

ARTHUR RIBEIRO AZOR IENSEN

**A INFLUÊNCIA DE FATORES SOCIOECONOMICOS SOBRE O DESEMPENHO
INDIVIDUAL NO EXAME NACIONAL DO ENSINO MÉDIO (ENEM)**

**Monografia apresentada ao Curso de Ciências
Econômicas, Universidade de São Paulo, como
requisito parcial para a obtenção do título de
Bacharel em Economia.**

Orientadora: Prof. Dra. Fabiana Fontes Rocha

**SÃO PAULO
2022**

ARTHUR RIBEIRO AZOR IENSEN

**A INFLUÊNCIA DE FATORES SOCIOECONOMICOS SOBRE O DESEMPENHO
INDIVIDUAL NO EXAME NACIONAL DO ENSINO MÉDIO (ENEM)**

**Monografia apresentada ao Curso de Ciências
Econômicas, Universidade de São Paulo, como
requisito parcial para a obtenção do título de
Bacharel em Economia.**

Orientador: Prof. Dra. Fabiana Fontes Rocha

SÃO PAULO

2022

FICHA CATALOGRÁFICA

Iensen, Arthur

A influência de fatores socioeconômicos sobre o desempenho individual no Exame Nacional do Ensino Médio (Enem) – São Paulo, 2022.

58 páginas

Área de concentração: Economia da educação.

Orientador: Prof. Dra. Fabiana Fontes Rocha.

Trabalho de Conclusão de Curso – Faculdade de Economia, Administração, Ciências Contábeis e Atuariais da Universidade de São Paulo

1. Desempenho escolar; 2. Características socioeconômicas; 3. Efeito quantitativo

Aos meus pais, Geraldo e Eliane, por terem me proporcionado a base sobre a qual
tudo foi possível

Dedico

AGRADECIMENTOS

A minha mãe, Eliane, por ter me incentivado sempre à busca da felicidade e pelo suporte emocional sempre que necessário

Ao meu pai, Geraldo, pela grande inspiração à busca do conhecimento e fascínio pelas artes

A minha irmã, Joyce, pela companhia diária e exemplo de que é sempre possível se reinventar

A Prof.^a Dra. Fabiana Fontes Rocha pela orientação e apresentação do mundo fascinante da econometria

Aos meus amigos de infância por me mostrarem que existem afetos que nunca se esvaem

Aos meus amigos de São Paulo por me mostrarem que é possível encontrar um lar em novos lugares

Aos meus colegas de graduação por tornarem mais leve a jornada e sempre estarem dispostos à discussão

A todos os professores que cruzaram meu caminho e me mostraram o poder da educação

“Se a educação sozinha não transforma a sociedade, sem ela tampouco a sociedade muda”

Paulo Freire.

SUMÁRIO

LISTA DE TABELAS	VII
LISTA DE GRÁFICOS	VIII
RESUMO.....	IX
ABSTRACT.....	X
1 INTRODUÇÃO	11
2 REVISÃO DE LITERATURA	14
2.1 ECONOMIA E EDUCAÇÃO	14
2.2 DETERMINANTES DO DESEMPENHO ESCOLAR NO BRASIL	16
3 DADOS E METODOLOGIA	22
3.1 DADOS.....	22
3.2 METODOLOGIA MODELO <i>DUMMIES</i> MQO	27
3.3 METODOLOGIA MODELO KLEIN & VELLA	28
4 RESULTADOS.....	32
4.1 RESULTADOS MODELO <i>DUMMIES</i> MQO	32
4.2 RESULTADOS MODELO KLEIN & VELLA.....	38
5 DISCUSSÃO	43
5.1 COMPARAÇÃO DE RESULTADOS	43
5.2 POLÍTICAS VOLTADAS À EDUCAÇÃO	47
6 CONCLUSÕES	49
REFERÊNCIAS.....	51
APÊNDICE 1 - RESULTADO REGRESSÃO COMPLETA MODELO <i>DUMMIES</i> MQO	53

LISTA DE TABELAS

TABELA 1- STATUS DE CONCLUSÃO DO ENSINO MÉDIO	23
TABELA 2 - DISTRIBUIÇÃO PERCENTUAL DAS REGIÕES	24
TABELA 3 - ESTATÍSTICAS DESCRITIVAS DAS VARIÁVEIS	26
TABELA 4 - DESCRIÇÃO DE VARIÁVEIS MODELO KV	31
TABELA 5 – RESULTADOS CARACTERISTICAS PESSOAIS MODELO MQO DUMMIES	32
TABELA 6 - RESULTADOS CARACTERISTICAS ESCOLARES MODELO MQO DUMMIES	34
TABELA 7 – RESULTADOS CARACTERISTICAS FAMILIARES MODELO MQO DUMMIES	36
TABELA 8 - RESULTADOS MODELO KLEIN & VELLA I	39
TABELA 9 - RESULTADOS MODELO KLEIN & VELLA II	40

LISTA DE GRÁFICOS

GRÁFICO 1 - INFLUÊNCIA DA ESCOLARIDADE DO PAI - ENEM 2015	18
GRÁFICO 2 - INFLUÊNCIA DA ESCOLARIDADE DA MÃE - ENEM 2015.....	19
GRÁFICO 3 – NOTAS MÉDIAS	24
GRÁFICO 4 – NOTAS MÉDIAS POR REGIÃO	25
GRÁFICO 5 - INFLUÊNCIA DA ESCOLARIDADE DO PAI - ENEM 2020	45
GRÁFICO 6 - INFLUÊNCIA DA ESCOLARIDADE DA MÃE - ENEM 2020.....	45

RESUMO

A INFLUÊNCIA DE FATORES SOCIOECONOMICOS NO DESEMPENHO INDIVIDUAL NO EXAME NACIONAL DO ENSINO MÉDIO (ENEM)

Objetivo: Estimar quantitativamente a influência dos fatores socioeconômicos dos indivíduos, identificados por características pessoais, escolares e familiares, no desempenho dos mesmos na prova do Enem. **Dados e Método:** A partir dos Microdados do Enem 2020 foram estimados dois modelos econométricos tendo como variável dependente as notas obtidas em cada prova do exame, bem como uma média geral calculada a partir de uma média aritmética. Como variáveis explicativas foram utilizados dados sobre a idade, a raça e o sexo do aluno, o tipo de administração e localização da escola, a renda familiar, a escolaridade da mãe e do pai, a posse de computadores e acesso à internet. **Resultados:** De maneira geral observou-se um melhor desempenho de alunos homens brancos, provenientes de escolas privadas ou federais urbanas, com famílias de alta renda e alta escolaridade dos pais. Os resultados apresentaram variações entre as provas e os modelos. **Conclusão:** O ambiente socioeconômico exerce grande influência no desempenho escolar dos alunos a partir do nível de investimento em capital humano na família, das características da escola e das oportunidades apresentadas.

Palavras-chave: educação; desempenho escolar; características socioeconômicas; efeitos quantitativos; Enem 2020

Códigos JEL: I24, C31, J32

ABSTRACT

THE INFLUENCE OF SOCIOECONOMIC FACTORS ON INDIVIDUAL PERFORMANCE ON THE NATIONAL HIGH SCHOOL EXAM (ENEM)

Purpose: To quantitatively estimate the influence of socioeconomic factors of individuals, identified by personal, school and family characteristics, on their performance in the Enem exam. **Data and Methods:** Based on the Enem 2020 microdata, two econometric models were estimated, having as dependent variable the grades obtained in each test of the exam, as well as a general average calculated from an arithmetic mean. As explanatory variables were used data on age, race and gender of the student, the type of administration and location of the school, family income, mother's and father's education, computer ownership and internet access. **Results:** In general, we observed better performance by white male students, coming from urban private or federal schools, with high-income families and high parental education. Results varied across tests and models **Conclusion:** The socioeconomic environment has a major influence on the school performance of students based on the level of investment in human capital in the family, the characteristics of the school, and the opportunities presented.

