavam em escolas urbanas também apresentaram pior desempenho em relação à aprovação na 1ª série do EM, sendo também maior a proporção de alunos que abandonaram a escola (Ver Tabela A.2.2 no Apêndice 2).

Na Tabela 6, é analisado o fluxo escolar dos alunos matriculados nas 1ª e 2ª séries do EM, em 2013, ou seja, dos alunos reprovados e aprovados em 2012, respectivamente. Dos alunos reprovados em 2012, 49,50% foram aprovados em 2013, 20,73% não conseguiram avançar, sendo reprovados novamente e 29,77% abandonaram a escola.

Em relação aos alunos matriculados na 2ª série, em 2013, 84,57% foram aprovados, 5,77% reprovados e 9,66% abandonaram a escola. Do perfil dos alunos evadidos em 2013, mais uma vez, a proporção foi maior entre os alunos matriculados no período noturno (15,89% contra 10,32% do diurno) e entre pessoas do sexo masculino (14,04% contra 10,60% das meninas). As chances de um aluno da zona rural que estuda em escolas urbanas ser reprovado e abandonar é maior do que daqueles que estudam no próprio meio rural (Ver Tabela A.2.2 no Apêndice 2).

Tabela 6 - Fluxo escolar dos alunos matriculados na 1ª e 2ª série EM, em 2013.

Fluxo escolar – 1ª série			
Total	Aprovados ¹	Reprovados ²	Evasão ³
	9.929	4.159	5.971
%	49,50%	20,73%	29,77%
Fluxo escolar – 2ª série			
Total	Aprovados ⁴	Reprovados ⁵	Evasão ⁶
	119.444	8.144	13.646
%	84,57	5,77%	9,66%

Fonte: RODRIGUES (2017) com base nos dados do Censo Escolar de 2011, 2012, 2013 e 2014

Nota: ¹ alunos matriculados na 1ª série do EM em 2012 e 2013 e na 2ª série em 2014. ² alunos matriculados em 2013 e 2014 na 2ª série do EM. ³ Aluno não matriculado em 2014. ⁴ alunos matriculados na 2ª série em 2013 e matriculados na 3ª série do EM, em 2014. ⁵ alunos matriculados em 2013 e 2014 na 2ª série do EM. ⁶ Aluno não matriculado em 2014.

O fluxo escolar dos alunos matriculados em 2014 é exibido na Tabela 7. Nota-se uma evasão expressiva entre os alunos que reprovados na 1ª série, tanto em 2012 como em 2013 - 45,90% deixaram a escola após duas reprovações seguidas. Os

demais, 35,32%, conseguiram avançar para a 2ª série em 2015 e 18,78% foram reprovados novamente. Dos alunos que cursaram a 2ª série em 2014, 60,74% foram aprovados, 12,80% reprovados e 26,46% se evadiram. Já para os alunos matriculados na 3ª série, não é possível, pelos microdados do Censo Escolar de 2015, saber quantos foram aprovados ou se evadiram, só é possível tirar conclusões sobre a taxa de reprovação, ou seja, a proporção de alunos que estavam matriculados em 2014 e 2015 na 3ª série do EM. Dos alunos que cursaram o 3º ano em 2014, 4,91% foram reprovados.

Tabela 7 - Fluxo escolar dos alunos matriculados na 1ª, 2ª e 3ª série do EM, em 2014

Fluxo escolar – 1ª série			
Total	Aprovados¹	Reprovados²	Evasão³
	1.469	781	1.909
%	35,32%	18,78%	45,90%
Fluxo escolar – 2ª série			
Total	Aprovados⁴	Reprovados⁵	Evasão⁶
	6.031	1.271	2.627
%	60,74%	12,80%	26,46%
Fluxo escolar – 3ª série			
Total	Aprovados	Reprovados⁷	Evasão
	-	5.868	-
%	-	4,91%	-

Fonte: RODRIGUES (2017) com base nos dados do Censo Escolar de 2011, 2012, 2013, 2014 e 2015

Nota: ¹ alunos matriculados na 1ª série em 2014 e na 2ª série em 2015. ² alunos matriculados em 2014 e 2015 na 1ª série do EM. ³ Aluno reprovados em 2013 e não matriculado em 2015. ⁴ alunos matriculados na 2ª série em 2014 e matriculados na 3ª série do EM em 2015. ⁵ alunos matriculados em 2014 e 2015 na 2ª série do EM. ⁶ Aluno matriculado em 2014 e 2015 na 3ª série.

Portanto, pelos resultados aqui mostrados, existe uma redução de matrículas dos alunos do meio rural que ingressaram no Ensino Médio em todo o período. A maior taxa de desistência acontece, principalmente, entre os alunos que reprovaram em pelo menos uma série, entre os alunos do período noturno, sendo maior entre os homens (as meninas são mais prováveis de serem aprovadas e de continuarem os estudos) e entre os estudantes que estudavam em escolas urbanas. Resumidamente, de cada 100 alunos residentes do meio rural que ingressaram no Ensino Médio regular em 2012, apenas 77 permaneceram no sistema de ensino sem abandonar a escola no período analisado. Somente 64,6% chegaram a 3ª série sem nenhuma reprovação.

3 BASE DE DADOS E TRATAMENTO DAS VARIÁVEIS

Os dados utilizados neste estudo fazem parte dos microdados do Exame Nacional do Ensino Médio (Enem) de 2014. O Enem foi criado em 1998 com o objetivo de avaliar o desempenho escolar dos estudantes ao final da Educação Básica, ou seja, na terceira série do Ensino Médio. E desde 2004, passou a ser utilizado também como mecanismo de seleção para o ingresso no Ensino Superior.

Em 2014, o Enem disponibilizou informações de mais de 8,7 milhões de participantes, no entanto, o banco de dados desta pesquisa é constituído apenas por alunos que atendem aos seguintes critérios; i) que ainda estavam frequentando a escola da rede pública²³ e privada em atividade no ano do exame (2014), ii) que estavam cursando a última série do Ensino Médio, matriculados em turmas de Ensino Regular²⁴; e iii) que participaram das quatro avaliações. Para os propósitos desta análise, foi selecionada uma amostra total de 1.084.015 alunos, distribuídos entre escolas da zona rural e urbana.

É importante ressaltar o fato de que os alunos que fazem parte da base de dados no Enem não são os mesmos da seção anterior. Claro que é possível alguns deles estejam nessa amostra, mas não há possibilidade de os identificar, pois, o número de inscrição na base de dados do Enem é diferente do código do aluno do Censo Escolar.

Além das variáveis de desempenho escolar classificadas nas quatro áreas de conhecimento – Ciências da Natureza (NOTA_CN), Ciências Humanas (NOTA_CH), Linguagens e Códigos (NOTA_LC), Matemática (NOTA_MT) e a Média Geral (MEDIA_GERAL) das quatro avaliações, o banco de dados inclui também variáveis contextuais dos alunos, tais como: idade, sexo, raça, se exerce atividade remunerada, tempo que levou para concluir o Ensino Fundamental e Ensino Médio, escolaridade dos pais, renda do domicílio, entre outras. Para a contextualização do ambiente escolar, foram incluídas informações sobre o nível socioeconômico dos alunos, média de alunos por turma, índice de regularidade docente, índice de esforço docente, indicador de adequação da formação docente e o índice de infraestrutura escolar. As variáveis utilizadas nos modelos econométricos estão listadas na Tabela 8.

²³ Exceto alunos de escolas federais.

²⁴ Foram excluídos os alunos de turmas de ensino de Jovens e Adultos e Ensino Especial.

Tabela 8 - Descrição das variáveis

Variável	Descrição
Dependentes	
NOTA_CN	Nota de desempenho do aluno na avaliação de Ciências da Natureza
NOTA_CH	Nota de desempenho do aluno na avaliação de Ciências Humanas
NOTA_LC	Nota de desempenho do aluno na avaliação de Linguagens e Códigos
NOTA_MT	Nota de desempenho do aluno na avaliação de Matemática
MEDIA_GERAL	Média Geral das quatro áreas
Covariadas	
Características do Aluno	
Idade1	<i>Dummy</i> indicativa se o aluno tem idade entre 15 e 17 anos
Idade2	<i>Dummy</i> indicativa se o aluno tem idade entre 18 e 25 anos
Idade3	<i>Dummy</i> indicativa se o aluno tem mais de 26 anos
Sexo	<i>Dummy</i> indicativa de gênero do aluno (referência = masculino)
Raça	<i>Dummy</i> indicativa de que o aluno é da cor branca
Estado Civil	Estado civil do aluno: 1 - se é solteiro; 0 - caso contrário (c.c)
Trabalho	Se o aluno exerce trabalho remunerado: 1 - sim; 0 - c.c
Conclusão do EF	Se o aluno concluiu o ensino fundamental em 8 ano: 1 - sim; 0 - c.c
Conclusão do EM	Se o aluno concluiu o ensino médio em 3 ano: 1 - sim; 0 - c.c
Características Familiar	
Educação do Pai	<i>Dummies</i> indicativas de escolaridade do pai (referência = “nunca estudou ou da 1 a 4ª série do EF”)
Educação do Mãe	<i>Dummies</i> indicativas de escolaridade da mãe (referência = “nunca estudou ou da 1 a 4ª série do EF”)
Renda00	Se a família não possui renda: 1 - sim; 0 - c.c
Renda01	Se a família possui até 1 salário-mínimo: 1 - sim; 0 - c.c
Renda02	Se a família possui de 1 até 2 salários-mínimos: 1 - sim; 0 - c.c
Renda03	Se a família possui de 2 até 5 salários-mínimos: 1 - sim; 0 - c.c
Renda04	Se a família possui de mais de 5 salários-mínimos: 1 - sim; 0 - c.c
CEF	Índices de Condição Socioeconômica Familiar
Características da Escola	
Esc_profiss	<i>Dummy</i> indicativa se escola é de ensino profissionalizante
Publico	<i>Dummy</i> indicativa se pertence a rede pública de ensino
INSE	Indicador do Nível Socioeconômico da Escola
Aluno_Turma	Média de alunos por turma no Ensino Médio
IRD	Índice de Regularidade Docente
IED	Índice de Esforço Docente
IAFD	Indicador de Adequação da Formação Docente
Infra Adequada	<i>Dummy</i> indicativa de infraestrutura adequada

Fonte: RODRIGUES (2017).

No que diz respeito às variáveis que caracterizam as famílias, a escolaridade dos pais é representada por um conjunto de *dummies*, cuja referência é o pai e/ou a mãe que nunca estudaram ou que frequentaram o EF da 1ª a 4ª série (antigo Primário). As demais *dummies* são os pais que estudaram da 5ª a 8ª série e não concluíram o Ensino Médio; pais que concluíram o Ensino Médio e com Ensino Superior incompleto e aqueles que possuem Ensino Superior completo ou pós-graduação. Os alunos que responderam “não sei” foram excluídos da base de dados.

Para a renda familiar dos estudantes, foram incluídas quatro *dummies* que capta a renda do aluno entre nenhuma renda e com rendimentos superiores a cinco salários-mínimos. Além disso, foi criado o Índice de Condição Socioeconômica Familiar (CSF)²⁵ que é uma adaptação do Critério de Classificação Econômica Brasil (CCEB) desenvolvido pela Associação Brasileira de Empresas de Pesquisa (ABEP) para o questionário do Enem (a tabela de pontuação pode ser consultada no Apêndice 2).

Para a caracterização do ambiente das escolas, foram incluídas *dummies* sobre tipo de ensino (profissionalizante ou regular) e rede de ensino (público ou privado). Para mensurar o nível socioeconômico dos alunos que estudam na escola, será utilizado o Indicador do Nível Socioeconômico da Escola (INSE), criado pelo Inep, que é mensurado pela posse de bens domésticos, renda e contratação de serviços pela família dos alunos e pelo nível de escolaridade dos pais²⁶ (INEP, 2014a). Também será considerada a média de alunos por turma na escola, considerando apenas as turmas do Ensino Médio.

Para ampliar as possibilidades de comparações, será considerado três indicadores nas escolas relacionadas aos professores o Índice de Regularidade Docente, Índice de Esforço Docente e o Indicador de Adequação da Formação Docente, todos disponibilizados pelo Inep, sendo definidos assim²⁷:

- ✓ **Índice de Regularidade Docente (IRD):** O índice varia de 0 a 5 pontos e mostra a permanência dos docentes nas escolas nos últimos cinco anos (no caso, a referência é 2014). Cada professor recebe uma pontuação que inclui as seguintes variáveis: (1) total de anos que o professor lecionou na unidade de ensino nos últimos 5 anos; (2) atuação do docente na escola em anos mais recentes e (3) atuação em anos consecutivos. O Inep também classifica o IRD em quatro categorias, são elas: Baixa regularidade (IRD médio igual ou menor que 2) Média-baixa (IRD

²⁵ O índice varia entre 0 e 46 pontos e a classificação em cada classe é definida da seguinte forma: A1 (42 – 46 pontos); A2 (35 a 41 pontos), B1 (29 a 34 pontos), B2 (23 a 28 pontos), C1 (18 a 22 pontos), C2 (14 a 17 pontos), D (8 a 13 pontos) e E (0 a 7 pontos).

²⁶ Informações adicionais podem ser consultadas na nota técnica disponível no endereço eletrônico: <http://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2011_2013/nivel_socioeconomico/nota_tecnica_indicador_nivel_socioeconomico.pdf>.

²⁷ É importante ressaltar que os indicadores são para professores apenas do Ensino Médio

médio maior que 2 até 3); Média-alta (IRD médio maior que 3 até 4); Alta (IRD médio maior que 4 até 5). (INEP, 2014b).

- ✓ **O Indicador de Esforço Docente (IED):** classifica o docente em níveis de 1 a 6 de acordo com o esforço empreendido no exercício da profissão. Sendo consideradas as seguintes características do docente: (1) número de escolas em que atua; (2) número de turnos de trabalho; (3) número de alunos atendidos e (4) número de etapas nas quais leciona. As variáveis criadas para representar tais atributos são do tipo ordinal, nas quais as categorias mais elevadas indicam maior esforço por parte do professor (INEP, 2014c). Para este estudo foi selecionado a proporção de professores por escola, classificados nos níveis 4, 5 e 6.
- ✓ **Indicador de Adequação da Formação Docente (IAFD):** este indicador classifica os docentes em exercício na Educação Básica considerando sua formação acadêmica e a (s) disciplina (s) que leciona. Sendo classificados em cinco perfis de regência das disciplinas (INEP, 2014d). Neste estudo foi selecionado a proporção de professores por escola, que lecionava no ensino médio, que se enquadravam na categoria 1 e 2, isto é, docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina com curso de complementação pedagógica concluído.

A variável infraestrutura adequada é uma *dummy* baseada no Plano Nacional de Educação (PNE), que assume valor um se a escola possuir todos os sete itens de uma infraestrutura adequada, que são: acesso à energia elétrica, abastecimento de água tratada, esgotamento sanitário da rede pública, quadra esportiva, laboratório de ciências, biblioteca ou sala de leitura e acesso à internet de banda larga (essas informações foram retiradas do Censo Escolar de 2014).