Key words: education; school performance; socioeconomic characteristics; quantitative effects; Enem 2020

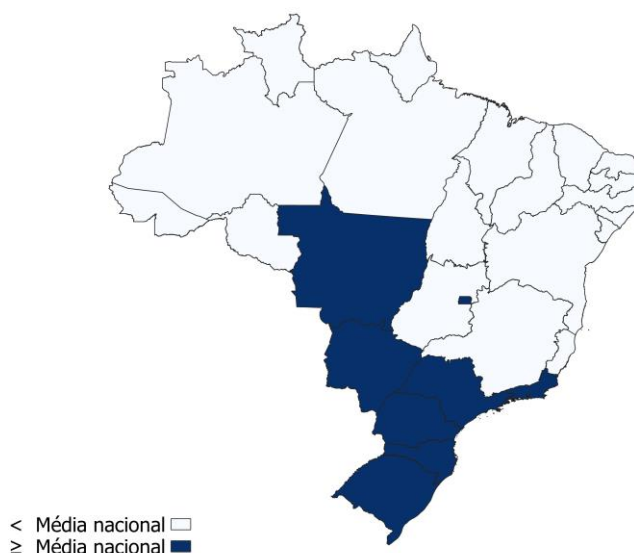
Jel Codes: I24, C31, J32

1 INTRODUÇÃO

A educação é tema de grande relevância na questão do desenvolvimento socioeconômico de um país. No que diz respeito à discussão do capital humano, o nível de instrução da força de trabalho é uma questão central à medida que maiores níveis de produtividade e de redução de custo podem ser atingidos com o avanço da educação da população. Ao se ter um nível de escolaridade elevada é possível acelerar o desenvolvimento e diminuir os gaps de desigualdade (Viana e Lima, 2010).

O Brasil é um país desigual. A desigualdade de renda pode ser observada em diferentes níveis no território, seja entre macrorregiões, unidades da federação (UF) e até entre as famílias. A partir de dados do IBGE observa-se que, no ano de 2019 a região mais rica, o Sudeste, concentrou 53% do PIB, enquanto a segunda mais rica, o Sul, teve participação de 17,2% no indicador. Em termos de PIB per capita, nota-se que os estados mais ao Sul possuem melhores indicadores comparados com estados mais ao Norte. A Figura 1 mostra a comparação entre o PIB per capita dos estados com o indicador nacional.

FIGURA 1 – COMPARAÇÃO PIB PER CAPITA BRASIL E UFS



Fonte: Elaboração própria a partir de dados do IBGE

Dentro de cada UF a desigualdade pode ser observada a partir do índice de Gini para renda, calculado pelo IBGE. Para 2019 os valores variaram entre 0,87 no

estado de São Paulo e 0,68 no estado do Mato Grosso. Enquanto São Paulo é o estado com maior PIB é também o mais desigual nesse quesito.

Em 2019, segundo o IBGE, o Brasil teve uma média de 9,8 anos de estudo, ou seja, o brasileiro médio não completa a educação básica proposta de 12 anos ao fim do Ensino Médio. Existe também uma diferença entre tal indicador para homens e mulheres, tendo as últimas em média 10 anos de estudo, 0,4 ano a mais que os homens. A diferença é ainda maior entre brancos e pretos ou pardos, chegando a 1,6 ano. Enquanto os brancos estudam em média 10,7 anos, os pretos ou pardos estudam 9,1. Diante deste contexto de desigualdade na renda e na escolaridade, uma educação de qualidade se faz essencial para um desenvolvimento equilibrado.

O Exame Nacional do Ensino Médio (Enem) é uma avaliação aplicada anualmente capaz de prover informações para se quantificar o desempenho dos estudantes brasileiros ao fim da educação básica. O exame cobre todo o conteúdo programático do ensino médio e avalia seus participantes em todas as áreas do conhecimento propostas pela estrutura curricular da educação básica brasileira. A avaliação é dividida em cinco provas, Ciências da Natureza, Ciências Humanas, Matemática, Linguagens e Códigos e Redação, obtendo-se uma nota para cada prova. Além das provas de conhecimentos, é respondido também um questionário socioeconômico, trazendo informações pessoais, escolares e familiares sobre os inscritos.

Em sua maioria, os estudos realizados com base no Enem tratam da qualidade da educação e desempenho dos alunos (Lima et al., 2019). A partir das respostas do questionário socioeconômico é possível diferenciar os alunos com base em características como idade, raça, sexo, renda familiar, escolaridade dos pais e características da escola, como localização e administração. Visto que o Brasil é um país heterogêneo e as desigualdades estão atreladas ao *background* socioeconômico dos indivíduos o presente trabalho se dedica a aferir como e quanto o ambiente socioeconômico do estudante influencia o seu desempenho acadêmico.

A monografia tem como objetivo investigar a influência dos fatores socioeconômicos individuais e familiares, além das características da escola, nas notas atingidas no ENEM, no intuito de mensurar quais características interferem no desempenho escolar dos alunos. Objetivos específicos são:

1. Avaliar a influência das características individuais, fatores socioeconômicos e características escolares no desempenho individual no Enem
2. Estimar modelos econométricos para avaliação quantitativa da influência das variáveis selecionadas
3. Identificar os fatores que mais influenciam no desempenho individual no ENEM
4. Discutir possíveis focos para políticas públicas na área de educação.

Para isso foram elaborados dois modelos econométricos com base nos Microdados do Enem 2020, disponibilizados pelo Instituto Nacional de estudos e Pesquisas Educacionais Anísio Teixeira (Inep), a instituição realizadora do exame. Foi avaliada quantitativamente a influência das características socioeconômicas dos indivíduos, divididas em pessoais, escolares e familiares, em cada nota no Enem, bem como em uma média geral.

O trabalho é dividido em, além deste, mais 5 capítulos. No segundo capítulo é apresentada uma revisão bibliográfica acerca da relação entre a economia e a educação bem como resultados para estudos sobre a correlação do *background* socioeconomico e o desempenho escolar no Brasil. No terceiro capítulo se apresentam com detalhes os dados utilizados e as metodologias para estimação de cada modelo. No quarto capítulo são descritos os resultados das estimativas com enfoque nas diferentes estimativas obtidas para cada prova. No quinto é feita uma discussão entre os resultados obtidos no presente trabalho e os expostos na revisão bibliográfica, bem como uma discussão de políticas voltadas à educação. Por último conclui-se.

2 REVISÃO DE LITERATURA

2.1 ECONOMIA E EDUCAÇÃO

Nos desenvolvimentos clássicos da teoria econômica inicialmente se atrelava a produção econômica a três fatores: o trabalho, o capital e a terra. Até a década de 1950 o desenvolvimento era atrelado à dotação de cada país destes fatores (Solow, 1957), porém com o desenvolvimento de estudos empíricos notou-se uma lacuna no poder explicativo ao se considerar somente tais fatores, assim emerge a discussão de um novo tipo de capital: o capital humano. Os estudos de Mincer (1958), Schultz (1964) e Becker (1964) iniciam a nova fase de entendimento acerca desse novo fator e o atrelam majoritariamente à educação (Viana e Lima, 2010). O investimento em capital humano ocorre quando o indivíduo escolhe dedicar-se ao acúmulo de conhecimento e expansão da formação para posterior aplicação destes no trabalho em detrimento à dedicação instantânea para o trabalho. Isso leva ao desenvolvimento à medida que trabalhadores melhores qualificados apresentam maior produtividade acarretando uma diminuição dos custos e minimizando os rendimentos decrescentes do capital (Viana e Lima, 2010). Daí então surge uma nova questão relacionada a trajetória do desenvolvimento e como a distribuição da educação afeta os indivíduos diferentemente na sociedade.