4 METODOLOGIA E SELEÇÃO DO GRUPO DE TRATAMENTO E CONTROLE

Como explicado anteriormente, o objetivo principal deste trabalho é identificar o efeito das características das famílias e das escolas localizadas em áreas rurais e urbanas sobre o desempenho escolar. Para alcançar esse objetivo, a análise empírica se propõe a decompor o diferencial de desempenho escolar dos alunos que estudam em escolas rurais e urbanas em dois efeitos: **efeito-escola** e **efeito-família**. De modo direto, aplica-se uma combinação de duas técnicas de análise estatística e econométrica. A primeira tem por objetivo fazer o pareamento entre dois grupos, dadas suas características observáveis, conhecido na literatura por algoritmo *Coarsened Exact Matching* (CEM). Já a segunda analisa a decomposição da diferença de notas entre estudantes de escolas rurais e urbanas em duas partes: uma, que é **diferença explicada**, dada pelas características observáveis utilizadas no modelo, e a segunda que é a **diferença não explicada**, que são características não mensuráveis.

4.1 Algoritmo CEM

Um problema associado aos métodos de *matching* baseados no *propensity score*²⁸ é o seu fraco desempenho em encontrar grupos de comparação equilibrados. De fato, o alvo principal dos métodos de *matching* consiste em eliminar observações, de modo a obter melhor equilíbrio comparativo entre grupo tratados e o grupo de controle (IACUS; KING; PORRO, 2011).

Para tentar contornar esses problemas, Iacus, King and Porro (2008) desenvolveram o algoritmo *Coarsened Exact Matching* (CEM), o qual não exige nenhuma hipótese à geração de dados (DGP), com exceção da ignorabilidade. Além disso, tal método garante que os desequilíbrios entre os grupos de tratados e de controle após a realização do *matching* não sejam maiores do que um limite previamente selecionado.

Iacus, King and Porro (2009, 2011) mostraram, por meio de simulações de um grande número de DGP's, as estimativas obtidas com o algoritmo CEM, que

²⁸ Propensity Score Matching (PSM) segundo Dehejia and Wahba (2002).

possuem propriedades estatísticas superiores aos métodos tradicionais, como menor desequilíbrio entre grupos pós-*matching*, menor dependência do modelo, dentre outros. Além disso, o CEM permite a melhora da análise de outros métodos²⁹.

O algoritmo do CEM segue os seguintes passos:

- 1) Dado as covariadas X , é realizado uma cópia de X , denominada X^* .
- 2) X^* é “engrossado” (*coarsened*) ou de acordo com cortes amostrais previamente selecionados ou utilizando um método automático, desenvolvido por Iacus, King and Porro (2008);
- 3) Cria-se um estrato por unidade de observação de X^* e se coloca cada observação no estrato.
- 4) Relaciona-se o estrato com os dados originais, X , e eliminam-se as observações do estrato que não contem ao menos um tratado ou uma unidade de controle.

Dito de forma diferente, após o *coarsened*, o algoritmo CEM cria um conjunto de estratos, isso é, $s \in S$, cada um com os mesmos valores atribuíveis de X . As unidades nos estratos que contêm pelo menos uma unidade de tratamento e uma unidade de controle são mantidas.

O método CEM também produz pesos que podem ser utilizados na avaliação de medidas de desequilíbrio e estimativas do efeito causal. Denotado por T^s , as unidades tratadas no estrato s e por $m_T^s = \#T^s$ o número de unidade tratadas no estrato, similarmente para as unidades de controle, tem-se C^s e $m_C^s = \#C^s$. Desse modo, o número de unidades correspondentes é, respectivamente, para os tratados e controles $m_T = \bigcup_{s \in S} m_T^s$ e $m_C = \bigcup_{s \in S} m_C^s$. Assim, para cada unidade correspondente i no estrato s , o algoritmo CEM atribui os seguintes pesos:

$$w_i = \begin{cases} 1, & i \in T^s \\ \frac{m_C}{m_T} \frac{m_T^s}{m_C^s}, & i \in C^s \end{cases} \quad (1).$$

As unidades sem pareamento recebem peso $w_i = 0$, sendo excluídas da base de dados (IACUS; KING; PORRO, 2008).

Segundo Blackwell *et al.* (2009), o peso gerado pelo CEM pode ser aplicado

²⁹ Ver estudos desenvolvidos por Aroca Brida e Volo (2014), Datta (2015) e Schurer et al (2015).

em qualquer modelo econométrico para controlar o desequilíbrio remanescente entre o grupo de controle e tratamento, desde que seja possível incorporá-lo na regressão, tais como: MQO, *logit*, *probit* e *etc.* Como o método original desenvolvido Oaxaca (1973) e Blinder (1973) permite que pesos sejam utilizados, os dois métodos podem ser combinados, gerando, assim, estatísticas mais robustas. A combinação entre os dois modelos permite que os grupos sejam mais homogêneos em termos de características observáveis. Assim, será possível encontrar o diferencial de rendimento dos alunos que é decorrente do **efeito-família** e do **efeito-escola**.

4.2 Funções de produção educacional e decomposição de Oaxaca-Blinder (1973)

Tendo como referência os estudos consolidados na literatura sobre os fatores associados ao rendimento escolar dos estudantes, as diferenças no desempenho escolar entre alunos do Ensino Médio que estudam em escolas no meio rural e em áreas urbanas são estimadas por uma Função de Produção Educacional (FPE), que relaciona ao indicador de desempenho uma série de “inputs”, que incluindo as características observáveis e não observáveis dos alunos, de seus familiares e da escola que frequentam. Cada característica considerada possui determinada taxa de retorno ou efeito marginal sobre os resultados de interesse, isso é, nos níveis de aprendizado alcançado pelos estudantes (RODRIGUES, 2009). A FPE pode ser representada pela seguinte equação:

$$D = f(X, F, E) + \varepsilon \quad (2)$$

Onde, D representa uma variável de resultado, nesse caso, é o desempenho ou nível de habilidade acadêmica adquirida pelo aluno, que está em função de um conjunto de fatores agrupados em três categorias: características individuais e motivação aos estudos (X), características familiares (F), características da escola onde estuda (E) e o termo do erro (ε), que representa todas as variáveis ou características não observáveis que impactam no desempenho escolar do estudante que não foram capturadas pelo conjunto de informações disponíveis³⁰. A equação (1) pode ser expressa linearmente como:

³⁰ Mais informações sobre a Função de Produção Educacional podem ser obtidas nos trabalhos de Hanushek e Woessmann (2011, 2012) e Wobmann (2003)

$$D_{ijk} = \beta_0 + \beta_1 X_{ij} + \beta_2 F_{ij} + \beta_3 E_{ij} + \beta_4 P_{ij} + \varepsilon_i \quad (3)$$

onde, $i = 1, \dots, n$, $j = 1 \dots J$ $k = r, u$

Em que, D_{ijk} é a nota do aluno i , na escola j e na zona k . E β_0 a β_4 são os parâmetros a serem estimados.

Com base na FPE de cada grupo, é aplicada a metodologia de decomposição de Oaxaca (1973) e Blinder (1973) - OB com o objetivo de explorar os fatores associados às diferenças nos resultados educacionais entre os grupos de alunos que estudam em escolas rurais e urbanas.

A técnica de decomposição OB permite identificar as causas que dão origem às diferenças no desempenho escolar dos estudantes por meio de dois componentes: uma parte explicada pelas características observadas, obtidas com base nas características dos indivíduos de cada grupo, e outra não explicada, onde são incluídos todos os fatores que não são diretamente mensuráveis ou que foram omitidos do modelo. Portanto, a técnica de decomposição de OB para o rendimento escolar entre alunos do meio rural e urbano pode ser expressa como:

$$D_Y - D_Z = [\hat{\beta}_Y(\bar{X}_Y - \bar{X}_Z)] + [\bar{X}_Z(\hat{\beta}_Y - \hat{\beta}_Z)] \quad (4)$$

Nesta expressão, o primeiro e o segundo termos do lado direito representam os componentes explicados e não explicados da diferença nos resultados médios, respectivamente. E Y e Z correspondem aos grupos analisados.

4.3 Estratégia e seleção do grupo de tratamento e controle

Considerando que o principal objetivo da pesquisa é encontrar o diferencial de rendimento escolar entre alunos que residem e frequentam escolas no meio rural, com aqueles que moram no meio rural, mas estudam em escolas urbanas, e entre os residentes da zona rural e os que moram na zona urbana, estudando em escolas urbanas. O estudo é dividido em duas análises independentes. A primeira, realizada apenas para aqueles residentes no meio rural, aqui chamados de Grupo 1 (Alunos de escolas rurais vs alunos da zona rural que estudam em escolas urbanas) e a segunda, para o grupo denominado de Grupo 2 (Alunos da zona rural que estudam em escolas urbanas vs alunos urbanos).

A descrição dos grupos pode ser visualizada no quadro seguinte, onde a amostra para o Grupo 1 é formada por 155.360 alunos, sendo que 24.397 estudavam em

escolas rurais (definido como o grupo de tratamento) e 130.963 da zona rural que estudavam em escolas urbanas (o grupo de controle.) A segunda restringiu-se somente àqueles que estudavam em escolas urbanas, sendo que o grupo de tratamento é o dos estudantes localizados na zona rural (130.963) e o grupo de controle os que residiam na zona urbana (928.655 alunos).

Quadro 3 - Descrição dos grupos de tratamento e controle.

Grupos	Descrição	Tratados	Controle
Grupo 1	Alunos de escolas rurais vs alunos da zona rural que estudam em escolas urbanas	Alunos das escolas rurais	Alunos das escolas urbanas
Grupo 2	Alunos estudantes de escolas urbanas	Alunos residentes na zona rural	Alunos residentes na zona urbana

Fonte: RODRIGUES (2017).

O método CEM foi aplicado a fim de permitir que o equilíbrio empírico entre os alunos tratados e não tratados seja escolhido *ex-ante*, de modo a reduzir monotonicamente o desequilíbrio da distribuição entre os grupos, conforme descrito por Blackwell *et al.* (2009). Dessa maneira, foram selecionadas variáveis que, segundo a literatura, influenciam no desempenho escolar do aluno.

Para o Grupo 1, as variáveis selecionadas no pareamento foram estabelecidas de acordo com as características individuais e familiares do aluno, de modo que a seleção da amostra resultante não tenha desequilíbrio/diferenças em razão dessas características. Este tipo de pareamento permite que o modelo isole os efeitos na nota que são resultantes das características observáveis, individuais e familiares, do aluno, de maneira que o pareamento seja feito apenas entre aqueles que possuem características semelhantes; com isso, pela aplicação do modelo Oaxaca-Blinder (1973), com a ponderação dos pesos gerados pelo CEM, é possível encontrar o diferencial da nota entre os grupos que são decorrentes do **efeito-escola**. A aplicação do método de Oaxaca e Blinder permite inclusive a decomposição detalhada, ajudando na identificação da contribuição de subconjuntos das variáveis explicativas, permitindo estimar o efeito que pode ser atribuído às escolas que frequentam (no meio rural e urbano).

Para o Grupo 2, o pareamento entre os alunos restringiu-se às variáveis indicativas das escolas, pois o objetivo é encontrar as diferenças de rendimento escolar decorrente do **efeito-família**. Nesse caso, deduz-se que os alunos residentes do meio

rural que estudam em escolas urbanas frequentam as mesmas escolas dos estudantes do meio urbano, ou seja, o desempenho entre eles não pode ser diferenciado pela escola, pois ambos estão expostos aos mesmos fatores escolares. Isso porque, o que é ofertado para o aluno do meio rural é também ofertado aos estudantes urbanos. Uma vez realizado o controle, a diferença de rendimento escolar entre estudantes do meio rural e urbano, dada pela decomposição de Oaxaca-Blinder com ponderação, indica o quanto do desempenho do aluno, em número de pontos, é em função do **efeito-família**.

5 RESULTADOS

Nesta etapa do estudo, faz-se uma análise descritiva dos dados e apresentam-se os resultados dos modelos estimados.

5.1 Análises descritivas

Ao analisar as estatísticas descritivas das variáveis utilizadas neste estudo, é possível conhecer o perfil dos estudantes pertencentes aos Grupos 1 e 2. A Tabela 9 exibe as médias e os desvios-padrões, levando-se em consideração a amostra total. Para simplificação do texto, alunos do meio urbano que estudavam em escolas urbanas serão chamados de **Urbano-Urbano**; os residentes na zona rural que frequentavam escolas urbanas serão chamados de **Rural-Urbano**; e os de escolas rurais serão denominados de **Rurais-Rurais**.

De acordo com os dados, observa-se que, tanto o desempenho dos alunos avaliados nas quatro áreas de conhecimento e habilidades (Ciências da Natureza, Ciências Humanas, Linguagem de Códigos e Matemática), como na Média Geral do Enem, as melhores notas são dos alunos Urbanos-Urbanos, seguidas pelas notas obtidas pelos alunos Rurais-Urbanos. A diferença na MEDIA_GERAL é de mais de 38 e 35 pontos entre alunos do meio Urbano-Urbano em relação às notas obtidas pelos estudantes Rural-Rural e Rural-Urbano, respectivamente.

Em relação à idade, a proporção de alunos em idade escolar para a frequência do Ensino Médio (de 15 a 17 anos de idade)³¹ é menor entre os estudantes no grupo Rural-Rural - 79%. Já os estudantes do Rural-Urbano, aproximadamente 81% se encontravam na idade adequada para a conclusão do Ensino Médio, enquanto aos alunos do Urbano-Urbano a proporção era de 87%.

³¹ De acordo com o sistema educacional brasileiro, um aluno que inicia os estudos aos seis anos de idade e sem nenhum atraso escolar, chegaria, normalmente, ao Ensino Médio com 15 anos de idade, no entanto, como alguns podem estar adiantados, considerou-se a faixa de 15 a 17 anos adequado à conclusão do ensino médio.

Tabela 9 - Estatística descritiva Média e Desvio-Padrão

Variáveis	Alunos de escolas Urbanas residentes em na Zona Urbana (Urbano-Urbano)		Alunos de escolas Urbanas residentes em na Zona Rural (Rural-Urbano)		Alunos de escolas Rurais (Rural-Rural)	
	Média	Desvio-padrão	Média	Desvio-padrão	Média	Desvio-padrão
NOTA_CN	492,71	75,64	464,96	62,93	462,86	61,34
NOTA_CH	553,13	72,41	517,92	69,31	516,72	67,65
NOTA_LC	518,11	68,97	482,94	66,91	478,28	67,3
NOTA_MT	488,67	112,17	445,51	87,33	441,82	83,34
MEDIA_GERAL	513,16	70,31	477,83	57,75	474,92	55,53
<i>Características do Aluno</i>						
Idade1	87,9%	32,7%	81,0%	39,2%	79,4%	40,4%
Idade2	11,4%	31,8%	17,6%	38,1%	18,2%	38,6%
Idade3	0,7%	8,5%	1,3%	11,5%	2,3%	15,0%
Sexo	41,8%	49,3%	37,3%	48,4%	41,4%	49,3%
Raça	47,0%	49,9%	37,9%	48,5%	36,7%	48,2%
Estado Civil	98,9%	10,5%	98,2%	13,4%	97,4%	15,9%
Trabalho	19,6%	39,7%	13,2%	33,8%	10,2%	30,3%
Conclusão do EF	62,9%	48,3%	55,9%	49,7%	58,7%	49,2%
Conclusão do EM	46,5%	49,9%	45,9%	49,8%	48,8%	50,0%
<i>Características Familiares</i>						
Edu_pai1	24,6%	43,1%	59,2%	49,2%	55,8%	49,7%
Edu_pai2	27,7%	44,8%	23,6%	42,5%	24,6%	43,1%
Edu_pai3	31,9%	46,6%	13,5%	34,2%	15,4%	36,1%
Edu_pai4	15,7%	36,4%	3,7%	18,9%	4,2%	20,0%
Edu_mae1	30,4%	46,0%	66,5%	47,2%	62,8%	48,3%
Edu_mae2	42,7%	49,5%	40,6%	49,1%	40,6%	49,1%
Edu_mae3	50,5%	50,0%	24,9%	43,2%	28,6%	45,2%
Edu_mae4	27,2%	44,5%	9,3%	29,1%	10,9%	31,2%
Renda00	1,2%	10,7%	4,3%	20,3%	3,8%	19,1%
Renda01	23,3%	42,2%	51,4%	50,0%	50,1%	50,0%
Renda02	30,6%	46,1%	26,7%	44,2%	28,3%	45,0%
Renda03	28,2%	45,0%	13,8%	34,5%	14,2%	34,9%
Renda04	16,7%	37,3%	3,9%	19,3%	3,6%	18,7%
CSF	22,42	7,26	17,68	6,37	17,60	6,21
<i>Características da Escola</i>						
Esc_profiss	11,2%	31,5%	8,1%	27,3%	10,1%	30,2%
Publico	74,7%	43,5%	94,2%	23,3%	92,9%	25,7%
INSE	52,314	7,747	45,572	7,01	43,828	6,486
Aluno_Turma	33,51	7,801	31,245	6,791	27,215	7,995
IRD	3,286	0,55	3,264	0,514	3,12	0,628
IED	67,9%	18,7%	69,0%	18,7%	72,0%	23,4%
IAFD	63,2%	16,5%	54,9%	17,7%	50,7%	19,7%
Infra Adequada	75,5%	43,0%	60,9%	48,8%	44,4%	49,7%
Total de Obs.	928.655		130.963		24.397	

Fonte: RODRIGUES (2017) com base nos dados do Enem e Censo Escolar, 2014.