Segundo Barros, Henriques e Mendonça (2002), características relacionadas à educação contribuem para uma desigualdade salarial a partir de dois fatores: o nível de desigualdade educacional e a valoração monetária do mercado para cada ano a mais de escolaridade. O primeiro diz respeito à diferença de remuneração entre trabalhos qualificados e não qualificados, assim uma força de trabalho heterogênea em termos educacionais reflete uma desigualdade salarial. E o segundo relaciona-se com os valores absolutos da diferença salarial, pois uma maior valoração monetária dos anos extra de estudo se traduzem em maiores diferenças nas remunerações dos trabalhos de diferentes qualificações. Barros, Henriques e Mendonça (2002, p. 6) ilustram essas relações:

Uma simples analogia pode ilustrar esse argumento causal. Consideremos que a desigualdade salarial representa a imagem da desigualdade educacional projetada através de um espelho curvo. Nesse caso, a imagem projetada (desigualdade salarial) seria tão maior quanto maior fosse o objeto original (heterogeneidade educacional da força de trabalho) e quanto maior

fosse a curvatura do espelho (o valor que o mercado de trabalho atribui a cada ano a mais de escolaridade)

Para o caso do Brasil, Lorel (2008) apresenta resultados esclarecedores sobre a desigualdade educacional. A partir do cálculo de um índice de Gini para a educação o autor mostra uma redução da desigualdade educacional durante a segunda metade do século XX atrelado a um fortalecimento da educação básica no país. Tal tendência pôde ser observada para todas as regiões do país, sendo o Nordeste a região com os maiores índices de desigualdade em todo o período e as regiões Sul e Sudeste as menores. Ao se excluirmos os indivíduos sem nenhuma educação observou-se uma diferença expressiva na tendência, não houve grandes mudanças nos índices de concentração, logo conclui-se que a melhoria captada pelo autor ocorreu devido a um esforço de alfabetização da população. Os resultados indicaram também uma correlação negativa entre a escolaridade média da população e o índice de Gini educacional além de um indício, mesmo que fraco, de uma curva de Kuznets relacionando anos de estudo e desigualdade de renda, ou seja, com o aumento da escolaridade média a desigualdade de renda tende a subir em um primeiro estágio até um limite de 6,59 anos de estudo, ponto em que ocorre uma inversão. Tais resultados sugerem a necessidade de uma mobilidade intergeracional da escolaridade, fazendo com que uma melhoria na educação se propague.

Ferreira e Veloso (2003), baseados em dados da PNAD de 1996, caracterizam padrões não-lineares da mobilidade intergeracional da educação no Brasil. Os autores apresentam resultados relacionados à probabilidade de os filhos manterem a escolaridade dos pais. O padrão de mobilidade difere entre grupos raciais e regiões. Para negros a probabilidade de um indivíduo com pai de baixa escolaridade permanecer com baixa escolaridade é superior que esta mesma para os brancos, enquanto a situação se inverte ao se considerar pais com alta escolaridade, ou seja os indivíduos brancos tem maior probabilidade de manter escolaridade alta. Em se tratando de regiões, a diferença observada pelos autores se deu entre a região Nordeste e Sudeste, tendo a primeira maior probabilidade de se manter uma baixa escolaridade. De maneira geral os autores observaram também uma menor mobilidade no Brasil em comparação com países desenvolvidos e em desenvolvimento. Mahlmeister et al. (2019) replicaram o estudo para dados da

PNAD de 2014 obtendo resultados similares, porém atenuados, como exposto pelos autores:

Especificamente, a menor mobilidade entre os negros continua fortemente associada à maior probabilidade, nesse grupo, de o filho de um pai sem escolaridade permanecer sem escolaridade. Enquanto a probabilidade de um filho negro de um pai sem escolaridade permanecer na mesma categoria de educação do pai é cerca de 23%, a probabilidade análoga para brancos é pouco abaixo de 11%. Já a persistência de alta escolaridade é significativamente mais elevada entre indivíduos de cor branca, ainda que a diferença em relação a essa persistência entre os negros seja menos pronunciada hoje do que outrora. Comparando os dois anos estudados, a probabilidade de o filho de um pai com ensino superior completo também completar seus estudos universitários passou de 40% para cerca de 62%, se o indivíduo reporta ser negro, e de aproximadamente 62% para 74% para brancos. (Mahlmeister et al., 2019, p. 160)

Para que se atinja então níveis superiores de desenvolvimento econômico e se diminuam os níveis de desigualdade no país, uma educação de qualidade é essencial. Albernaz, Ferreira e Franco (2002) se dedicaram a estudar o tema a partir de dados do Sistema Nacional de Avaliação da Educação Básica (Saeb) de 1999 em um esforço para identificar quais fatores estão associados à qualidade do ensino brasileiro. Os autores concluíram que cerca de 80% da variância do desempenho de alunos da 8ª série na avaliação do Saeb se deu por conta de diferentes níveis socioeconômicos, além de que alunos negros obtiveram piores desempenhos mesmo ao se controlar pelas demais características individuais. Os resultados apontaram também um pior desempenho de escolas públicas e atrelaram a variação dos rendimentos escolares a variáveis como escolaridade dos professores, qualidade física e financeira da escola.

Ainda sobre a qualidade da educação, a seguir apresenta-se trabalhos dedicados a medir seus determinantes utilizando notas nas provas do Enem e do Exame Nacional de Desempenho dos Estudantes (Enade) como indicadores do desempenho escolar dos indivíduos.

2.2 DETERMINANTES DO DESEMPENHO ESCOLAR NO BRASIL

O desempenho no Enem não é homogêneo entre as regiões do país. Viggiano e Mattos (2013), ao realizar uma análise exploratória dos dados do Enem 2010 evidenciaram uma diferença entre as cinco macroregiões. Os autores foram capazes de classificar 3 grupos a partir das notas médias por região obtidas nas

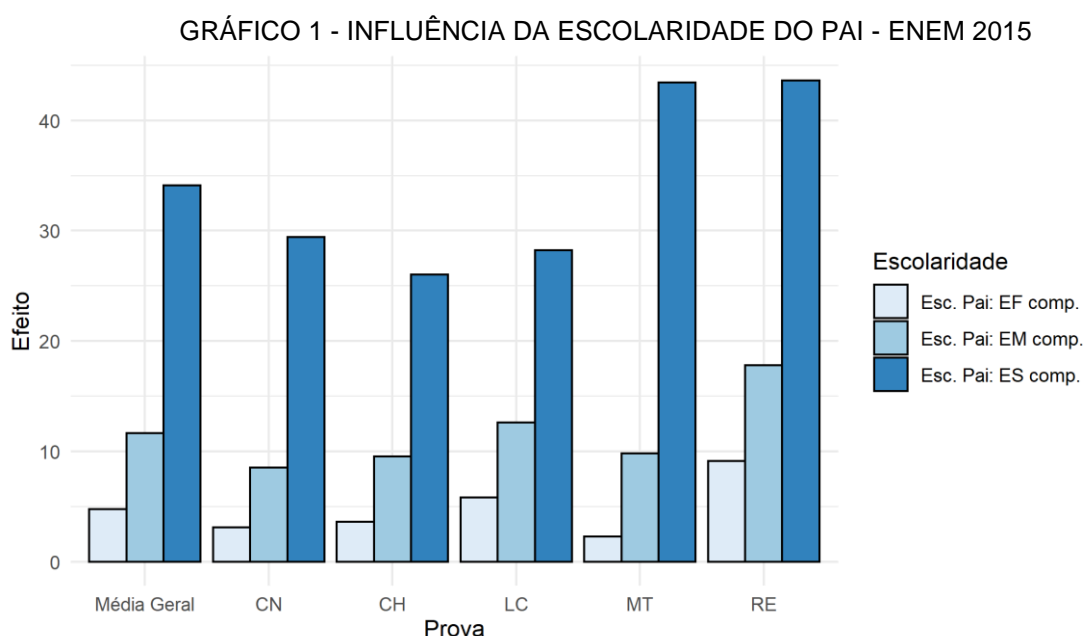
cinco provas e de uma nota final calculada como uma média aritmética. As regiões Sul e Sudeste compõe o grupo de desempenho superior, o Centro-Oeste de desempenho médio e Nordeste e Norte de desempenho inferior. Tais resultados foram observados na nota final e nas provas de Ciências da Natureza, Ciências Humanas, Linguagens e Código e Matemática em que a ordem decrescente de desempenho foi Sudeste, Sul, Centro-oeste, Nordeste e Norte, enquanto na prova de Redação as notas decresceram na ordem Sudeste, Nordeste, Norte, Sul, Centro-Oeste. De maneira geral os autores apontam a necessidade de uma melhora na educação brasileira para que se atinja um desempenho minimamente aceitável, visto que em nenhuma das provas se obteve rendimento superior a 59%.