Quanto à presença de alunos dos sexos masculino e feminino entre os alunos Urbano-Urbano, em comparação com os Rural-Rural, quase não há diferença, sendo a participação de aproximadamente 41% de meninos e 59% de meninas. A proporção de meninos do meio rural, no entanto, que frequentava escolas urbanas, é

menor, sendo 37% garotos e de 63% garotas. Assim como foi mostrado na seção 2, a taxa de abandono é sempre maior entre os meninos, sendo ainda maior quando eles estudam em escolas urbanas.

Assim como explicam Artes e Carvalho (2010), isso ocorre porque os meninos não apenas entrariam no mundo de trabalho com maior frequência e mais cedo do que as meninas, mas também porque cumpririam tarefas³² que, muitas vezes, impossibilitariam a frequência escolar.

Quanto à raça, a proporção de alunos que se determinam da cor branca é maior entre o grupo Urbano-Urbano (47%), enquanto, nos casos dos estudantes do Rural-Urbano e Rural-Rural, a participação é de 37% e 36%, respectivamente. Em relação ao estado civil, as estatísticas se comportaram de modo semelhante entre os grupos, pois em torno de 98% dos estudantes se declararam ser solteiros. A variável, porém, referente à situação de atividade do aluno é maior no grupo Urbano-Urbano, visto que 19,6% responderam que estavam trabalhando contra 13,2% do grupo Rural-Urbano e 10,2% do Rural-Rural. No entanto, essa variável pode estar subestimando a proporção de alunos no meio rural que trabalham, já que a pergunta realizada pelo questionário do Enem é sobre trabalho remunerado, e sabe-se que grande parte dos jovens do meio rural exerce atividades não remuneradas, muitas das vezes relacionadas à agricultura familiar.

A proporção de alunos que concluiu o Ensino Fundamental em oito anos é maior entre os alunos do Urbano-Urbano. Já no ensino médio, os estudantes do Rural-Rural expressaram melhor desempenho - cerca de 48,8% concluiriam o Ensino Médio em três anos.

Dos alunos analisados da área urbana, aproximadamente 24% têm o pai e 30% a mãe com o menor nível de escolarização (analfabeto ou primeiro ciclo do EF). Estes valores aumentam substancialmente quando se analisam estudantes residentes na zona rural, em que 59% e 55% dos pais e 66% e 62% das mães, respectivamente, são analfabetos ou têm apenas do 1º ao 5º ano do EF. A proporção de estudantes com pai com nível de Ensino Superior foi de 15%, 3% e 4% para alunos do grupo Urbano-

³² Segundo os autores, para as meninas, não haveria tanto prejuízo à escolarização, pois elas estariam inseridas principalmente, em atividades domésticas que, aparentemente, melhor se adequariam às demandas escolares, em razão, principalmente pela flexibilidade de horários.

Urbano, Rural-Urbano e Rural-Rural, respectivamente. Para as mães, essa proporção é de 27,2%, 9,3% e 10,9% para alunos do grupo Urbano-Urbano, Rural-Urbano e Rural-Rural, respectivamente.

Em relação à renda domiciliar declarada pelos alunos, para os três grupos, os alunos estavam mais concentrados entre domicílios com rendimento familiar de um ou até dois salários-mínimos. A condição socioeconômica familiar estava em torno de 22,4 pontos para alunos do Urbano-Urbano e perto de 17,6 para alunos do Rural-Urbano e Rural-Rural (informações adicionais sobre a proporção de alunos em cada classe podem ser consultadas na Tabela A.2.3 no Apêndice 2).

Da amostra, 11,2% dos alunos do grupo Urbano-Urbano estudavam em escolas profissionalizantes e 74,7% em escolas públicas. Dos alunos do Rural-Urbano, 8,1% e 92,9% frequentavam escolas profissionalizantes e da rede pública, respectivamente, no grupo Rural-Rural, 10,1% estavam em escolas de Ensino Médio integrado ao Ensino Profissionalizante e 92,9% em escolas públicas.

Estudos também indicam relação positiva entre a composição socioeconômica do alunato de uma escola e o rendimento escolar alcançado pelos estudantes de baixo *status* socioeconômico. Isso acontece porque a concentração do corpo estudantil de composição socioeconômica mais elevada pode facilitar a instrução, por criar um ambiente no qual as normas e procedimentos sejam favoráveis ao aprendizado. (AIKEN; BARBARIN, 2008). É importante, entretanto, ressaltar, que, embora essa hipótese seja confirmada na literatura³³, existem situações em que escolas com nível socioeconômico mais homogêneo e considerado baixo conseguem se sobressair mais do que outras. Isso é, escola que, mesmo tendo como clientela crianças de nível socioeconômico desfavorável, conseguem oferecer um ensino de qualidade. Foi o que mostrou o estudo desenvolvido por Faria e Guimarães (2015), sendo que essas escolas possuem como diferencial bons gestores (diretores) e um ambiente favorável ao aprendizado. Portanto, a fim de controlar o efeito dos pares, neste estudo, é utilizado o Indicador do Nível Socioeconômico da Escola (INSE) para controlar os possíveis efeitos relacionados ao contexto socioeconômico dos colegas onde o aluno está

³³ Por exemplo, o estudo de Aiken e Barbarin (2008) mostra que as condições socioeconômicas da escola e da vizinhança contribuíam mais nas taxas de aprendizagem das crianças em leitura do que as características socioeconômicas da família; de modo que a composição da população estudantil, ligada pela concentração da pobreza, estava associada a crianças com baixo nível de aprendizado.

inserido.

O INSE médio das escolas frequentadas por alunos do Urbano-Urbano é em torno de 52,3 pontos, isto é, estão inseridos no grupo de escolas que possuem nível socioeconômico dos alunos considerado médio. Já os alunos do grupo Rurais-Urbano e do Rural-Rural apresentaram INSE médio de 45,5 e 43,8 pontos, respectivamente. Este se enquadra no estrato de alunos com nível socioeconômico médio baixo. Na Tabela A.2.4, é apresentada a proporção de alunos de acordo com cada estrato da distribuição do INSE. Já a média de aluno por turma varia de 33, no grupo Urbano-Urbano a 27 no grupo Rural-Rural. Em relação às variáveis de controle dos docentes, observa-se que a IRD não varia significativamente entre os grupos, mas o Índice de Esforço Docente (IED) exibiu uma diferença maior para o grupo Rural-Rural. Este resultado pode ser consequência da falta de professores para escolas rurais, o que resulta em uma carga horária maior para aqueles docentes alocados nessas escolas.

Além disso, pelo Indicador de Adequação da Formação Docente (IAFD), observa-se que apenas 50% dos professores que ministram aulas para o grupo Rural-Rural possuem formação adequada para as disciplinas que lecionam. No grupo Urbano-Urbano, a proporção é de 63,2%. Enquanto isso, 75,5% das escolas frequentadas pelo grupo Urbano-Urbano possui infraestrutura adequada. Entre as escolas frequentadas pelo grupo Rural-Urbano, a proporção é de 54,9%, e do grupo Rural-Rural é ainda inferior - 50,7%.

Para alcançar os objetivos deste estudo, entretanto, não seria possível simplesmente comparar as médias, variâncias e desvios-padrão entre os grupos Urbano-Urbano, Rural-Rural e Urbano-Rural. É necessário que os alunos pertencentes aos grupos possuam características de seleção similares, de modo que se possa obter resultados mais robustos na estimação dos efeitos dos fatores considerados. Dessa forma, a subseção a seguir busca discutir os resultados com suporte nas metodologias apresentadas anteriormente.

5.2 Diferencial de desempenho escolar entre alunos do Rural-Rural e Rural-Urbano (Grupo 1)

Inicialmente, faz-se o pareamento e as regressões para o Grupo 1 (Rural-

Rural e Rural- Urbano). Alunos que residem no meio rural e estudam em escolas tanto da zona rural como na zona urbana.

5.2.1 Pareamento pelo CEM

Para o pareamento entre os dois grupos, pelas suas características observáveis, no primeiro momento, é aplicado o algoritmo CEM. O objetivo é encontrar alunos com características individuais e familiares iguais que possam ser comparados entre os grupos de controle e tratamento. Assim, é possível isolar o efeito dessas variáveis sobre o desempenho dos alunos e encontrar a diferença no rendimento no qual decorre do **efeito-escola**. Portanto, as variáveis pré-selecionadas estão relacionadas aos alunos e suas respectivas famílias.

A Tabela 10 traz as medidas de desequilíbrios antes e após o pareamento. Nota-se que o valor de \mathcal{L}_1 multivariado reduziu de 0.39646111 para 0.25364464 e que algumas observações foram excluídas dos dois grupos após o pareamento, restando 23.147 observações do grupo de tratamento e 109.465 do grupo de controle, que podem ser comparados com base nas das características pré-selecionadas de modo a isolar o **efeito-família** sobre a nota do aluno.

Tabela 10 - Resultados do algoritmo CEM para o Grupo 1

Distância multivariada	Grupo 1	
\mathcal{L}_1 - Antes do CEM	0.39646111	
\mathcal{L}_1 - Após o CEM	0.25364464	
Algoritmo CEM	Tratados ¹	Controle ²
	1	0
Total de Observações	24.397	130.963
Comparáveis – Após o CEM ³	23.147	109.465
Não comparáveis – Após o CEM	1.250	21.498

Fonte: RODRIGUES (2017) com base nos resultados da pesquisa

Nota: ¹ Alunos que estudam em escolas rurais. ² Alunos residentes no meio rural que estudam em escolas urbanas. ³ Variáveis selecionadas para o pareamento: **Individuais:** idade2 idade3 sexo branca solteiro trabalha conc_ef conc_em. **Família:** Edu_pai2 Edu_pai3 Edu_pai4 Edu_mae2 Edu_mae3 Edu_mae4 Renda01 Renda02 Renda03 Renda04 csf.

5.2.2 Resultados da Decomposição de Oaxaca e Blinder (1973) sem ponderação

A Tabela 11 mostra a decomposição de rendimento escolar dos dois grupos pelo método desenvolvido por Oaxaca-Blinder (1973), inicialmente sem ponderação, ou seja, sem isolar as características individuais e familiares dos alunos, com vista a se obter os diferenciais sem o balanceamento dos tratados e controles.

Primeiramente, os resultados apontam que o diferencial de rendimento entre alunos do grupo Rural-Rural e Rural-Urbano é estatisticamente significativo ao nível de 5% para quase todas as avaliações, sempre em favor dos alunos do meio rural que estudam em escolas urbanas, com exceção do desempenho em Ciências Humanas. De acordo com este modelo, o maior diferencial acontece na avaliação de Linguagem e Códigos, em que a diferença entre os grupos é, em média, de aproximadamente 3,3 pontos. Vale ressaltar que a maior parte do diferencial é atribuída aos fatores explicados, ou seja, pelas variáveis observáveis do modelo.

Com relação à diferença explicada, observa-se que as características da escola obtiveram o maior peso no diferencial de notas entre os grupos, denotando, assim, que estas variáveis são importantes para explicar o diferencial de rendimento escolar dos alunos dos grupos Rural-Rural e Rural-Urbano. Portanto, o diferencial no desempenho aumenta em função das características observadas na escola. Por outro lado, as características individuais e familiares, são responsáveis pela redução da diferença entre os grupos.

Tabela 11 - Decomposição Oaxaca-Blinder do diferencial de rendimento escolar Grupo 1¹ para o Brasil - Sem ponderação

Variáveis	Ciências da Natureza	Ciências Humanas	Linguagem e Códigos	Matemática	Média Geral
Rural_Urbano (1)	464.0005 *	516.7578 *	482.0894 *	443.4353 *	476.5708 *
Rural_Rural (2)	463.0143 *	516.8549 *	478.7449 *	442.0192 *	475.1583 *
<i>Diferença total (1 - 2)</i>	0.9862 **	-0.0970	3.3445 *	1.4160 **	1.4124 *
<i>Diferença explicada</i>	1.2351 *	2.8414 *	4.5878 *	1.2345 **	2.4747 *
<i>Diferença não explicada</i>	-0.2490	-2.9385 *	-1.2433 **	0.1816	-1.0623 **
<i>Diferença Explicada detalhado</i>					
Características individuais	-0.3945 *	0.1038	1.0817 *	-1.1472 *	-0.0890
Características familiares	-1.1024 *	-1.3682 *	-1.4772 *	-1.6745 *	-1.4056 *
Características da Escola	2.7321 *	4.1058 *	4.9833 *	4.0561 *	3.9693 *
<i>Diferença não Explicada detalhado</i>					
Características individuais	-2.8556	-1.0198	-1.1628	-2.4779	-1.8790
Características familiares	-2.1542	-2.8707	-5.9869 **	5.0614	-1.4876
Características da Escola	-15.7926 *	-14.3899 **	-25.8773 *	-19.7754 *	-18.9588 *
Constante	20.5534 *	15.3419	31.7838 *	17.3735 **	21.2631 *

Fonte: RODRIGUES (2017) com base nos resultados da pesquisa

Nota: ¹Alunos que estudam em escolas rurais vs alunos residentes no meio rural que estudam em escolas urbanas.

Indivíduos: idade2 idade3 sexo branca solteiro trabalha conc_ef conc_em. **Família:** Edu_pai2 Edu_pai3 Edu_pai4 Edu_mae2 Edu_mae3 Edu_mae4 Renda01 Renda02 Renda03 Renda04 csf. **Escola:** esc_profiss publico inse aluno_turma ird ied iafd iie.

(*) significante a 1%; (**) significante a 5%.

Quanto às características não explicadas, nota-se que estas apresentaram

sinal negativo em todas as avaliações, porém, no geral, apenas as características das escolas foram estatisticamente significantes ao nível de confiança de 95%.

5.2.3 Resultados da Decomposição de Oaxaca e Blinder (1973) com ponderação

Conforme foi mencionado, o objetivo do estudo é encontrar o **efeito-escola** isolando os efeitos das demais variáveis do modelo. Para tal propósito, a decomposição de Oaxaca-Blinder é estimada novamente, considerando a ponderação ou utilizando o peso gerado pelo método CEM, tornando a amostra mais homogênea, de modo que as comparações são feitas entre alunos que possuem características similares.

Assim sendo, na Tabela 12 encontram-se os resultados da decomposição de Oaxaca-Blinder com ponderação, considerando as mesmas variáveis explicativas utilizadas anteriormente e demonstradas na Tabela 11. No geral, os resultados do diferencial de rendimento entre alunos do grupo Rural-Rural e Rural-Urbano com ponderação é superior aos valores exibidos anteriormente. Ou seja, comparando alunos com características individuais e familiares similares, o **efeito-escola** é superior e estatisticamente significativo em todas as disciplinas. Assim, os resultados indicam que os alunos da zona rural e estudantes de escolas urbanas, considerando a Média Geral das avaliações, tem desempenho médio superior aos alunos de escolas rurais, de aproximadamente 3,7 pontos, sendo que grande parte desse diferencial é dado pelas características explicadas no modelo.

Tabela 12 - Decomposição Oaxaca-Blinder do diferencial de rendimento escolar Grupo 1¹ para o Brasil - Com ponderação

Variáveis	Ciências da Natureza	Ciências Humanas	Linguagem e Códigos	Matemática	Média Geral
Rural_Urbano (1)	465.8909 *	518.8742 *	483.2902 *	447.3970 *	478.8631 *
Rural_Rural (2)	463.0143 *	516.8549 *	478.7449 *	442.0192 *	475.1583 *
<i>Diferença total (1 - 2)</i>	2.8766 *	2.0193 *	4.5453 *	5.3778 *	3.7048 *
<i>Diferença explicada</i>	3.6257 *	5.1242 *	5.9644 *	5.2251 *	4.9848 *
<i>Diferença não explicada</i>	-0.7491	-3.1049 *	-1.4190 *	0.1527	-1.2801 *
<i>Diferença Explicada Detalhada</i>					
Características individuais	0.0000	0.0000	0.0000	0.0000	0.0000
Características familiares	0.0016	-0.0013	0.0008	0.0036	0.0012
Características da Escola	3.6240 *	5.1255 *	5.9636 *	5.2215 *	4.9837 *
<i>Diferença não Explicada Detalhada</i>					
Características individuais	-5.9076	-3.8160	-2.5170	-5.3770	-4.4044
Características familiares	-3.6793	-4.5958	-5.7593 **	2.9722	-2.7656
Características da Escola	-6.0608	-5.6242	-19.5097	-8.2209	-9.8539
Constante	14.8987	10.9312	26.3671	10.7785	15.7439 *

Fonte: RODRIGUES (2017) com base nos resultados da pesquisa

Nota: ¹Alunos que estudam em escolas rurais vs alunos residentes no meio rural que estudam

em escolas urbanas.

Indivíduos: idade2 idade3 sexo branca solteiro trabalha conc_ef conc_em. **Família:** Edu_pai2 Edu_pai3 Edu_pai4 Edu_mae2 Edu_mae3 Edu_mae4 Renda01 Renda02 Renda03 Renda04 csf. **Escola:** esc_profiss publico inse aluno_turma ird ied iafd iie.

(*) significante a 1%; (**) significante a 5%.

Portanto, ao comparar alunos da zona rural que estudam em escolas no próprio meio rural com aqueles que estudam em escolas urbanas, é possível afirmar, para esta amostra, que o **efeito-escola** é positivo para o grupo Rural-Urbano. Em outras palavras, os resultados expressam que é mais vantajoso para alunos do meio rural estudarem em escolas urbanas. Das características explicadas que mais contribuem com o diferencial entre os grupos, destacam-se o Indicador de Nível Socioeconômico (INSE) dos alunos nas escolas, o índice de esforço docente e o Índice de Adequação da Formação Docente (IAFD) (Ver Tabela A.2.5 no Apêndice 2).

5.3 Diferencial de desempenho escolar entre alunos do Rural-Urbano e Urbano-Urbano (Grupo 2)

Nessa subseção, é analisado o diferencial de desempenho escolar entre alunos do meio rural que estudam em escolas urbanas em relação aos estudantes que vivem e estudam na zona urbana. Aqui o objetivo é encontrar o **efeito-família**, isso é, isolando o **efeito-escola** sobre o desempenho dos alunos, para encontrar o impacto das características individuais e familiares no diferencial entre os alunos dos meios rural e urbano.

5.3.1 Pareamento pelo CEM

Inicialmente, é necessário fazer o pareamento entre os grupos. Nesse caso, assume-se o argumento que os estudantes residentes no meio rural e urbano estudam nas mesmas escolas, ou escolas com características muito semelhantes, o que significa dizer que as qualidades dos professores e da escola que serve aos alunos é a mesma, tanto para discentes do meio rural como do urbano, ou seja, a diferença de desempenho entre eles não pode ser atribuída à diferença das características da escola frequentada.

Portanto, a primeira etapa é a definição dos grupos de tratamento e controle pelo método de pareamento das características observáveis da escola pelo algoritmo CEM. Na Tabela 13 encontram-se as medidas de desequilíbrio pelo método, antes e

após o pareamento. Verifica-se que os valores do \mathcal{L}_1 multivariado reduziu-se para valores próximos de zero (de 0.58189444 para 0.04394802). Do grupo de tratamento restaram 129.211 observações e do grupo de controle 797.608, que poderão ser comparadas a base nas características pré-selecionadas de modo a isolar o **efeito-escola** sobre o desempenho dos alunos.

Tabela 13 - Resultados do algoritmo CEM para o Grupo 2

Distância multivariada	Grupo 2	
\mathcal{L}_1 - Antes do CEM	0.58189444	
\mathcal{L}_1 - Após o CEM	0.04394802	
Algoritmo CEM	Tratados ¹	Controle ²
	1	0
Total de Observações	130.963	928.655
Comparáveis – Após o CEM ³	129.211	797.608
Não comparáveis – Após o CEM	1.752	131.047

Fonte: RODRIGUES (2017) com base nos resultados da pesquisa

Nota: ¹ Alunos do meio rural que estudam em escolas urbanas. ² Alunos que residem e estudam em escolas da zona urbana. ³ Variáveis selecionadas para o pareamento: **Escola:** esc_profiss publico inse aluno_turma ird ied iaafd iie

5.3.2 Resultados da Decomposição de Oaxaca e Blinder (1973) sem ponderação

Após a definição do grupo de controle e tratamento, a segunda análise é feita pela decomposição entre os grupos pelo método de Oaxaca-Blinder sem ponderação. A Tabela 14 identifica o diferencial de rendimento escolar para os alunos Rural-Urbano e Urbano-Urbano nas quatro áreas de conhecimentos e habilidades do Enem, bem como a média geral das avaliações. Nesse caso, não se considera o fato de os alunos estudarem em escolas semelhantes.

O diferencial de rendimento entre alunos residentes no meio rural e urbano é estatisticamente significativo ao nível de 1% e favorável aos alunos do grupo Urbano-Urbano em todas as avaliações e na média geral. A diferença da nota entre os grupos Rural-Urbano e Urbano-Urbano foi menor em Ciências da Natureza, de 24,20 pontos, indicando que os alunos que residem no meio urbano têm desempenho maior do que aqueles moradores em áreas rurais. A maior diferença é observada na prova de Matemática, a diferença total entre os grupos foi de 38,08 pontos; mais uma vez, os alunos do meio urbano expressaram vantagem em relação aos alunos residentes no meio rural.

Observa-se que a parte explicada, referente as características observáveis, é responsável por explicar a maior parte do diferencial. Por exemplo, em Ciências da Natureza, as características explicadas são responsáveis por 90,7%³⁴ do diferencial total entre os dois grupos. Em matemática ele explica 91,4% da diferença total.

Tabela 14 - Decomposição Oaxaca-Blinder do diferencial de rendimento escolar Grupo 2¹ para o Brasil - Sem ponderação

Variáveis	Ciências da Natureza	Ciências Humanas	Linguagem e Códigos	Matemática	Média Geral
Urbano Urbano (1)	489.2522 *	550.0911 *	515.2744 *	483.7105 *	509.5820 *
Rural Urbano (2)	465.0463 *	518.0269 *	483.0297 *	445.6302 *	477.9333 *
<i>Diferença total (1 - 2)</i>	24.2059 *	32.0642 *	32.2447 *	38.0803 *	31.6488 *
<i>Diferença explicada</i>	21.9589 *	25.3128 *	24.1556 *	34.8181 *	26.5613 *
<i>Diferença não explicada</i>	2.2471 *	6.7514 *	8.0891 *	3.2622 *	5.0874 *
<i>Diferença Explicada detalhado</i>					
Características individuais	1.7895 *	2.3711 *	1.9849 *	3.6060 *	2.4379 *
Características familiares	6.0681 *	8.1295 *	8.1862 *	11.5393 *	8.4808 *
Características da Escola	14.1012 *	14.8122 *	13.9845 *	19.6727 *	15.6427 *
<i>Diferença não Explicada detalhado</i>					
Características individuais	-1.0617	-1.8664	1.2891	1.4725	-0.0416
Características familiares	-2.8392 **	0.1226	1.4066	-3.6032 **	-1.2283
Características da Escola	69.5398 *	43.6508 *	46.2965 *	112.2524 *	67.9349 *
Constante	-63.3919 *	-35.1557 *	-40.9031 *	-106.8595 *	-61.5775 *

Fonte: RODRIGUES (2017) com base nos resultados da pesquisa

Nota: ¹ Alunos residentes no meio rural que estudam em escolas urbanas vs alunos de escolas urbanas.

Individuais: idade2 idade3 sexo branca solteiro trabalha conc_ef conc_em. **Família:** Edu_pai2 Edu_pai3 Edu_pai4 Edu_mae2 Edu_mae3 Edu_mae4 Renda01 Renda02 Renda03 Renda04 csf. **Escola:** esc_profiss publico inse aluno_turma ird ied iafd iie

(*) significativa a 1%; (**) significativa a 5%.

Nota-se, ainda, na Tabela 14, dentro do detalhamento da diferença explicada, que os três grupos de variáveis (características individuais, familiares e da escola) são responsáveis por aumentar essa diferença entre os grupos em todas as avaliações, sendo que a maior diferença é explicada pelas características da escola. Como o objetivo é verificar, no entanto, o **efeito-família**, ou seja, o diferencial de rendimento escolar entre alunos do meio rural e urbano que estudam em escolas com as mesmas características é necessário isolar o **efeito-escola** pelo uso da ponderação, que é desenvolvido na próxima subseção.

5.3.3 Resultados da Decomposição de Oaxaca e Blinder (1973) com ponderação

As comparações realizadas aqui são feitas considerando que os alunos

³⁴ (Diferença explicada/Diferença total) *100 ou (Diferença não explicada /Diferença total) *100

residentes no meio rural estudam em escolas semelhantes aos alunos da área urbana. Com isso, os resultados encontrados são mais robustos, de modo que é possível encontrar o diferencial que é decorrente do **efeito-família**. Busca-se explicar se alunos residentes da zona rural que estudam em escolas possuidoras das mesmas características de infraestrutura e professores qualificados tendem a ter desempenhos mais parecidos com os obtidos por alunos do meio urbano.

Em suma, o objetivo é verificar se, quando há melhora (ou piora) nas condições da escola, a diferença de resultado entre grupos de alunos, considerando suas características individuais e familiares, diminui. Em outras palavras, o que se busca saber é se os alunos são afetados pelo ambiente ou pelas condições da escola, diminuindo assim as desigualdades entre os grupos.

Os dados da Tabela 15 mostram que, mesmo após o rígido controle exercido em relação às características da escola, o diferencial de rendimento escolar entre aluno do grupo Rural-Urbano e Urbano-Urbano ainda é favorável ao último grupo. A diferença, porém, se reduziu significativamente. Enquanto a diferença total na média geral era de 31,64 pontos (Tabela 14) entre os grupos sem ponderação, ao considerar que alunos do meio rural e urbano estudam em escolas similares, o diferencial total de rendimento reduziu para uma média de 8,32 pontos, sendo que 57,7% do diferencial podem ser atribuídos a fatores não explicados, ou seja, a variáveis não observáveis diretamente.

No modelo com ponderação, o maior valor expresso pela diferença total é observado na prova de Linguagem e Códigos, com diferença total de 10,67 pontos entre os grupos. Também é notório que a diferença explicada tem peso maior nas avaliações de Ciências da Natureza e de Matemática. Sendo assim, a diferença explicada (ou efeito característica) contribui com 50,4% e 58,9%, na avaliação de ciências da natureza e de matemática, respectivamente, no aumento do hiato de rendimento escolar entre alunos do grupo Rural-Urbano e Urbano-Urbano.

Por outro lado, para as avaliações de Ciências Humanas e Linguagem e Códigos, os resultados apontam para maior importância relativa aos componentes não observáveis ou a diferença não explicada, sendo de 63,2% e 69,0% para Ciências Humanas e Linguagem e Códigos, respectivamente.

Tabela 15 - Decomposição Oaxaca-Blinder do diferencial de rendimento escolar Grupo 2¹ para o Brasil - Com ponderação

Variáveis	Ciências da Natureza	Ciências Humanas	Linguagem e Códigos	Matemática	Média Geral
Urbano_Urbano (1)	469.3935 *	527.9337 *	493.7042 *	454.0089 *	486.2601 *
Rural_Urbano (2)	465.0463 *	518.0269 *	483.0297 *	445.6302 *	477.9333 *
<i>Diferença total (1 - 2)</i>	4.3472 *	9.9068 *	10.6745 *	8.3786 *	8.3268 *
<i>Diferença explicada</i>	2.1895 *	3.6424 *	3.3080 *	4.9327 *	3.5182 *
<i>Diferença não explicada</i>	2.1577 *	6.2644 *	7.3665 *	3.4459 *	4.8086 *
<i>Diferença Explicada Detalhado</i>					
Características individuais	0.3677 *	0.3995 *	0.0003	1.1652 *	0.4831 *
Características familiares	1.8138 *	3.2334 *	3.2978 *	3.7544 *	3.0248 *
Características da Escola	0.0080	0.0095	0.0100	0.0132	0.0102 *
<i>Diferença não Explicada Detalhado</i>					
Características individuais	0.1744	-0.4614	1.1748 *	2.6577	0.8864
Características familiares	1.1479	4.0279 *	3.7336	4.3908 *	3.3251 *
Características da Escola	-3.3859	-4.3359	-1.6109 *	-4.0092	-3.3355
Constante	4.2213	7.0338 *	4.0690	0.4065	3.9327

Fonte: RODRIGUES (2017) com base nos resultados da pesquisa

Nota: ¹ Alunos residentes no meio rural que estudam em escolas urbanas vs alunos de escolas urbanas.

Indivíduos: idade2 idade3 sexo branca solteiro trabalha conc_ef conc_em. **Família:** Edu_pai2 Edu_pai3 Edu_pai4 Edu_mae2 Edu_mae3 Edu_mae4 Renda01 Renda02 Renda03 Renda04 csf. **Escola:** esc_profiss publico inse aluno_turma ird ied iafd iie

(*) significante a 1%; (**) significante a 5%.

Assim, pelo que foi mostrado nesta subseção, pela decomposição de Oaxaca-Blinder com ponderação, os alunos do grupo Urbano-Urbano ainda apresentam desempenho superior aos alunos do Rural-Urbano, mesmo após o controle das características das escolas, embora essa diferença tenha sido reduzida significativamente; ou seja, ainda existem diferenças entre os grupos explicadas pelas características individuais e familiares observáveis e não observáveis no modelo, oriundas de particularidades e capacidades inatas de cada aluno. Na Tabela A.2.6 no Apêndice 2, mostra o detalhamento das variáveis dos alunos e das famílias, individualmente.