Curi e Menezes Filho (2013) se dedicaram a analisar as influências das características escolares e familiares dos alunos no seu desempenho na parte objetiva do Enem, ou seja, excluindo-se a Redação. Os autores utilizaram uma amostra do Enem de 2006 composta por alunos de escolas paulistas que concluíram o ensino médio neste ano e criaram um *ranking* de escolas a partir dos efeitos escola – a influência das características escolares na nota - estimados. A educação da mãe, bem como a renda familiar se mostrou positivamente correlacionadas com a posição no *ranking*, ou seja, escolas com alunos de mãe com alta escolaridade e família de alta renda mostraram desempenhos melhores na prova. As estimativas obtidas pelos autores para os efeitos familiares indicaram que homens não negros com mãe de alta escolaridade e alta renda familiar obtém notas maiores no exame. Os resultados indicaram também que “a idade é negativamente correlacionada com o desempenho escolar, isto é, alunos mais velhos têm, em média, notas inferiores” (Curi e Menezes Filho, 2013, p. 251).

Ao analisar os dados do Enem 2015, Henares, Menezes Filho e Komatsu (2021) trazem resultados sobre como as características socioeconômicas se relacionam com o desempenho dos alunos. Os autores utilizaram como amostra os participantes que concluíram o ensino médio no ano de 2015 e como variáveis explicadas as notas nas cinco provas do exame e uma média geral calculada a partir da média aritmética das provas. As características analisadas foram: escolaridade do pai, escolaridade da mãe, idade, raça, sexo, turno de estudo e características escolares como a localização e esfera administrativa. A escolaridade dos pais e das mães dos alunos se mostrou positivamente correlacionada com o desempenho em todas as provas, ou seja, uma maior escolaridade dos pais refletiu notas maiores e

ainda, a educação dos pais teve efeito ligeiramente maior que a das mães Henares, Menezes Filho e Komatsu, 2021) como pode ser visto no Gráfico 1 e no Gráfico 2.

Em relação à idade e à raça, Henares, Menezes Filho e Komatsu (2021) observaram que alunos mais jovens, brancos ou amarelos obtiveram um resultado superior. Sobre o sexo, na média geral, Ciências da Natureza, Ciências Humanas e Matemática os homens se saíram melhor e em Linguagens e Códigos e Redação as mulheres foram superiores. Alunos que estudaram apenas no turno diurno, em escolas urbanas e que possuíam Ensino Médio profissionalizante tiveram notas maiores que as suas contrapartes. Ainda, os alunos de escolas de administração federal apresentaram melhores rendimentos, seguidos pelos de escolas privadas.

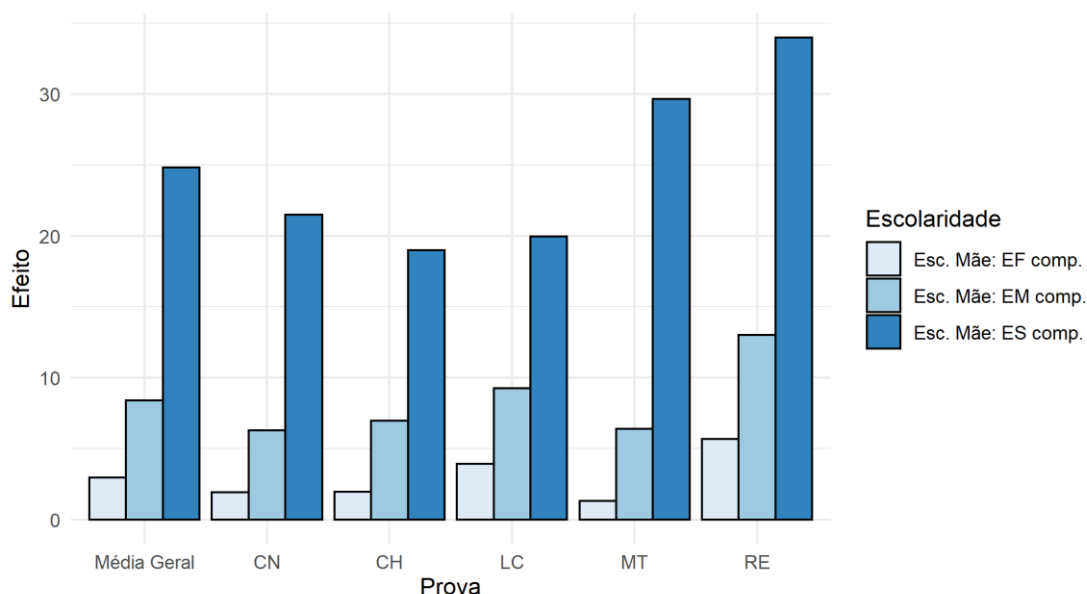


Fonte: Henares, Menezes Filho e Komatsu (2021)

Nota: Em comparação ao Ensino Fundamental incompleto

Por fim, Henares, Menezes Filho e Komatsu (2021) observaram que as influências dos fatores estudados não foram iguais para todas as provas, sendo a prova de Redação aquela que apresentou as maiores diferenças entre os grupos.

GRÁFICO 2 - INFLUÊNCIA DA ESCOLARIDADE DA MÃE - ENEM 2015



Fonte: Henares, Menezes Filho e Komatsu (2021)

Nota: Em comparação ao Ensino Fundamental incompleto

Com base nos dados do Enem 2018, Jaloto e Primi (2021) investigaram o efeito de características pessoais, socioeconômicas e escolares sobre o desempenho dos alunos nas quatro provas objetivas do exame, Ciências Humanas, Ciências da Natureza, Matemática e Linguagens e Códigos. A amostra utilizada pelos autores foi composta por participantes com pelo menos 16 anos, que concluiriam o Ensino Médio em 2018. Foram excluídos das análises de cada prova os que não obtiveram notas nas respectivas provas, bem como alunos que deixaram a prova em branco. Para o nível socioeconômico (NSE) dos indivíduos foi utilizado um índice composto por renda familiar, ocupação e escolaridade dos pais, assim, “as variáveis preditoras de primeiro nível (estudante) foram: atraso escolar (considerado neste trabalho como equivalente a possuir mais de 18 anos no fim de 2018), cor/raça, sexo e NSE” (Jaloto e Primi, 2021, p. 130), e para o nível da escola foi utilizado a esfera administrativa como variável explicadora. A partir de regressões lineares multiníveis, os autores foram capazes de concluir que todas as características estudadas guardaram relações com o desempenho do estudante. O atraso escolar apresentou um efeito negativo na nota do aluno enquanto o NSE e ser da raça branca apresentaram um efeito positivo. Os homens tiveram um desempenho superior nas quatro provas analisadas, com uma diferença expressiva apenas na prova de matemática (Jaloto e Primi, 2021). Os autores observaram

também um melhor desempenho de escolas privadas e federais em relação às escolas municipais e estaduais, sendo estas últimas aquelas com piores rendimentos dentre as 4 quatro provas. Por fim, Jaloto e Primi (2021, p. 138) destacam que “a prova de MT foi a que apresentou o maior efeito de escola privada, de sexo e a maior variância do efeito do NSE. Isso pode indicar que essa área possui algumas especificidades associadas a essas diferenças.”

Figuerêdo, Nogueiray e Santanaz (2014) se dedicaram a estudar o papel das oportunidades sobre o desempenho dos alunos no Enem 2010. Segundo os autores, tal desempenho depende de dois fatores: o esforço do indivíduo, controlado pelo mesmo e o conjunto de circunstâncias enfrentando, definido “por variáveis como raça e região de nascimento, além do background familiar do estudante” (Figuerêdo, Nogueiray e Santanaz, 2014, p. 374), que não está sob controle dele. Para isso, os autores estimaram dois modelos, o primeiro em que consideraram o esforço independente das circunstâncias e o segundo, baseado na metodologia de Klein e Vella (2010) – adotada também neste trabalho e explicada na seção de metodologia -, para aferir a magnitude da influência das variáveis que compõe as circunstâncias. Os resultados do primeiro modelo apontaram para uma enorme desigualdade social expressa pela desigualdade educacional visto que um indivíduo de baixo background social – definido por renda familiar, escolaridade dos pais e tipo de escola, pública ou privada – tem probabilidade quatro vezes menor de atingir a nota média nacional em comparação com indivíduos de alto background social (Figuerêdo, Nogueiray e Santanaz, 2014). Já o segundo modelo de Figuerêdo, Nogueiray e Santanaz (2014), trouxe como as variáveis que definem as circunstâncias a escolaridade dos pais, renda familiar, tipo de escola, sexo, raça, localização da moradia e qualidade da escola, esta última advinda do Sistema Nacional de Avaliação da Educação Básica (SAEB). A metodologia adotada pelos autores foi capaz de aferir um efeito indireto da educação dos pais sobre o desempenho dos alunos, advindos da premissa de uma transferência intergeracional de habilidade não observada, porém tal medida pode ser afetada por outros fatores não observados como motivação e sorte. Os resultados obtidos por Figuerêdo, Nogueiray e Santanaz (2014) apontaram um efeito indireto da educação da mãe 2,36 vezes maior que o efeito direto, porém para a educação do pai o efeito indireto não foi observado. De maneira geral, os autores encontraram um melhor desempenho associados a mulheres não negras, residentes em áreas urbanas,

advindas de escolas particulares com boa pontuação no SAEB, de alta renda familiar e com pais de alto nível educacional.

Saindo da esfera do Enem, Gutierrez, Moreira e Marquetti (2021) analisaram o efeito do *background* familiar sobre o Ensino Superior a partir de dados do Enade de 2017. Para a estimação dos efeitos, os autores se utilizaram da metodologia de Klein e Vella (2010) com variável explicada sendo o logaritmo da nota e variáveis explicativas o tipo de escola frequentada no Ensino Médio (pública ou privada), a raça (branca ou não branca), o sexo, se o indivíduo trabalha ou não, número de pessoas que moram na residência, renda familiar e anos de estudos dos pais. De maneira geral os resultados apontaram um efeito negativo sobre o desempenho para escolas públicas, sexo masculino, estar trabalhando e mais moradores na residência, enquanto ser branco, ter alta renda familiar e pai com alta escolaridade apresentaram efeitos positivos. No que diz respeito aos efeitos indiretos da educação dos pais e ao efeito direto da educação da mãe estimados pela metodologia de Klein e Vella (2010), Gutierrez, Moreira e Marquetti (2021) não obtiveram resultados significativos.

Como pode ser visto, trabalhos que se dedicam a estimar influências de características socioeconômicas sobre o desempenho escolar no Brasil apontam para a existência de efeitos significativos de variáveis como sexo, raça, renda familiar, escolaridade dos pais e tipo de escola. De maneira geral, observa-se um desempenho superior associado a alta renda e escolaridade dos pais, bem como à raça branca.

Na próxima seção são apresentados os dados e metodologias utilizadas no presente trabalho.

3 DADOS E METODOLOGIA

3.1 DADOS

O banco de dados utilizado foi composto pelos Microdados do Enem 2020 adquirido no site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), a entidade responsável pela aplicação da prova.

O Enem tem como objetivo principal aferir o nível de conhecimento de seus participantes adquirido ao fim do ensino médio. Com os resultados do exame deve ser possível uma autoavaliação do participante, uma avaliação do currículo nacional do ensino médio, a seleção de ingressantes em diversas faculdades públicas e em programas de apoio financeiro aos estudos e o desenvolvimento de estudos acerca da educação brasileira (INEP, 2020).

No ano de 2009 o exame passou por uma reformulação, passando a avaliar os participantes em quatro áreas do conhecimento: Linguagens, Códigos e suas tecnologias, Matemática e suas tecnologias, Ciências da Natureza e suas tecnologias e Ciências Humanas e suas tecnologias. A prova passou a ser composta por 180 questões objetivas - sendo 45 de cada área citada acima - e uma redação, aplicada em dois dias, a partir de 2017, dois domingos seguidos. Com esta reformulação o Enem passou a ser a principal porta de entrada para o ensino superior no Brasil.

No ano de 2020 devido à pandemia do coronavírus o exame foi oferecido também em formato digital, pela primeira vez. Foram oferecidas 100 mil vagas para a realização do Enem Digital, distribuídas em municípios específicos. O intuito é que a modalidade seja implementada de forma progressiva.

À cada realização do exame o Inep disponibiliza os Microdados do Enem. O banco de dados é composto por informações divididas em seis seções, são elas: dados do participante, dados da escola, dados do local da aplicação da prova, dados da prova objetiva, dados da redação, dados do questionário socioeconômico. Assim, informações pessoais, da escola em que estudou, das condições socioeconômicas familiares e dos resultados no exame estão disponíveis sobre cada um dos inscritos.

Em 2020, ao todo foram disponibilizados os dados de 5.783.109 indivíduos. Este número não foi igual ao total de inscritos pois, seguindo a LGPD, foram excluídos do banco alunos que poderiam ser identificados indevidamente a partir de

tipos de provas realizadas que contaram com um número de participantes muito pequeno (INEP, 2020). Destes, 2.588.681 participantes estiveram presentes em todas as provas aplicadas, resultando em um índice de abstenção de 55,24% dos inscritos (sem contabilizar os excluídos do banco de dados).

De acordo com o status de conclusão do ensino médio os alunos são divididos em quatro grupos, como mostrado na Tabela 1. Dos presentes, 55% já haviam concluído o ensino médio no momento da inscrição e 31% concluiriam em 2020.

TABELA 1- STATUS DE CONCLUSÃO DO ENSINO MÉDIO

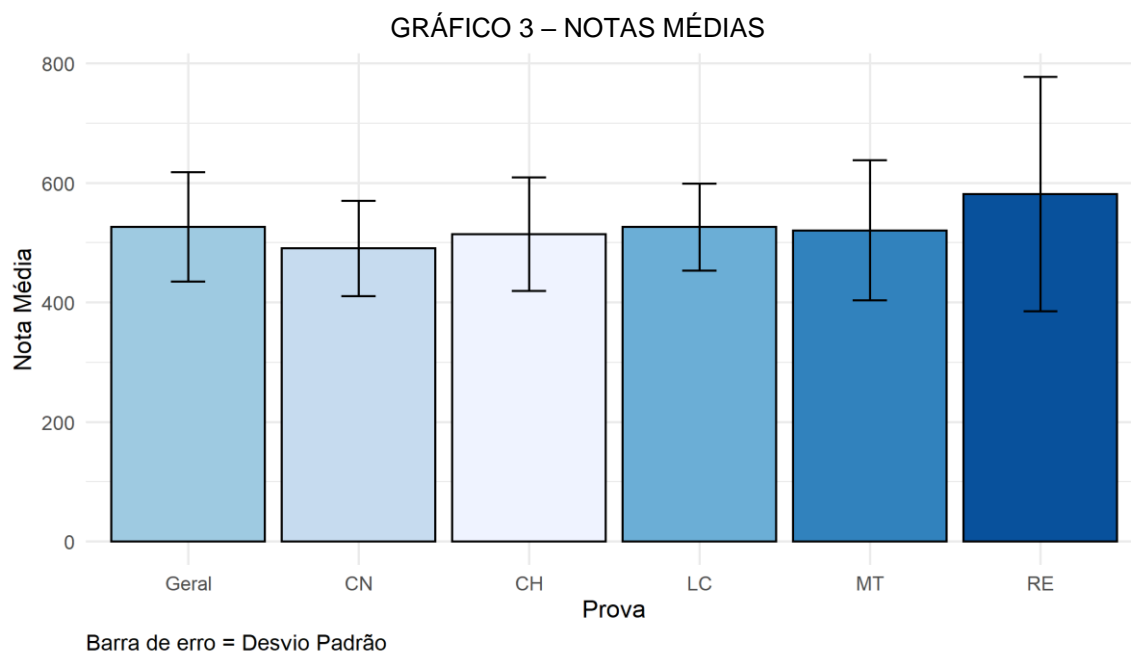
Status de conclusão EM	Média
Já concluí o Ensino Médio	0,55
Estou cursando e concluirei o Ensino Médio em 2020	0,31
Estou cursando e concluirei o Ensino Médio após 2020	0,13
Não concluí e não estou cursando o Ensino Médio	0,003