De modo que pelos resultados obtidos fica evidente que o ambiente escolar é importante na redução do diferencial de desempenho dos alunos que residem na zona rural em relação aos que vivem em áreas urbanas. Isto é, quando há melhora (ou piora) das condições da escola, a diferença de resultado entre grupos de alunos diminui. Em suma, quando frequentam escolas com características de infraestrutura e com professores com as mesmas características, tendem a der rendimentos semelhantes. No entanto, mesmo quando a escola faz a diferença, as características individuais e famílias, ainda são responsáveis em colocar os alunos da zona rural em desvantagem em relação aos seus colegas que moram na própria zona urbana.

6 CONSIDERAÇÕES FINAIS

Este estudo teve como objetivo trazer para literatura o tema referente ao efeito da migração educacional de alunos do meio rural para escolas urbanas. No Brasil, que acontece, sobretudo, entre o Ensino Fundamental e Ensino Médio. Para tal, foi estimado o diferencial de desempenho escolar dos alunos da 3ª série do Ensino Médio que estudavam em escolas rurais e urbanas e realizadas a decomposição deste diferencial em dois efeitos: **efeito-escola** e **efeito-família**. Pelo método de pareamento do algoritmo *Coarsened Exact Matching* (CEM) que controla as características observáveis dos alunos, combinado com método de decomposição desenvolvido por Oaxaca (1973) e Blinder (1973), foi possível estimar os dois efeitos, empregando dados do Enem.

Na seção 2, foi exibida uma análise do fluxo escolar dos alunos residentes em áreas rurais que foram aprovados no Ensino Fundamental em 2011 e ingressaram no Ensino Médio em 2012. Observou-se que houve redução do número de matrículas dos alunos do meio rural que ingressaram no ensino médio em todo o período. A maior taxa de desistência foi entre os alunos que reprovados em pelo menos uma série, entre os que frequentavam aulas no período noturno, entre alunos do sexo masculino e entre os estudantes matriculados em escolas urbanas. Do total de alunos que ingressaram no Ensino Médio, em 2012, apenas 77% permaneceram no sistema de ensino sem abandonar a escola e somente 64,6% chegaram à 3ª série sem nenhuma reprovação.

Os resultados para os alunos dos grupos denominados como Rural-Rural e Rural-Urbano mostraram que, após o controle do *background* familiar dos alunos, as diferenças de desempenho entre os grupos se tornam ainda mais significativas do que em uma análise sem o uso de qualquer pareamento (sem a ponderação). Tais resultados também apontam para um efeito favorável aos alunos residentes do meio rural que estudavam em escolas urbanas; ou seja, o **efeito-escola** foi positivo para o grupo Rural-Urbano. Em outras palavras, os resultados dizem que é mais vantajoso para alunos do meio rural estudarem em escolas urbanas, visto que essas denotam melhor infraestrutura, têm professores mais qualificados e o *status* socioeconômico dos alunos proporciona um ambiente mais favorável ao aprendizado.

Analisando o grupo de alunos do Rural-Urbano e Urbano-Urbano observou-se que, após o controle de variáveis relacionadas à escola, pela decomposição de

Oaxaca-Blinder com ponderação, ou seja, combinada com o método de pareamento, os alunos do grupo Urbano-Urbano ainda apresentaram desempenho superior ao dos alunos do grupo Rural-Urbano, embora essa diferença tenha sido reduzida significativamente em consideração ao modelo sem ponderação. O que significa que as características individuais e familiares observáveis e não observáveis no modelo, derivadas de particularidades e capacidades inatas de cada estudante, ainda são responsáveis pelo diferencial entre os grupos, que é o denominado **efeito-família**. Isto é, mesmo exposto às mesmas condições de infraestrutura escolar e de ensino, os alunos que residem na zona rural e estudam em escolas urbanas, ainda tem desempenho inferior aos seus colegas por apresentarem características individuais e familiares que os colocam em desvantagens em relação aos seus pares.

Embora os resultados encontrados sejam bastante interessantes e intrigantes, é necessário observar que o Enem não é uma avaliação obrigatória, portanto, os alunos participam voluntariamente. Com isso, estudantes de baixo desempenho escolar podem estar fora da amostra. Sendo assim, uma precaução com relação aos resultados é que estes não podem ser considerados representativos de todo o universo dos alunos do Ensino Médio brasileiro.

Para a amostra utilizada nesse estudo, no entanto, os resultados indicam que os diferenciais de desempenho entre os alunos pesquisados estão fortemente relacionados com a localização ou zona em que estes estudam ou residem. Em relação ao Grupo 1 (Rural-Rural e Rural-Urbano), a zona da escola é responsável por aumentar o diferencial entre os grupos. E, no Grupo 2 (Rural-Urbano e Urbano-Urbano), a zona de moradia é que contribuí com este diferencial, isso é, mesmo estudando em escolas semelhantes, ainda existem diferenças individuais importantes entre os grupos, oriundas da origem familiar e ou habilidades inatas dos alunos.

Portanto, embora seja necessário um número maior de escolas de Ensino Médio no meio rural, principalmente para atender àqueles jovens que concluem o Ensino Fundamental, mas encontram dificuldades em se deslocarem para escolas urbanas. É importante se pensar na qualidade da infraestrutura e na oferta de capital humano, formada, principalmente, pelo corpo docente da escola, e em outros fatores que são determinantes para a melhoria na qualidade da educação.

CONCLUSÃO GERAL

Essa dissertação é composta de dois artigos que abordaram assuntos distintos com respeito a diferenças de desempenho entre alunos do meio rural e urbano, utilizando metodologias de decomposição que mensuram a diferença entre alunos, tanto na média como ao longo da distribuição de desempenho.

No primeiro artigo, os resultados encontrados mostraram que alunos de escolas urbanas apresentam desempenho superior aos de escolas rurais ao longo de toda a distribuição. O exercício de decomposição por *quantis* evidenciou que o papel do efeito-característica é crescente ao longo da distribuição, ou seja, quanto mais elevado é o *quantil* de desempenho, mais importantes são as características na explicação do hiato educacional entre os grupos, principalmente aqueles relacionados aos atributos escolares e familiares. E o efeito estrutural (fatores não observáveis), embora tenha apresentado peso relativamente menor, também contribui para o aumento do diferencial entre os grupos.

De tal sorte, os resultados encontrados indicam que as escolas do meio rural ainda deixam muito a desejar em termos de infraestrutura e qualificação dos professores, o que influencia diretamente no desempenho escolar dos alunos do meio rural. Esses, além de estarem em desvantagem em relação às características familiares, não recebem um ensino de qualidade da escola que frequentam suficiente para reduzir o hiato em relação aos alunos do meio urbano.

Com relação ao segundo artigo, uma das contribuições do estudo reside tanto em relação ao tema, que parece não ter sido ainda explorado na literatura pelo foco abordado nesta dissertação, como também pela metodologia utilizada. Isso porque, por intermédio do pareamento pelo algoritmo *Coarsened Exact Matching* (CEM), combinado com o método de decomposição de Oaxaca (1973) e Blinder (1973), foi possível comparar alunos que possuem características individuais e familiares iguais, mas estudam em escolas diferentes, e estudantes que estudam em escolas semelhantes, mas que possuem características individuais e familiares distintas. Assim, foi possível tornando os grupos mais equilibrados e permitindo identificar as variáveis que mais contribuem para explicar as diferenças de resultados educacionais entre os alunos.

Os resultados para os alunos da zona rural e estudando em escolas rurais em

relação aos alunos da zona rural que estudavam em escolas urbanas mostraram que, após o controle do *background* familiar dos alunos, as diferenças entre os grupos foram ainda maiores do que sem a ponderação e favorável aos alunos residentes do meio rural que estudavam em escolas urbanas. O **efeito-escola** foi positivo para o grupo de alunos da zona rural que estudavam em escolas urbanas, fato indicativo de que seria mais vantajoso para alunos do meio rural estudarem em escolas urbanas, visto que essas detêm melhor infraestrutura, têm professores mais qualificados e o *status* socioeconômico dos alunos proporciona um ambiente mais favorável ao aprendizado.

Para mensurar o **efeito-escola**, foram comparados alunos residentes da zona rural e urbana, estudando em escolas semelhantes. Observou-se que, mesmo após o controle de variáveis relacionadas à escola, os alunos da zona urbana ainda registraram desempenho superior aos alunos do meio rural. O que significa que as características individuais e familiares observáveis e não observáveis no modelo que são oriundas de particularidades e capacidades inatas de cada estudante, ainda são responsáveis pelo diferencial entre os grupos, que é o **efeito-família**.

Resumidamente, mesmo exposto às mesmas condições de infraestrutura escolar e de ensino, os alunos que residem na zona rural, mas estudam em escolas urbanas, têm desempenho inferior aos seus colegas por apresentarem características individuais e familiares que os colocam em desvantagens em relação aos seus pares que moram na zona urbana.

Dada à importância deste tema, e reconhecendo as diferenças ou heterogeneidade entre as regiões do Brasil, como agenda futura de estudo, pretende-se analisar, tanto o primeiro artigo, como o segundo, conforme as regiões e alguns estados, em particular, o Ceará, já que nos últimos anos, este Estado, vem se destacando nacionalmente pelos avanços na qualidade do ensino, sobretudo, nos anos iniciais do ensino fundamental.

REFERÊNCIAS

AIKENS, N, L.; BARBARIN, O, Socioeconomic differences in Reading trajectories: the contribution of family, neighborhood, and school contexts, **Journal of Educational Psychology**, v.100, n. 2, p. 235-251, Washington, 2008.

ALMEIDA, A. T. C. de. Determinantes dos piores e melhores resultados educacionais dos alunos da rede pública de Ensino Fundamental no Brasil. **Planejamento e Políticas Públicas** – PPP. n. 42. Brasília, 2014

ALBERNAZ, A.A.; FERREIRA, F.H.G.; FRANCO, C. Qualidade e equidade no ensino fundamental brasileiro. **Pesquisa e Planejamento Econômico**. v. 32, n.3.p. 453-476, Rio de Janeiro: IPEA, 2002.

AMINI C.; NIVOROZHKIN E. The urban–rural divide in educational outcomes: Evidence from Russia. **International Journal of Educational Development**. v.44, p. 118–133, Washington, 2015.

ARISTIZABAL, G. C.; ESTEBAN, G. G.; EMBÚN, D. P. X. de. Desigualdades educativas en america latina, PISA 2012: Causas de las Diferencias en Desempeño Escolar entre los Colégios públicos y privados. **Documentos de Trabajo FCEA**. n.19, Santiago, 2016.

_____. El desempeño educativo escolar em Colombia: factores que determinan la diferencia en rendimiento académico entre las escuelas públicas y privadas. **Investigaciones de Economía de la Educación**. n.9, Madrid, 2014.

AROCA, P., BRIDA, J.G, VOLO, S. Tourism statistics: correcting data inadequacy using coarsened exact matching. (Working Paper). **School of Economics and Management at the Free University of Bozen**, Bolzano, 2014.

ARTES, C. A. A.; CARVALHO, M. P. O trabalho como fator determinante da defasagem escolar dos meninos no Brasil: mito ou realidade? **Cadernos pag**. v.34, p. 41-74, São Paulo, 2010.

BADR, M.; MORRISSEY, O.; APPLETON, S. Gender differentials in maths test scores in MENA countries. **CREDIT research paper**, v.12 n.04, Malaysia: University of Nottingham, 2012.

BARBOSA, M.E.F.; FERNANDES, C. A escola brasileira faz diferença? uma investigação dos efeitos da escola na proficiência em Matemática dos alunos da 4a série. *In*: FRANCO, C. **Promoção, ciclos e avaliação educacional**. Porto Alegre: ArtMed, 2001.

BEZERRA, M. G.; KASSOUF, A. L. Análise dos fatores que afetam o desempenho escolar nas escolas das áreas urbanas e rurais do Brasil. *In*: XLIV congresso da SOBER, 2006, Fortaleza. **Anais...**Fortaleza, 2006.

BIONDI, R. L.; FELÍCIO, F. de. **Atributos escolares e o desempenho dos estudantes: uma análise em painel dos dados do Saeb.** (Texto para discussão). p. 19, Brasília: INEP, 2007.

BLACKWELL, M. *et al.* CEM: Coarsened exact matching in Stata. **The Stata Journal**. v. 9, n. 4, p. 524-546, College Station, 2009.

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

BORAH, B. J.; BASU, A. Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. **Health Econ**. v. 22, n.9, p. 1052–1070, Medford, 2013.

CERQUEIRA, C. A. Determinação de fatores ligados às taxas de distorção idade/série, taxa de evasão escolar e taxa de repetência. *In*: RIOS-NETO, L. G.; RIANI, J. de L. R. (Org.). **Associação Brasileira de Estudos Populacionais**. Parte IV, São Paulo, 2004.

CHERNOZHUKOV, V.; FERNANDEZ-VAL, I; MELLY, B. Inference on counterfactual distributions. **Econometrica**. v. 81, n.6, p. 2205-2268, New York, 2013.

COLEMAN, J. *et al.* Equality of educational opportunity. **Government Printing Office**, Washington, 1966.

DATTA, N. evaluating impacts of watershed development program on agricultural productivity, income, and livelihood in bhalki watershed of bardhaman district, west bengal. **World Development**. v. 66, p. 443-456, Washington, 2015.

DEHEJIA, R. H.; WAHBA, S. Propensity score matching methods for non-experimental causal studies. **Review of Economics and Statistics**, v. 84, p. 151–161, Cambridge, 2002.

DINARDO, J; FORTIN, N. M.; LEMIEUX, T. Labor market institutions and the distribution of wages, 1973-1992: A Semiparametric Approach. **Econometrica**, v. 64, p. 1001-1044, Chicago, 1996.

DONAHUE, D.J. Human capital and income inequality. *In*: O'BRIEN, D.J.; VEGREN, S.K. (Org.), **Rural Reform in Post-Soviet Russia**, Washington, 2002.

DONALD, S. G.; GREEN, D. A.; PAARSCH, H. P. Differences in wage distributions between Canada and the united states: an application of a flexible estimator of distribution functions in the presence of covariates source. **Review of Economic Studies**, v. 67, p. 609-633, Chicago, 2000.

EDOKA; I. P. **Decomposing differences in cotinine distribution between children and adolescents from different socioeconomic backgrounds**. HEDG. Working Paper, v. 12 n.29, York: University of York, 2012.

FARIA, E. M.; GUIMARÃES. R. R. de M. Excelência com equidade: fatores escolares para o sucesso educacional em circunstâncias desfavoráveis, **Estudos em Avaliação, Educacional**. v. 26, n. 61, p. 192-215, São Paulo, 2015.

FIRPO, S.; FORTIN, N.; LEMIEUX, T. Decomposing wage distributions using recentered influence function regressions. **Mimeo**. Columbia: University of British Columbia, 2007.

FIRPO, S.; FORTIN, N.; LEMIEUX, T. Unconditional Quantile Regressions, **Econometrica**, v. 77, n.3, p. 953-973, Chicago: University of Chicago, 2009.

FISHER, J.; MARCHAND, J. Does the retirement consumption puzzle differ across the distribution? **The Journal of Economic Inequality**, v.12, p. 279-296, New York, 2014.

FORTIN, N M. The Gender Wage Gap among Young Adults in the United States: The Importance of Money vs. People. **Journal of Human Resources**. v.43, p. 886-920, Madison, 2008.