Fonte: Elaboração própria a partir dos Microdados do Enem 2020

Para avaliar o desempenho dos participantes foram analisadas as notas em cada uma das cinco provas (Ciências Humanas e suas tecnologias – CH, Ciências Naturais e suas tecnologias – CN, Matemática e suas tecnologias – MT, Linguagens, Códigos e suas tecnologias – LC e Redação – RE) e uma média geral calculada como a média aritmética das cinco notas.

$$Média\ Geral = \frac{1}{5} \times (CH + CN + MT + LC + RE)$$

Como ilustrado no Gráfico 3 a prova com maior nota média foi a Redação, atingindo 581,34 de um total de 1.000 pontos, e aquela com a menor média foi Ciências da Natureza, com 490,53 pontos. Em termos de desvios padrão, representados pela barra de erro no gráfico, as provas com maior e menor desvio foram, respectivamente, Redação e Linguagem e Códigos. A média geral entre todos os presentes no exame foi de 526,58 pontos.



Fonte: Elaboração própria a partir dos Microdados do Enem 2020

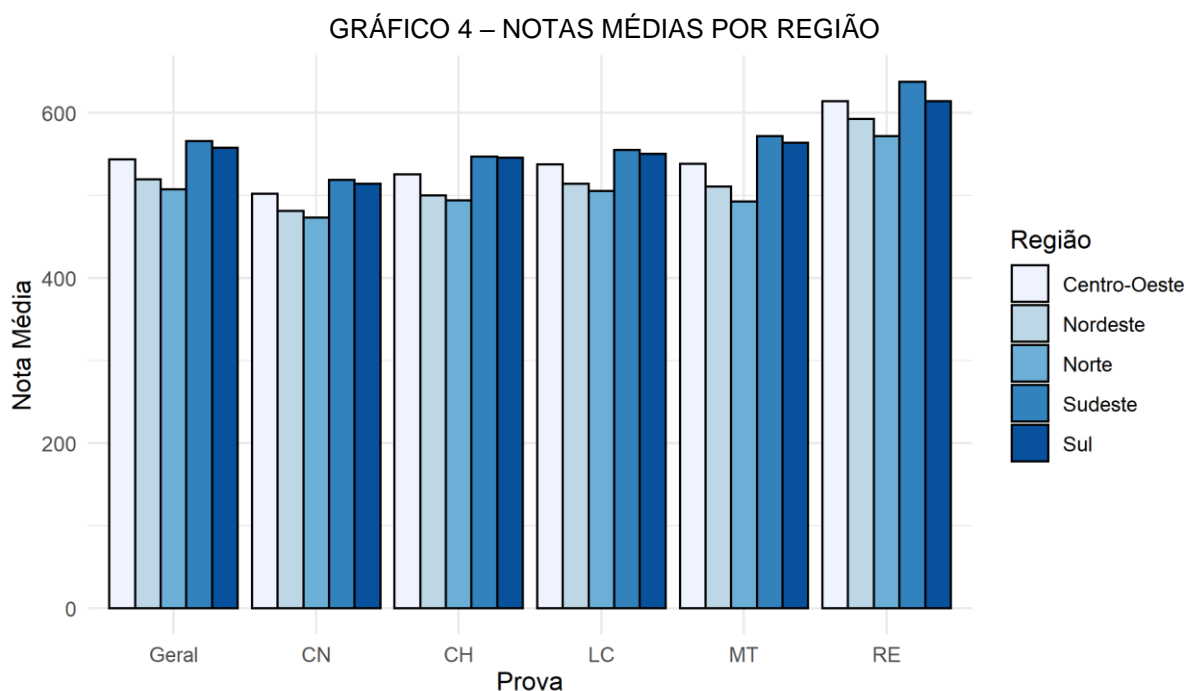
Em se tratando da localização, pode-se extrair do banco de dados a região em que cada participante estudou. Para isso foi utilizado a variável de Unidade da Federação da escola para cada participante. Ao se excluir as linhas que não contém a informação necessária restaram 537.909 indivíduos, distribuídos nas regiões conforme exposto na Tabela 2. As regiões Sudeste e Nordeste foram aquelas com mais alunos no exame, representando respectivamente 37% e 33% dos presentes, enquanto as regiões Norte e Centro-Oeste foram aquelas com menos alunos, ambas com a parcela de 9%. Tal distribuição está dentro do esperado conforme a distribuição populacional do país.

TABELA 2 - DISTRIBUIÇÃO PERCENTUAL DAS REGIÕES

Região	Média
Norte	0,09
Nordeste	0,33
Centro-Oeste	0,09
Sudeste	0,37
Sul	0,12

Fonte: Elaboração própria a partir dos
Microdados do Enem 2020

As regiões diferem também nas notas médias em cada prova. É possível observar um padrão no Gráfico 4. As regiões Sudeste e Sul apresentaram as notas mais altas em todas as cinco provas, sendo o Sudeste mais bem sucedido, enquanto as regiões Nordeste e Norte apresentaram as notas mais baixas em todas as provas, sendo o Norte a região com os piores resultados.



Fonte: Elaboração própria a partir dos Microdados do Enem 2020

A partir do banco de dados foram estimados dois modelos. O primeiro utilizando o método de Mínimos Quadrados Ordinários – MQO – com variáveis explicativas *dummies* para caracterização dos indivíduos criadas a partir das informações originais da base. E o segundo a partir do método desenvolvido por Klein e Vella (2010) e Farré, Klein e Vella (2013) para estimação não viesada de modelos com endogeneidade presente.

A Tabela 3 resume as estatísticas descritivas das variáveis utilizadas. Após a exclusão das linhas com valores faltantes para as variáveis de interesse, a amostra foi composta por 480.397 observações, 8,31% do banco de dados, e resultou apenas em alunos que estavam cursando o ensino médio e o concluiriam em 2020. Percebe-se que a amostra do modelo apresentou notas mais altas em média ao se comparar com a totalidade dos presentes, a Redação, por exemplo, teve média de 621,71 pontos, frente aos 581,34 do total de presentes no exame. Em

relação às características pessoais, 56% da amostra foi composta por mulheres, 90% possuíam entre 17 e 18 anos, 48% se declaram como brancos, 50% como pretos, pardos ou indígenas (aqueles que não declaram a cor/raça foram excluídos da amostra). Quanto às escolas, 57% da amostra estudou em escolas públicas de administração estadual e 35% em escolas privadas, além disso 97% pertenciam a escola situada em área urbana. Finalmente, acerca das características socioeconômicas familiares dos indivíduos, percebe-se uma diferença entre as distribuições das escolaridades do pai e da mãe dos indivíduos na amostra, 41% dos pais têm até o Ensino Fundamental completo, enquanto 29% das mães têm esse nível de escolaridade, 35% dos pais e 38% das mães têm o Ensino Médio completo e 24% dos pais e 33% das mães possuem Ensino Superior completo. Quanto à renda familiar 67% da amostra tem até 3 salários mínimos. Além disso 64% dos indivíduos possuem computador em casa e 88% têm acesso à internet.