FORTIN, N. M; LEMIEUX, T. Rank Regressions, Wage Distributions, and the Gender Gap. **Journal of Human Resources**. v. 33, p. 610-643, Madison, 1998.

FOURNIER, J.; KOSKE, I. Less income inequality and more growth—Are they compatible? The drivers of labour earnings inequality—An analysis based on conditional and unconditional quantile regressions. **OECD Economics Department Working Papers**. n. 930, New York, 2012.

HAMPEL, F. R et al. **Robust Statistics: The Approach Based on Influence Functions**. New York: John Wiley and Sons, 1986.

HANUSHEK, E. A. Schooling, educational achievement, and the Latin American growth puzzle. **Journal of Development Economics**, Issue 99, p. 497–512, Washington, 2012.

_____. Economic growth in developing countries: The role of human capital. **Economics of Education Review**, v. 37, p. 204-212, Wollongong: University of Wollongong, 2013.

_____. Incentives for Efficiency and Equity in the School System. **Perspektiven der Wirtschaftspolitik**, v. 9, p. 5-27, Barcelona, 2008.

HANUSHEK, E. A.; WÖBMAN, L. **Education Quality and Economic Growth**. The International Bank for Reconstruction and Development. The World Bank. Washington, 2007.

_____. Education and Economic Growth. **International Encyclopedia of Education**, v. 2, p. 245-252, Amsterdam: Elsevier, 2010.

HANUSHEK, E. A.; WOESSMANN, L. The Economics of International Differences in

Educational Achievement. *In*: HANUSHEK, E. A.; MACHIN S.; WOESSMANN L. **HandBooks in Economics of Education**, v. 3. p. 89-200, Amsterdam, 2011.

HELFAND; S, M.; PEREREIRA; V, de F. O capital humano no Brasil rural: o déficit educacional, *In*: MIRANDA, C.; TIBURCIO, B. **A nova cara da pobreza rural: desafios para as políticas públicas: IICA**, (Série desenvolvimento rural sustentável), v.16, Brasília, 2012.

IACUS, S. M.; KING, G; PORRO, G. Causal Inference without Balance Checking: Coarsened Exact Matching. **Political Analysis**, v. 20, p. 1-24. Cambridge: Cambridge University Press, 2011.

_____. **Matching for causal inference without balance checking**. Working paper series. Cambridge, 2008. Disponível em: <<http://gking.harvard.edu/files/cem.pdf>>. Acesso em: dezembro de 2016.

IBGE. **Pesquisa Nacional por Amostra de Domicílios, 2015**. Rio de Janeiro, 2016. Disponível em: <<http://www.ibge.gov.br>>. Acesso em: outubro de 2016.

INEP. **Microdados do Saeb (Aneb/Prova Brasil), 2015**. Disponível em: <<http://portal.inep.gov.br/microdados>>. Acesso em: dezembro de 2016.

_____. **Brasil no PISA 2015**. Sumário Executivo. Brasília, 2016. Disponível em: <http://download.inep.gov.br/acoes_internacionais/pisa/documentos/2016/pisa_brasil_2015_sumario_executivo.pdf>. Acesso em: outubro de 2016.

_____. **Indicador de Nível Socioeconômico (Inse) das Escolas**. Brasília, 2014a, Disponível em: <http://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2011_2013/nivel_socioeconomico/nota_tecnica_indicador_nivel_socioeconomico.pdf>, Acesso em: janeiro de 2017.

_____. **Índice de Regularidade Docente**. Brasília, 2014b, Disponível em: <http://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_regularidade_vinculo/nota_tecnica_indicador_regularidade_2015.pdf>, Acesso em: janeiro de 2017.

_____. **O Indicador de Esforço Docente**. Brasília, 2014c Disponível em: <http://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_esforco/nota_tecnica_indicador_docente_esforco.pdf>, Acesso em: janeiro de 2017.

_____. **Indicador de adequação da formação do docente da educação básica**, Brasília, 2014d, Disponível em: <http://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_formacao_legal/nota_tecnica_indicador_docente_formacao_legal.pdf>, Acesso em: janeiro de 2017.

IPEA. **PNAD 2009 - Primeiras análises**: Situação da educação brasileira - avanços e problemas. Comunicados do Ipea. Brasília, 2010.

JANN, B. The Oaxaca-Blinder Decomposition for Linear Regression Models, **Stata Journal**, v. 8, p. 435-479, New York, 2008.

JUHN, C.; MURPHY, K. M.; PIERCE, B. Wage Inequality and the Rise in Returns to Skill. **Journal of Political Economy**, v. 101, p. 410-442, Chicago, 1993.

KASSENBOEHMER, S.C.; SINNING, M.G. Distributional changes in the gender wage gap. **Industrial & Labor Relations Review**, v. 67, p. 335-361, Ithaca, 2014.

LAVOR, D. C.; ARRAES, R. de A. de. Qualidade da educação básica e uma avaliação de política educacional para o Ceará. *In*: X Encontro Economia do Ceará em Debate, 2014, Fortaleza. **Anais...** Fortaleza: Instituto de Pesquisa e Estratégia Econômica do Ceará – IPECE, 2014.

LE, H.T.; BOOTH, A.L. Inequality in Vietnamese Urban–Rural Living Standards, 1993–2006. **Review of Income and Wealth**, v. 60, p. 862–886, Medford, 2013.

LOMBARDI-FILHO, S. C.; CRUZ, M. S. da; IRFFI, G. D. Estimação da Desigualdade de Contingência entre os Estudantes do Ensino Fundamental da rede Pública de João Pessoa – PB. *In*: XX Fórum BNB de Desenvolvimento. XIX Encontro Regional de Economia, 2014, Fortaleza. **Anais...** Fortaleza: Banco do Nordeste – BNB, 2014

LOUNKAEW, K. Explaining urban–rural differences in educational achievement in Thailand: Evidence from PISA literacy data. **Economics of Education Review**, v.37. p. 213–225, Amsterdam, 2013.

MACHADO, J. F.; MATA, J. Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. **Journal of Applied Econometrics**, v.20, p. 445-465, Medford, 2005.

MELLY, B. Decomposition of differences in distribution using quantile regression. **Labour economics**, v.12, p. 577-590, Amsterdam, 2005.

MORIN, L.P. Do Men and Women Respond Differently to Competition? Evidence from a Major Education Reform. **Journal of Labor Economics**, v. 33, p. 443-491, Amsterdam, 2015.

NEY, M. G; SOUSA, P. M de; PONCIANO, N. J. Desigualdade de acesso à educação e evasão escolar entre ricos e pobres no Brasil rural e urbano. **Revista Científica Internacional Indexada**. v. 13. p. 33-55l, Rio de Janeiro, 2010.

NGUYEN; H. T. **The Evolution of The Gender Test Score Gap Through Seventh Grade**: New Insights From Australia Using Quantile Regression and Decomposition. Working Paper. v. 15/7, Bentley: Bankwest Curtin Economics Centre, 2015.

NIETO, S.; RAMOS, R. Decomposition of Differences in PISA Results in Middle Income Countries. **Working Paper** 2014/08, p. 38. Barcelona: Institut de Recerca en Economia Aplicada Regional i Pública, 2014.

OAXACA, R. L. Male-Female Wage Differentials in Urban Labor Markets. **International Economic Review**, v. 14, p. 693-709, Medford, 1973.

OCDE - **Organização para a Cooperação e Desenvolvimento Econômico. Programme for International Student Assessment (PISA) - 2015.** Resultados Clave, 2016. Paris, 2016.

OSORIO, F. B.; MORENO, V. A. G.; PATRINOS, H. A.; PORTA, E. E. Using the Oaxaca-Blinder decomposition technique to analyze learning outcomes changes over time: an application to Indonesia's results in PISA mathematics. **World Bank Policy Research Working Paper Series**, v. 5584, Washington 2011.

PALERMO, G.A.; SILVA, D.B.N.; NOVELLINO, M.S.F. Fatores associados ao desempenho escolar: uma análise da proficiência em matemática dos alunos do 5º ano do ensino fundamental da rede municipal do Rio de Janeiro. **R. bras. Est. Pop.** v. 31, n.2, p. 367-394, Rio de Janeiro, 2014.

PASSADOR, C.S.; LOPES, J. E. F. Educação do campo no Estado de São Paulo: análise do nível de ruralidade das escolas no desempenho escolar. **Revista do Serviço Público.** v. 65, p. 87-113, Brasília, 2014.

RAMOS, R.; DUQUE J. C.; NIETO, S. Decomposing the Rural-Urban Differential in Student Achievement in Colombia using PISA Microdata. **Estudios de Economía Aplicada.** v. 34 -2. p. 379 – 412, Santiago, 2016.

RIANI, J. de L. R.; RIOS-NETO, E. L. G. *Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros?* **R. bras. Est. Pop.**, v. 25, n. 2, p. 251-269, São Paulo, 2008.

RIVKIN, S.; HANUSHEK, E.; KAIN, J. Teachers, Schools, and Academic Achievement. **Econometrica**, v. 73, n. 2, p. 417–458, Chicago: University of Chicago, 2005.

RODRIGUES, C. G. **A Relação entre a expansão do acesso ao ensino e o desempenho escolar no Brasil:** evidências com base no SAEB para o período de 1997 a 2005. Belo Horizonte: UFMG/Cedeplar, 2009. 202 f. Tese (Doutorado) - Centro de Desenvolvimento e Planejamento Regional da Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

RUBIL; I. The Great Recession and the Public-Private Wage Gap: Distributional Decomposition Evidence from Croatia 2008-2011. **Munich Personal RePEc Archive paper**, n. 46798, Munich, 2013.

SCHURER, S. *et al.* The Medical Care Costs of Mood Disorders: A Coarsened Exact Matching Approach. **IZA Discussion Paper**, n. 8814, Bonn, 2015.

SILVA, N. V.; HASENBALG, C. Recursos familiares e transições educacionais. **Cad. Saúde Pública**. v. 18, p. 67-76, Rio de Janeiro, 2002.

SILVA, C. B. da.; KICHOW, I. V. T. Transição escola rural para a escola urbana e seus reflexos no ensino de matemática: um caso na cidade de laguna Carapã. *In*: XI Encontro Nacional de Educação Matemática, Curitiba. **Anais...** Curitiba: Sociedade Brasileira de Educação Matemática, 2013.

SOUZA *et al.* Fatores associados ao fluxo escolar no ingresso e ao longo do ensino médio no Brasil. **Pesquisa e Planejamento Econômico – PPE**. v. 42, n. 1, Brasília: IPEA, 2012

SOARES, S. **Aprendizado e Seleção**: Uma Análise da Evolução Educacional Brasileira de Acordo com uma Perspectiva de Ciclo de Vida. Texto para Discussão nº 1185, Brasília: IPEA, 2006.

SOARES, S. S. D.; FONTOURA, N. D. O.; PINHEIRO, L. Tendências recentes na escolaridade e no rendimento de negros e de brancos. *In*: BARROS, R. P. D.; FOGUEL, M. N.; ULYSSEA, G. **Desigualdade de renda no Brasil**: uma análise da queda recente. v. II, p. 401-415, Brasília: IPEA, 2007

SOARES, S.; RAZO, R.; FARIÑAS, M. Perfil Estatístico da Educação Rural: Origem Socioeconômica Desfavorecida, Insumos Escolares Deficientes e Resultados Inaceitáveis. *In*: BOF, A. M. **A educação no Brasil rural**. Brasília: INEP, 2006.

SOARES-NETO, J. J. *et al.* Uma escala para medir a infraestrutura escolar. **Estudos em Avaliação Educacional**. v. 24, n. 54, p. 78-99, São Paulo, 2013.

VALENZUELA, J. P. *et al.* ¿Qué explica las diferencias de resultados PISA Matemática entre Chile y algunos países de la OCDE y América Latina? **Estudios Internacionales, SIMCE**. p.105-148, Santiago, 2009.

WÖBMANN, L. Schooling resources, educational institutions and student performance: the international evidence. **Oxford bulletin of economics and statistics**, v. 65, n. 2, p. 117-170, Malden, 2003.

APÊNDICE 1

Tabela A.1 - Estimativas da Regressão Quantílica Incondicional por Zona – Português

Classe	Urbano		Rural	
	Freq.	%	Freq.	%
A1	3438	0.48	109	0.15
A2	42948	6.04	1417	1.93
B1	140394	19.75	6073	8.27
B2	234369	32.98	16049	21.85
C1	178321	25.09	20376	27.75
C2	82362	11.59	17246	23.48
D	26423	3.72	10094	13.74
E	2425	0.34	2076	2.83

Fonte: RODRIGUES (2017).

Tabela A.1.1 - Estimativas da Regressão Quantílica Incondicional por Zona – Português

Variáveis	MQO ¹		q10		q50		q90	
	Rural	Urbano	Rural	Urbano	Rural	Urbano	Rural	Urbano
idade	0,00130 (0,000961)	0,00302*** (0,000398)	0,000878 (0,00226)	-0,00234** (0,000912)	-0,000678 (0,00142)	0,00314*** (0,000466)	0,00487*** (0,00116)	0,00914*** (0,000476)
sexo	-0,0202*** (0,00159)	-0,0203*** (0,000497)	-0,00510* (0,00289)	-0,00905*** (0,00102)	-0,0250*** (0,00283)	-0,0277*** (0,000696)	-0,0253*** (0,00276)	-0,0180*** (0,000745)
branco	-0,00202 (0,00183)	0,0169*** (0,000541)	-0,0170*** (0,00318)	-0,00338*** (0,00106)	-0,00901*** (0,00245)	0,0219*** (0,000699)	0,0247*** (0,00341)	0,0304*** (0,000818)
nunca_reprov	0,0657*** (0,00205)	0,0854*** (0,000751)	0,0255*** (0,00384)	0,0631*** (0,00183)	0,0799*** (0,00342)	0,112*** (0,00116)	0,0803*** (0,00283)	0,0620*** (0,000670)
nunca_aband	0,0339*** (0,00289)	0,0347*** (0,00124)	0,0721*** (0,00695)	0,147*** (0,00343)	0,0466*** (0,00436)	0,00466*** (0,00145)	-0,0174*** (0,00269)	-0,0256*** (0,000857)
idade_certa	0,0570*** (0,00285)	0,0541*** (0,00111)	0,0489*** (0,00556)	0,0675*** (0,00268)	0,0695*** (0,00402)	0,0575*** (0,00134)	0,0440*** (0,00388)	0,0323*** (0,00114)
atitu_leitura	0,00254 (0,00161)	0,00206*** (0,000495)	0,00347 (0,00253)	-0,0122*** (0,000951)	-0,00268 (0,00213)	-0,000447 (0,000699)	0,0105*** (0,00239)	0,0221*** (0,000682)
possui_pc	0,0186*** (0,00188)	0,0296*** (0,000582)	-0,00424 (0,00322)	0,0194*** (0,00126)	0,0216*** (0,00310)	0,0366*** (0,000803)	0,0370*** (0,00314)	0,0281*** (0,000681)
faz_dever1	0,0519*** (0,00174)	0,0522*** (0,000566)	0,0379*** (0,00324)	0,0511*** (0,00134)	0,0637*** (0,00263)	0,0592*** (0,000839)	0,0480*** (0,00267)	0,0403*** (0,000831)
trab_fora	-0,0926*** (0,00193)	-0,116*** (0,000745)	-0,0767*** (0,00445)	-0,158*** (0,00224)	-0,122*** (0,00319)	-0,128*** (0,00114)	-0,0661*** (0,00251)	-0,0507*** (0,000703)
SEF	0,00380*** (0,000163)	0,00264*** (5,40e-05)	0,00340*** (0,000312)	0,00305*** (0,000112)	0,00460*** (0,000226)	0,00314*** (6,94e-05)	0,00288*** (0,000272)	0,00118*** (6,92e-05)
Edu_pai2	-0,00246 (0,00233)	0,00477*** (0,000750)	0,00249 (0,00417)	0,0113*** (0,00162)	-0,00654* (0,00340)	0,00458*** (0,00119)	-0,00265 (0,00382)	0,000380 (0,000974)
Edu_pai3	0,0308*** (0,00285)	0,0405*** (0,000728)	0,0182*** (0,00387)	0,0368*** (0,00131)	0,0353*** (0,00371)	0,0475*** (0,00113)	0,0422*** (0,00603)	0,0319*** (0,00112)
Edu_pai4	-0,0174*** (0,00313)	-0,000831 (0,000817)	0,000827 (0,00502)	0,00875*** (0,00150)	-0,0252*** (0,00453)	-0,00495*** (0,00114)	-0,0254*** (0,00524)	-0,000578 (0,00105)
Edu_mae2	0,00841*** (0,00240)	0,0213*** (0,000782)	0,00287 (0,00386)	0,0203*** (0,00174)	0,0102*** (0,00374)	0,0243*** (0,00105)	0,0130*** (0,00402)	0,0148*** (0,000861)
Edu_mae3	0,0485*** (0,00263)	0,0611*** (0,000718)	0,0206*** (0,00397)	0,0528*** (0,00129)	0,0596*** (0,00352)	0,0725*** (0,000907)	0,0582*** (0,00495)	0,0481*** (0,00109)