TABELA 3 - ESTATÍSTICAS DESCRITIVAS DAS VARIÁVEIS

(continua)

Variável	Média	Desv. Pad
Média Geral	547,70	95,04
Ciências da Natureza	528,75	95,36
Ciências Humanas	503,67	81,95
Linguagens e Códigos	538,17	72,56
Matemática	546,22	121,72
Redação	621,71	199,66
Sexo Feminino	0,56	0,50
Menor que 17 anos	0,02	0,13
17 anos	0,51	0,50
18 anos	0,39	0,49
19 anos	0,06	0,25
20 anos	0,01	0,12
21 anos	0,004	0,06
Maior que 21 anos	0,005	0,07
Branco	0,48	0,50
Preto, pardo ou indígena	0,50	0,50
Amarelo	0,02	0,14

Variável	(conclusão)	
	Média	Desv. Pad
Escola privada	0,35	0,48
Escola Federal	0,07	0,26
Escola Estadual	0,57	0,49
Escola Municipal	0,01	0,09
Escola em área urbana	0,97	0,17
Pai: EF incompleto	0,28	0,45
Pai: EF completo	0,13	0,33
Pai: EM completo	0,35	0,48
Pai: Graduação	0,14	0,35
Pai: Pós-graduação	0,10	0,30
Mãe: EF incompleto	0,18	0,39
Mãe: EF completo	0,11	0,32
Mãe: EM completo	0,38	0,48
Mãe: Graduação	0,17	0,38
Mãe: Pós-graduação	0,16	0,37
Renda familiar: até 1 salário mínimo	0,27	0,44
Renda familiar: 1 a 3 salários mínimos	0,40	0,49
Renda familiar: 3 a 5 salários mínimos	0,13	0,34
Renda familiar: 5 a 10 salários mínimos	0,12	0,32
Renda familiar: 10 a 20 salários mínimos	0,06	0,24
Renda familiar: mais de 20 salários mínimos	0,02	0,16
Possui computador	0,64	0,48
Acesso à internet	0,88	0,32

Fonte: Elaboração própria a partir dos Microdados do Enem 2020

Nota: Observações: 480.397

3.2 METODOLOGIA MODELO *DUMMIES* MQO

Baseado nos modelos de Curi e Menezes Filho (2013) e Henares, Menezes Filho e Komatsu (2021) foi estimado um modelo econométrico em que as características dos indivíduos são representadas por variáveis *dummies* construídas a partir das variáveis originais do banco de dados. As variáveis selecionadas permitem exprimir as características individuais do aluno – sexo, idade e raça -, as

características da escola – tipo de administração e localização (urbana ou rural) – e características socioeconômicas familiares – escolaridade do pai e da mãe, faixa de renda familiar, posse de computador na residência e acesso à internet.

O modelo foi elaborado a partir da equação:

$$\begin{aligned} Nota_i = & \tau + \alpha sexo_i + \beta idade_i + \gamma raça_i + \delta Adm.Esc_i + \epsilon Loc.Esc_i + \zeta Esc.Pai_i \\ & + \eta Esc.Mae_i + \theta Renda_i + \lambda Computador_i + \pi Internet_i + \varphi UF.Esc_i \\ & + \varepsilon_i \end{aligned}$$

Na equação τ representa a constante do modelo e ε_i o erro de cada observação. A variável $UF.Esc_i$, referente à Unidade da Federação em que a escola do indivíduo se localiza foi utilizada como controle, assim, o parâmetro φ não será apresentado nos resultados. A estimação dos coeficientes foi feita a partir do método dos mínimos quadrados ordinários. Foram realizadas seis regressões, uma para cada prova e uma para a média geral de cada indivíduo.

3.3 METODOLOGIA MODELO KLEIN & VELLA

Modelos econométricos que relacionam desempenho escolar com a escolaridade dos pais carregam endogeneidade. O argumento se baseia na existência de uma variável omitida, a habilidade inata do indivíduo, que estaria correlacionada com a habilidade dos pais gerando assim estimadores viesado ao se utilizar o método de Mínimos Quadrados Ordinários (Gutierrez, Moreira e Marquetti, 2021). Tradicionalmente métodos de estimação por Variáveis Instrumentais foram utilizados para lidar com essa questão, porém o método desenvolvido por Klein e Vella (2010) propõe uma estratégia de identificação do modelo na ausência de instrumentos a partir da estrutura de erros heterocedasticos. Os trabalhos de Figuerêdo, Nogueiray e Santanaz (2014) e Gutierrez, Moreira e Marquetti (2021) utilizam tal metodologia no contexto brasileiro para identificar a relação entre o desempenho estudantil e os fatores socioeconômicos dos alunos, o primeiro para o Enem de 2010 e o segundo para o Enade 2017.

A identificação do modelo se da a partir das equações

$$N_i = \beta_{10} + \beta_1 X_i + \beta_2 E_i^m + \beta_3 E_i^p + \mu_i \quad (1)$$

$$E_i^m = \beta_{20} + \gamma_2 X_i + v_i^m \quad (2)$$

$$E_i^p = \beta_{30} + \gamma_3 X_i + v_i^p \quad (3)$$

Em que N_i é o desempenho do indivíduo no Enem 2020, E_i^m a escolaridade da mãe, E_i^p a escolaridade do pai e X_i o vetor de variáveis exógenas, considerado o mesmo para as três equações. Os vetores β_s e γ_s são parâmetros desconhecidos a serem estimados e μ_i , v_i^m e v_i^p os erros dos modelos. A relação entre os erros é tal que:

$$E(\mu_i|X_i) = E(v_i^j|X_i) = 0 \quad j = m, p \quad (4)$$

$$\text{corr}(\mu_i, v_i^j) \neq 0 \quad j = m, p \quad (5)$$

De acordo com (5), os erros da equação (1) não são independentes dos erros das equações (2) e (3), evidenciando a endogeneidade presente. De acordo com Klein e Vella (2010) é possível reescrever o erro da equação (1) na forma:

$$\mu_i = \lambda^m v_i^m + \lambda^p v_i^p + \varepsilon_i \quad (6)$$

$$\lambda^j = \frac{\text{cov}(v_i^j, \mu)}{\text{var}(v_i^j)} \quad j = m, p \quad (7)$$

Sendo ε_i o termo de erro de média zero não correlacionado com as variáveis explicativas. Assim pode-se reescrever a equação original como:

$$N_i = \beta_{10} + \beta_1 X_i + \beta_2 E_i^m + \beta_3 E_i^p + \lambda^m v_i^m + \lambda^p v_i^p + \varepsilon_i \quad (8)$$

Porém a estimação da equação (8) não é possível sem a existência de restrições para a equação (6) devido à multicolinearidade das variáveis explicativas. Assim, Klein e Vella (2010) sugerem que, na presença de heterocedasticidade os coeficientes da equação (6) são dados por:

$$\lambda^j = \frac{\text{cov}(v_i^j, \mu|X)}{\text{var}(v_i^j|X)} = A^j(X_i) \quad j = m, p \quad (9)$$