Continua...								
Edu_mae4	0,00274 (0,00314)	0,0215*** (0,000850)	-0,00669 (0,00449)	0,0205*** (0,00156)	0,00867** (0,00429)	0,0205*** (0,00105)	0,00331 (0,00582)	0,0222*** (0,00126)
mora_mae_pai	0,00813*** (0,00170)	0,0130*** (0,000516)	0,00954*** (0,00312)	0,00781*** (0,00104)	0,00455* (0,00234)	0,0144*** (0,000750)	0,0117*** (0,00292)	0,0163*** (0,000679)
pais_incentiva	0,0712*** (0,00287)	0,0816*** (0,00121)	0,0722*** (0,00679)	0,141*** (0,00374)	0,0951*** (0,00377)	0,0825*** (0,00159)	0,0382*** (0,00353)	0,0194*** (0,00117)
Prof_superior	0,0791*** (0,00285)	0,0921*** (0,00135)	0,0688*** (0,00517)	0,112*** (0,00315)	0,0989*** (0,00410)	0,101*** (0,00177)	0,0586*** (0,00461)	0,0514*** (0,00160)
experiencia	-0,00116 (0,00259)	0,00480*** (0,00113)	-1,15e-05 (0,00530)	-0,00334 (0,00258)	-0,00243 (0,00348)	0,00718*** (0,00163)	-0,00291 (0,00425)	0,00988*** (0,00170)
salario	0,000496 (0,00234)	0,00536*** (0,000716)	-1,22e-05 (0,00412)	0,0176*** (0,00144)	0,00691* (0,00369)	0,00434*** (0,000969)	-0,00524 (0,00464)	-0,00416*** (0,00104)
rede_munic	0,00603* (0,00329)	-0,00459*** (0,000594)	0,00593 (0,00534)	0,00334*** (0,00106)	0,00680 (0,00492)	-0,00610*** (0,000750)	0,00755 (0,00630)	-0,00952*** (0,000945)
Tam_escola	-2,66e-05*** (3,19e-06)	-1,36e-05*** (6,45e-07)	-1,60e-05*** (4,51e-06)	-2,15e-05*** (1,32e-06)	-2,74e-05*** (4,99e-06)	-1,38e-05*** (9,43e-07)	-2,64e-05*** (4,60e-06)	-6,10e-06*** (8,68e-07)
aluno_prof	-0,00110*** (0,000118)	-0,000212*** (3,69e-05)	-8,35e-05 (0,000227)	0,000113 (7,66e-05)	-0,00133*** (0,000192)	-0,000410*** (5,26e-05)	-0,00159*** (0,000197)	-0,000246*** (5,29e-05)
m_duracao_aula	1,74e-05 (1,43e-05)	9,48e-05*** (4,69e-06)	5,34e-05** (2,47e-05)	6,15e-05*** (8,59e-06)	1,81e-05 (2,23e-05)	0,000112*** (6,64e-06)	1,69e-07 (2,51e-05)	9,77e-05*** (6,90e-06)
escass_material	-0,00380** (0,00162)	-0,00860*** (0,000506)	0,00110 (0,00297)	-0,00732*** (0,00103)	-0,000643 (0,00242)	-0,00908*** (0,000678)	-0,0122*** (0,00300)	-0,00898*** (0,000731)
falta_de_prof	-0,0200*** (0,00172)	-0,00689*** (0,000522)	-0,0108*** (0,00315)	-0,00411*** (0,00119)	-0,0238*** (0,00260)	-0,00864*** (0,000698)	-0,0268*** (0,00291)	-0,00706*** (0,000793)
sel_diretor	0,00920*** (0,00214)	-0,000393 (0,000532)	0,0103*** (0,00258)	0,00458*** (0,00104)	0,0129*** (0,00326)	-0,000772 (0,000825)	0,00337 (0,00358)	-0,00342*** (0,000781)
IRD	-0,00952*** (0,00124)	0,00265*** (0,000450)	-0,00681*** (0,00199)	0,00131 (0,000917)	-0,0135*** (0,00171)	0,00336*** (0,000661)	-0,00885*** (0,00182)	0,00450*** (0,000615)
IMCE	-0,0387*** (0,00348)	-0,0249*** (0,00133)	-0,0329*** (0,00592)	-0,0225*** (0,00284)	-0,0450*** (0,00484)	-0,0302*** (0,00201)	-0,0349*** (0,00567)	-0,0172*** (0,00168)
IIE	0,157*** (0,00554)	0,140*** (0,00235)	0,126*** (0,00884)	0,172*** (0,00491)	0,195*** (0,00922)	0,154*** (0,00348)	0,123*** (0,00923)	0,0772*** (0,00307)
Intercepto	4,825*** (0,0149)	4,729*** (0,00609)	4,501*** (0,0312)	4,281*** (0,0147)	4,764*** (0,0200)	4,713*** (0,00784)	5,234*** (0,0190)	5,204*** (0,00734)

Continua...								
Obs.	73.440	710.680	73.440	710.680	73.440	710.680	73440	710.680
r2	0,257	0,264	0,064	0,104	0,199	0,190	0,089	0,064
F	874,11	8329,36	126,04	1531,26	755,62	7231,91	186,96	1661,64

Fonte: RODRIGUES (2017) com base nos microdados do Censo Escolar e Prova Brasil/Saeb, 2015.

Erros padrões entre parênteses, * p < 0,05, ** p < 0,01, *** p < 0,001.

Nota: ¹ Estimação robusta para heterocedasticidade.

Tabela A.1.2 - Estimativas da Regressão Quantílica Incondicional por Zona – Matemática

Variáveis	MQO		q10		q50		q90	
	Rural	Urbano	Rural	Urbano	Rural	Urbano	Rural	Urbano
idade	-4,93e-05 (0,000828)	0,000659* (0,000350)	0,000535 (0,00184)	-0,00183** (0,000819)	-0,00130 (0,000960)	0,000899** (0,000410)	0,00247* (0,00142)	0,00309*** (0,000480)
sexo	0,0298*** (0,00141)	0,0325*** (0,000450)	0,0208*** (0,00203)	0,0265*** (0,000738)	0,0319*** (0,00204)	0,0377*** (0,000603)	0,0364*** (0,00293)	0,0304*** (0,000643)
branco	0,00521*** (0,00161)	0,0173*** (0,000494)	-0,00943*** (0,00290)	-0,00272*** (0,000849)	0,00332 (0,00205)	0,0219*** (0,000763)	0,0263*** (0,00319)	0,0305*** (0,000723)
nunca_reprov	0,0596*** (0,00176)	0,0798*** (0,000659)	0,0234*** (0,00341)	0,0580*** (0,00168)	0,0639*** (0,00216)	0,103*** (0,000921)	0,0856*** (0,00292)	0,0645*** (0,000717)
nunca_aband	0,0167*** (0,00250)	0,0154*** (0,00107)	0,0399*** (0,00567)	0,0768*** (0,00315)	0,0219*** (0,00355)	0,00498*** (0,00149)	-0,0183*** (0,00264)	-0,0223*** (0,000888)
idade_certa	0,0436*** (0,00246)	0,0457*** (0,000970)	0,0344*** (0,00539)	0,0504*** (0,00246)	0,0495*** (0,00293)	0,0553*** (0,00127)	0,0451*** (0,00374)	0,0272*** (0,00105)
atitu_leitura	0,00447*** (0,00142)	0,00208*** (0,000448)	0,00655*** (0,00230)	-0,00599*** (0,000893)	0,00207 (0,00197)	0,000948 (0,000659)	0,0101*** (0,00283)	0,0122*** (0,000612)
possui_pc	0,0234*** (0,00166)	0,0261*** (0,000526)	0,00552** (0,00250)	0,0186*** (0,00116)	0,0242*** (0,00241)	0,0326*** (0,000763)	0,0441*** (0,00379)	0,0229*** (0,000769)
faz_dever2	0,0380*** (0,00162)	0,0363*** (0,000551)	0,0233*** (0,00296)	0,0390*** (0,00116)	0,0439*** (0,00216)	0,0412*** (0,000860)	0,0412*** (0,00255)	0,0254*** (0,000681)
trab_fora	-0,0635*** (0,00169)	-0,0830*** (0,000665)	-0,0475*** (0,00347)	-0,104*** (0,00170)	-0,0738*** (0,00242)	-0,0968*** (0,00110)	-0,0586*** (0,00271)	-0,0452*** (0,000749)
SEF	0,00345*** (0,000143)	0,00372*** (4,87e-05)	0,00244*** (0,000250)	0,00389*** (0,000108)	0,00402*** (0,000185)	0,00459*** (6,28e-05)	0,00315*** (0,000274)	0,00227*** (6,81e-05)

Continua...

sel_diretor	0,0109*** (0,00190)	0,000776 (0,000485)	0,00722** (0,00284)	0,00484*** (0,000847)	0,0146*** (0,00279)	0,00142** (0,000669)	0,0119*** (0,00408)	-0,00472*** (0,000767)
IRD	-0,00517*** (0,00109)	0,00607*** (0,000407)	-0,00173 (0,00160)	0,00150** (0,000759)	-0,00701*** (0,00169)	0,00695*** (0,000534)	-0,00701*** (0,00208)	0,00818*** (0,000637)
IMCE	-0,0403*** (0,00304)	-0,0314*** (0,00119)	-0,0224*** (0,00616)	-0,0242*** (0,00265)	-0,0406*** (0,00457)	-0,0376*** (0,00172)	-0,0586*** (0,00554)	-0,0305*** (0,00137)
IIE	0,136*** (0,00487)	0,127*** (0,00212)	0,106*** (0,00881)	0,143*** (0,00440)	0,157*** (0,00639)	0,163*** (0,00242)	0,132*** (0,00941)	0,0593*** (0,00333)
Constant	4,963*** (0,0129)	4,843*** (0,00540)	4,733*** (0,0271)	4,534*** (0,0137)	4,915*** (0,0181)	4,758*** (0,00722)	5,257*** (0,0235)	5,268*** (0,00809)
Obs.	73.440	710.680	73.440	710.680	73.440	710.680	73.440	710.680
r2	0,217	0,238	0,044	0,082	0,159	0,176	0,088	0,064
F	686,42	7468,03	94,72	1290,68	538,52	6355,41	180,03	1650,19

Fonte: RODRIGUES (2017) com base nos microdados do Censo Escolar e Prova Brasil/Saeb, 2015.

Erros padrões entre parênteses, * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

Nota: ¹ Estimação robusta para heterocedasticidade.

Tabela A.1.3 - Resultados de decomposição de FIR para Português e Matemática

Português							
Estimativa	Média esperada			Diferença entre grupos			
	Urbano	Rural	Diferencial	Características	%	Estrutural	%
q10	4,886	4,792	0,094*	0,054*	58%	0,040*	42%
q20	4,988	4,873	0,115*	0,068*	59%	0,047*	41%
q30	5,066	4,933	0,133*	0,085*	64%	0,048*	36%
q40	5,134	4,983	0,151*	0,110*	73%	0,041*	27%
q50	5,201	5,036	0,165*	0,125*	75%	0,041*	25%
q60	5,269	5,091	0,179*	0,149*	83%	0,030*	17%
q70	5,342	5,162	0,180*	0,160*	89%	0,020*	11%
q80	5,423	5,246	0,177*	0,181*	102%	-0,004*	-2%
q90	5,525	5,371	0,154*	0,192*	125%	-0,038*	-25%

Matemática							
Estimativa	Média esperada			Diferença entre grupos			
	Urbano	Rural	Diferencial	Características	%	Estrutural	%
q10	5,013	4,938	0,074*	0,051*	69%	0,023*	31%
q20	5,097	5,010	0,087*	0,060*	69%	0,027*	31%
q30	5,161	5,065	0,097*	0,073*	76%	0,023*	24%
q40	5,217	5,111	0,106*	0,080*	76%	0,026*	24%
q50	5,271	5,158	0,113*	0,089*	79%	0,024*	21%
q60	5,326	5,203	0,123*	0,101*	82%	0,022*	18%
q70	5,388	5,256	0,131*	0,109*	83%	0,022*	17%
q80	5,463	5,321	0,142*	0,122*	86%	0,020*	14%
q90	5,565	5,415	0,150*	0,163*	109%	-0,013*	-9%

Fonte: RODRIGUES (2017) com base nos microdados do Censo Escolar e Prova Brasil/Saeb, 2015.

(*) significante a 1%; (**) significante a 5%; significante a 10%.

Tabela A.1.4 - Decomposição do efeito característico Português e Matemática

Português								
Quantil	Aluno	%	Família	%	Professores	%	Escola	%
q10	0,002**	4%	0,019*	34%	0,002	3%	0,032*	59%
q20	0,003*	5%	0,025*	36%	0,006*	9%	0,034*	50%
q30	0,008*	10%	0,026*	30%	0,017*	21%	0,033*	39%
q40	0,007*	6%	0,032*	29%	0,025*	23%	0,046*	42%
q50	0,011*	9%	0,036*	29%	0,027*	22%	0,050*	40%
q60	0,015*	10%	0,045*	30%	0,028*	19%	0,061*	41%
q70	0,020*	13%	0,047*	30%	0,029*	18%	0,063*	40%
q80	0,031*	17%	0,045*	25%	0,032*	18%	0,073*	40%
q90	0,034*	18%	0,049*	25%	0,031*	16%	0,079*	41%

Matemática								
Quantil	Aluno	%	Família	%	Professores	%	Escola	%
q10	0,005*	9%	0,015*	28%	0,010*	20%	0,022*	43%
q20	0,005*	8%	0,019*	32%	0,012*	20%	0,024*	39%
q30	0,005*	7%	0,026*	35%	0,016*	22%	0,027*	36%
q40	0,007*	8%	0,026*	33%	0,016*	21%	0,031*	38%
q50	0,011*	12%	0,025*	28%	0,017*	20%	0,036*	40%
q60	0,014*	13%	0,024*	24%	0,017*	17%	0,046*	46%
q70	0,018*	16%	0,023*	21%	0,020*	18%	0,049*	45%
q80	0,021*	18%	0,025*	20%	0,024*	19%	0,052*	43%
q90	0,028*	17%	0,034*	21%	0,031*	19%	0,071*	43%

Fonte: RODRIGUES (2017) com base nos microdados do Censo Escolar e Prova Brasil/Saeb, 2015.

Nota (*) significante a 1%; (**) significante a 5%.

Tabela A.1.5 - Decomposição do efeito estrutural - Português e Matemática

Português										
Quantil	Aluno	%	Fam.	%	Prof.	%	Escola	%	Constante	%
q10	-0.083*	-211%	0.049*	124%	0.055*	140%	0.081*	205%	-0.063*	-158%
q20	-0.114*	-243%	0.026*	57%	0.046*	98%	0.061*	131%	0.027	57%
q30	-0.133*	-276%	0.042*	87%	0.022*	46%	0.009	19%	0.109*	224%
q40	-0.039*	-94%	0.034*	82%	0.002*	23%	0.088*	213%	-0.051*	-125%
q50	-0.050*	-123%	0.028*	69%	0.007*	17%	0.051*	127%	0.004*	10%
q60	0.027	91%	0.006	22%	0.010*	35%	0.011	36%	-0.025	-83%
q70	0.036**	178%	-0.024*	-116%	0.007	32%	0.005	24%	0.00	-18%
q80	0.002	-60%	-0.038*	933%	0.002	-44%	0.018	-436%	0.01188	-293%
q90	0.017	-44%	-0.071*	188%	0.000	-1%	-0.023	60%	0.03883	-102%
Matemática										
Quantil	Aluno	%	Fam.	%	Prof.	%	Escola	%	Constante	%
q10	-0.050*	-216%	0.061*	264%	0.016*	69%	0.061*	263%	-0.065*	-279%
q20	0.046*	171%	0.062*	228%	0.006**	22%	0.084*	313%	-0.171*	-635%
q30	0.016	69%	0.033*	144%	0.009*	38%	0.073*	316%	-0.108*	-467%
q40	-0.032**	-123%	0.047*	183%	0.009*	33%	0.102*	397%	-0.100*	-390%
q50	0.000	0%	0.057*	238%	0.009*	39%	0.067*	284%	-0.109*	-461%
q60	0.060*	268%	0.046*	206%	0.014*	61%	0.029*	131%	-0.127*	-566%
q70	0.073*	324%	0.054*	239%	0.012*	54%	0.015	69%	-0.132*	-586%
q80	0.024	122%	0.033*	163%	0.019*	96%	0.023	114%	-0.079*	-395%
q90	0.034	-256%	-0.029*	224%	0.015*	-114%	-0.022	165%	-0.011	81%

Fonte: RODRIGUES (2017) com base nos microdados do Censo Escolar e Prova Brasil/Saeb, 2015.

Nota (*) significativa a 1%; (**) significativa a 5%.

APÊNDICE 2

Tabela A.2.1 - Descritiva dos alunos matriculados na 8ª série ou 9º ano do Ensino Fundamental - 2011

Idade	Média	15.79
	Desvio Padrão	3,23
Sexo	Feminino	137.852 (52,03%)
	Masculino	127.070 (47,97%)
Rede de ensino	Estadual	54.002 (20,3%)
	Municipal	209.631 (79,13%)
	Privada	1.289 (0,49%)
Turno	Diurno	236.306 (89,2%)
	Noturno	28.616 (10,8%)

Fonte: RODRIGUES (2017) com base nos dados do Censo Escolar de 2011

Nota: Entre parênteses valores percentuais.

Tabela A.2.2 – Evasão Escolar no Ensino Médio de 2012 a 2014

Ano	Variável		Matriculas	Evasão	Reprovados	% de alunos evadidos	% de alunos reprovados
2012	Turno	Diurno	131.217	15.155	144.89	11,55%	11,04%
		Noturno	53.815	8.584	5.570	15,95%	10,35%
	Sexo	Feminino	99.468	11.671	8.402	11,73%	8,45%
		Masculino	85.564	12.068	11.657	14,10%	13,62%
	Zona	Rural	55.269	6.277	5.059	11,36%	9,15%
Urbana		129.763	17.462	15.000	13,46%	11,56%	
2013	Turno	Diurno	107.932	11.136	8.197	10,32%	7,59%
		Noturno	53.361	8.481	4.106	15,89%	7,69%
	Sexo	Feminino	88.111	9.342	6.705	10,60%	7,61%
		Masculino	73.182	10.275	5.598	14,04%	7,65%
	Zona	Rural	46.564	5.055	3.069	10,86%	6,59%
Urbana		114.729	14.562	9.650	12,69%	8,41%	
2014	Turno	Diurno	89.364	2.301	5.508	2,57% ¹	6,16%
		Noturno	523.12	2.235	2.412	4,27% ¹	4,61%
	Sexo	Feminino	78.752	1.378	4.810	1,75% ¹	6,11%
		Masculino	62.924	2.030	3.110	3,23% ¹	4,94%
	Zona	Rural	39.830	1.079	1.836	2,71% ¹	4,61%
Urbana		101.846	3.457	6.084	3,39% ¹	5,97%	

Fonte: RODRIGUES (2017) com base do Censo Escolar de 2012, 2013, 2014 e 2015

Nota: (1) somente para alunos da 1ª e 2ª série do EM

Tabela A.2.1. Índice ABEP adaptado o Enem

Escolaridade		Pontos		
Escolaridade do chefe de família ¹	Analfabeto ou primeira fase do ensino fundamental incompleta	0		
	Primeira fase do ensino fundamental completa ou segunda fase incompleta	1		
	Segunda fase do ensino fundamental completa ou ensino médio incompleto	2		
	Ensino médio completo ou superior incompleto	4		
	Superior completo	8		
Itens	Quantidade de Itens e pontos			
	0	1	2	3 ou +
Televisão em cores	0	1	2	4
Rádio	0	1	2	4
Banheiro	0	4	5	7
Automóvel	0	4	7	9
Empregada mensalista	0	3	4	4
Máquina de lavar	0	2	2	2
Videocassete e/ou DVD	0	2	2	2
Geladeira	0	4	4	4
Freezer (aparelho independente ou parte da geladeira duplex)	0	2	2	2

Fonte: RODRIGUES (2017) com base no Índice ABEP (2014).

Nota: ¹ Assim como definido por Almeida (2014), chefe da família é a pessoa (pai, mãe ou responsável) com maior nível de escolaridade.

Tabela A.2.3 - Distribuição das condições Socioeconômico familiar dos Alunos

Classes	Urbano-Urbano	Rural - Urbano	Rural - Rural
Classe A1	0,86%	0,15%	0,07%
Classe A2	6,13%	1,26%	1,05%
Classe B1	12,5%	4,29%	4,17%
Classe B2	25,95%	15,11%	14,81%
Classe C1	27,35%	26,5%	26,99%
Classe C2	17,4%	22,58%	23,96%
Classe D	9,49%	27,77%	26,5%
Classe E	0,32%	2,34%	2,46%

Fonte: RODRIGUES (2017) com base nos resultados da pesquisa

Tabela A.2.4 - Distribuição do Nível Socioeconômico dos Alunos das Escolas

Estratos	Escolas Urbanas	Escolas Rurais
Muito Baixo	0,49	5,87
Baixo	6,89	24,23
Médio Baixo	13,38	28,72
Médio	20,57	22,42
Médio Alto	30,31	15,01
Alto	16,34	2,57
Muito Alto	12,02	1,18

Fonte: RODRIGUES (2017) com base nos resultados da pesquisa

Tabela A.2.5. Decomposição Oaxaca-Blinder do diferencial de rendimento Grupo 1¹ para o Brasil - Com ponderação

Variáveis	Ciências da Natureza	Ciências Humanas	Linguagem e Códigos	Matemática	Média Geral
Rural_Urbano (1)	464.0005 *	516.7578 *	482.0894 *	443.4353 *	476.5708 *
Rural_Rural (2)	460.9756 *	514.9245 *	477.7441 *	438.7573 *	473.1004 *
<i>Diferença total (1 - 2)</i>	3.0249 *	1.8334 *	4.3453 *	4.6780 *	3.4704 *
<i>Diferença explicada</i>	2.8415 *	4.5578 *	5.4519 *	4.5684 *	4.3549 *
<i>Diferença não explicada</i>	0.1833	-2.7244 *	-1.1066 *	0.1096	-0.8845 **
<i>Diferença Explicada Detalhada</i>					
Características individuais	0.0000	0.0000	0.0000	0.0000	0.0000
Características familiares	0.0016	-0.0008	0.0004	0.0056	0.0019
esc_profiss	-0.1833 *	-0.0801 *	-0.0352	-0.2152 *	-0.1284 *
publico	-0.3409 *	-0.2503 *	-0.1648 *	-0.2675 *	-0.2559 *
inse	1.7719 *	2.7547 *	2.6520 *	3.1362 *	2.5787 *
aluno_turma	0.4715 **	0.8561 *	0.9350 *	0.6384 **	0.7252 *
ird	0.2778 *	0.4045 *	0.6509 *	0.6248 *	0.4895 *
ied	0.3425 *	0.3768 *	0.3332 *	0.6561 *	0.4271 *
iafd	0.4367 *	0.6215 *	0.7919 *	0.2722 **	0.5306 *
iee	0.0629 *	-0.1245	0.2884 **	-0.2822 *	-0.0138
<i>Diferença não Explicada Detalhada</i>					
Características individuais	-5.9076	3.5554	-3.5494	-4.5331	-2.7619
Características familiares	-3.6793	-4.7779	-10.3273 **	-0.5909	-4.9468 **
esc_profiss	-0.5137 *	0.0620	0.2794	-0.4919	-0.1660
publico	-8.9765 *	-3.5106	-5.4856	-19.6466 *	-9.4049 *
inse	6.3317	-4.7310	-7.8066	-4.4231	-2.6573
aluno_turma	-6.4784 *	-11.3256 *	-7.0842 *	-3.4894	-7.0944 *
ird	7.9561 *	6.9161	2.1648	13.1114 *	7.5371 *
ied	-1.1742	1.8289	-2.3368	0.3233	-0.3397
iafd	0.0209	-0.2793	-2.2935	1.7079	-0.2110
iee	-0.4139	0.9598	-0.3583	0.5416	0.1823
Constante	14.8987	8.5779	35.6911 *	17.6005 *	18.9781 *

Fonte: RODRIGUES (2017) com base nos resultados da pesquisa

Nota: ¹Alunos que estudam em escolas rurais vs alunos residentes no meio rural que estudam em escolas urbanas. **Escola:** esc_profiss publico inse aluno_turma ird ied iaafd iie

(*) significante a 1%; (**) significante a 5%.

Tabela A.2.6. Decomposição Oaxaca-Blinder do diferencial de rendimento Grupo 2¹ para o Brasil - Com ponderação

Variáveis	Ciências da Natureza	Ciências Humanas	Linguagem e Códigos	Matemática	Média Geral
Urbano_Urbano (1)	469.394 *	527.934 *	493.704 *	454.009 *	486.260 *
Rural_Urbano (2)	465.046 *	518.027 *	483.030 *	445.630 *	477.933 *
<i>Diferença total (1 - 2)</i>	4.347 *	9.907 *	10.675 *	8.379 *	8.327 *
<i>Diferença explicada</i>	2.190 *	3.642 *	3.308 *	4.933 *	3.518 *
<i>Diferença não explicada</i>	2.158 *	6.264 *	7.366 *	3.446 *	4.809 *
<i>Diferença Explicada detalhado</i>					
idade2	0.2302 *	0.3558 *	0.4048 *	0.3595 *	0.3375 *
idade3	-0.0161 *	-0.0236 *	-0.0365 *	-0.0237 *	-0.0250 *
sexo	0.4173 *	0.1926 *	-0.2374 *	0.9725 *	0.3363 *
branca	0.0566 *	0.0653 *	0.0618 *	0.0931 *	0.0692 *
solteiro	0.0027	0.0007	-0.0053	0.0056	0.0009
trabalha	-0.4587 *	-0.4039 *	-0.3630 *	-0.4301 *	-0.4139 *
conc_ef	0.1135 *	0.1703 *	0.1407 *	0.1754 *	0.1500 *
conc_em	0.0222 *	0.0424 *	0.0351 *	0.0128	0.0281 *
Edu_pai2	-0.0409 *	0.0182	0.0565 **	0.0557	0.0224
Edu_pai3	0.0450 *	0.2057 *	0.3953 *	0.0429	0.1722 *
Edu_pai4	0.1904 *	0.2743 *	0.2790 *	0.3554 *	0.2748 *
Edu_mae2	-0.0673 *	-0.0117	-0.0113	-0.0859 *	-0.0440 *
Edu_mae3	-0.0140	0.8901 *	0.7416 *	0.5794 *	0.5493 *
Edu_mae4	0.2215 *	0.5816 *	0.5194 *	0.3679 *	0.4226 *
Renda01	-0.2299 *	-0.4639 *	-0.4849 *	-0.1171 *	-0.3239 *
Renda02	0.2432 *	0.4417 *	0.4185 *	0.2696	0.3433 *
Renda03	0.7849 *	1.1600 *	1.0674 *	1.0824 *	1.0237 *
Renda04	0.4868 *	0.5361 *	0.4723 *	0.7984 *	0.5734 *
csf	0.1941 *	-0.3987 *	-0.1561	0.4057 *	0.0112 *
Características da Escola	0.0080	0.0095	0.0100	0.0132	0.0102
<i>Diferença não Explicada detalhado</i>					
idade2	0.1451	0.2989 *	0.1137	-0.1023	0.1139 **
idade3	0.0546 **	0.0777 *	0.0594 **	0.0318	0.0559 *
sexo	-0.1491	0.4575 *	0.4328 *	0.2800	0.2553 **
branca	-0.5043 **	-0.7340 *	-0.5196 *	-1.2151 *	-0.7432 *
solteiro	0.3682	-0.6763	1.0615	2.9068	0.9151
trabalha	0.4147 *	0.2151 **	0.0567	0.5541 *	0.3101 *
conc_ef	-0.2268	-0.2176	0.1537	0.6939 **	0.1008
conc_em	0.0719	0.1174	-0.1834	-0.4915 **	-0.1214
Edu_pai2	0.6945 *	0.5902 **	0.5788 *	0.0073	0.4677 *
Edu_pai3	0.2279	0.3044	0.0654	0.4578 **	0.2638 **
Edu_pai4	0.1099	0.0749	0.1040	0.0777	0.0916
Edu_mae2	-0.1267	-0.0045	0.0332	0.4273	0.0823
Edu_mae3	1.4345 *	0.5891 **	0.8882 *	0.2690	0.7952 *
Edu_mae4	0.2571	-0.0585	-0.1464	0.2235	0.0689
Renda01	0.0763	0.7353	0.7368 **	0.9826	0.6328 **
Renda02	0.2204	0.7840 **	0.7985 **	0.7899 **	0.6482 **
Renda03	0.1237	0.6490 *	0.6513 *	0.6453	0.5173 **
Renda04	0.1058	0.3630 *	0.3366 *	0.0615	0.2167 **
csf	-1.9755 **	0.0010	-0.3129	0.4491	-0.4596
Características da Escola	-3.3859	-4.3359	-1.6109	-4.0092	-3.3355
Constante	4.2213	7.0338 **	4.0690	0.4065	3.9327

Fonte: RODRIGUES (2017) com base nos resultados da pesquisa

Nota: ¹Alunos que estudam em escolas rurais vs alunos residentes no meio rural que estudam em escolas urbanas.

Escola: esc_profiss público inse aluno_turma ird ied ia fd iie

(*) significante a 1%; (**) significante a 5%.