Para identificação do modelo, assume-se uma restrição na correlação dos erros. Sendo $S_\mu^2(X_i)$, $S_{vm}^2(X_i)$ e $S_{vp}^2(X_i)$ as variâncias condicionais de μ , v^m e v^p , respectivamente, temos:

$$\mu_i = S_\mu(X_i) \mu_i^* \quad (10)$$

$$v_i^m = S_{vm}(X_i) v_i^{m*} \quad (11)$$

$$v_i^p = S_{vp}(X_i) v_i^{p*} \quad (12)$$

Em que μ_i^* , v_i^{m*} , v_i^{p*} são erros homocedasticos de média zero tal que:

$$E(\mu_i^* v_i^{m*} | X) = E(\mu_i^* v_i^{m*}) = \rho^m \quad (13)$$

$$E(\mu_i^* v_i^{p*} | X) = E(\mu_i^* v_i^{p*}) = \rho^p \quad (14)$$

Sendo ρ^m e ρ^p constantes, a identificação é possível caso os coeficientes da equação (6) sejam expressos por

$$\lambda^j = A^j(X_i) = \rho^j \frac{S_\mu(X_i)}{S_{vj}(X_i)} \quad j = m, p \quad (15)$$

Considerando (15), a equação (8) pode ser estimada na forma:

$$N_i = \beta_{10} + \beta_1 X_i + \beta_2 E_i^m + \beta_3 E_i^p + \rho^m \frac{S_\mu(X_i)}{S_{vm}(X_i)} v_i^m + \rho^p \frac{S_\mu(X_i)}{S_{vp}(X_i)} v_i^p + \varepsilon_i \quad (16)$$

Para identificar a forma da heterocedasticidade dos erros, segue-se a abordagem paramétrica proposta por Farré, Klein e Vella (2013):

$$S_\mu = \sqrt{\exp(Z_\mu \theta_1)} \quad (17)$$

$$S_{vj} = \sqrt{\exp(Z_{vj} \theta_{2j})} \quad j = m, p \quad (18)$$

Em que Z_μ e Z_{vj} são os vetores de variáveis que causam heterocedasticidade nas equações (1), (2) e (3), e θ_1 e θ_{2j} , para $j = m, p$, são parâmetros desconhecidos. O procedimento de estimação utilizado é o mesmo sugerido por Farré, Klein e Vella (2013), adotado por Gutierrez, Moreira e Marquetti (2021), considerando para Z_μ e Z_{vj} as mesmas variáveis de X_i . Segue:

- i. Estima-se por mínimos quadrados ordinários as equações (2) e (3) para obtenção dos resíduos \widehat{v}_i^j para $j = m, p$
- ii. Estima-se a equação (18) pelo método de mínimos quadrados não lineares utilizando como variável dependente $\ln[(\widehat{v}_i^j)^2]$ para obtenção de θ_{2j} , para $j = m, p$
- iii. Computa-se as estimativas para $\widehat{S_{vj}}(X_i)$ a partir dos θ_{2j} calculados, para $j = m, p$
- iv. Assumindo (17) e os valores obtidos em (i) e (iii) estima-se a expressão abaixo pelo método de mínimos quadrados não lineares

$$\min_{\beta_0, \beta_1, \beta_2, \beta_3, \rho^m, \rho^p, \theta_1} \sum_{i=1}^n (N_i - \beta_0 - \beta_1 X_i - \beta_2 E_i^m - \beta_3 E_i^p - \rho^m \sqrt{\exp(X_i \theta_1)} \frac{\widehat{v}_i^m}{\widehat{S_{vm}}(X_i)} - \rho^p \sqrt{\exp(X_i \theta_1)} \frac{\widehat{v}_i^p}{\widehat{S_{vp}}(X_i)} - \varepsilon_i)^2$$

As variáveis utilizadas são descritas na Tabela 4.

TABELA 4 - DESCRIÇÃO DE VARIÁVEIS MODELO KV

Variável	Descrição
Escola privada	<i>Dummy</i> para escola privada
Escola urbana	<i>Dummy</i> para escola urbana
Raça: Branco	<i>Dummy</i> para indivíduo declarado como branco
Sexo Feminino	<i>Dummy</i> para indivíduos do sexo feminino
Renda familiar	Dividida em 5 estratos: <ol style="list-style-type: none"> 1. Até 1 salário mínimo 2. De 1 a 3 salários mínimos 3. De 3 a 6 salários mínimos 4. De 6 a 9 salários mínimos 5. Mais de 9 salários mínimos
Escolaridade dos pais	Anos de estudo por cada etapa de ensino: <ol style="list-style-type: none"> 1. Nunca estudou ou não completou até 4ª série/5ºano do ensino fundamental – 0 anos 2. Completou até 4ª série/5ºano do ensino fundamental – 4 anos 3. Completou até 4ª série/5ºano do ensino fundamental – 5 anos 4. Completou até 8ª série/9ºano do ensino fundamental – 9 anos 5. Completou o Ensino Médio – 12 anos 6. Completou graduação – 17 anos 7. Completou pós-graduação – 20 anos

Fonte: Elaboração própria a partir dos Microdados do Enem 2020

4 RESULTADOS

4.1 RESULTADOS MODELO *DUMMIES* MQO

Para a apresentação dos resultados do modelo de variáveis *dummies* as tabelas de coeficientes foram divididas em três partes buscando melhor visualização. Primeiramente se apresentam os coeficientes das características pessoais, em seguida escolares e por último familiares.

A Tabela 5 sumariza os resultados a respeito das características pessoais dos indivíduos. Foi observado que na nota média geral do exame as mulheres apresentaram desempenho ligeiramente pior que os homens, 2,32 pontos abaixo, porém tal resultado não é o mesmo entre todas as 5 provas. Nas provas de Ciências da natureza, Ciências Humanas e Matemática em média os homens obtêm um melhor desempenho. As provas de Matemática e Redação foram as que apresentaram maior disparidade entre as notas de homens e mulheres, enquanto na primeira os homens atingem 41,24 pontos a mais, na última as mulheres se sobressairam por 58,90 pontos.

TABELA 5 – RESULTADOS CARACTERISTICAS PESSOAIS MODELO MQO *DUMMIES*

(continua)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
Sexo Feminino	-2,32*** (0,22)	-18,75*** (0,20)	-11,64*** (0,24)	1,13*** (0,18)	-41,24*** (0,28)	58,90*** (0,51)
17 anos	-3,76*** (0,83)	-0,12 (0,76)	-1,62* (0,91)	-3,10*** (0,69)	-8,70*** (1,09)	-5,25*** (1,95)
18 anos	-12,68*** (0,84)	-4,80*** (0,76)	-7,35*** (0,91)	-9,20*** (0,69)	-20,44*** (1,10)	-21,59*** (1,96)
19 anos	-40,90*** (0,92)	-20,17*** (0,84)	-29,42*** (1,00)	-28,85*** (0,76)	-48,51*** (1,21)	-77,57*** (2,16)

(conclusão)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
20 anos	-54,68*** (1,17)	-24,94*** (1,06)	-39,65*** (1,27)	-39,75*** (0,96)	-58,34*** (1,53)	-110,71*** (2,73)
21 anos	-63,57*** (1,69)	-29,16*** (1,54)	-42,52*** (1,84)	-47,45*** (1,40)	-61,14*** (2,21)	-137,58*** (3,96)
Maior que 21 anos	-70,21*** (1,29)	-25,81*** (1,18)	-42,08*** (1,41)	-50,24*** (1,07)	-63,13*** (1,69)	-169,82*** (3,03)
Preto, pardo ou indígena	-14,47*** (0,25)	-11,20*** (0,23)	-12,73*** (0,27)	-9,03*** (0,20)	-18,76*** (0,32)	-20,66*** (0,58)
Amarelo	-2,81*** (0,75)	1,53** (0,68)	-5,34*** (0,82)	-5,27*** (0,62)	3,68*** (0,98)	-8,67*** (1,76)
Constante	516,56*** (1,20)	500,23*** (1,09)	510,23*** (1,30)	507,77*** (0,99)	549,82*** (1,57)	514,74*** (2,80)
Observações	480.397	480.397	480.397	480.397	480.397	480.397
R ²	0,40	0,33	0,30	0,30	0,37	0,26
R ² Ajustado	0,40	0,33	0,30	0,30	0,37	0,26
Erro pad. residual (df = 480342)	73,79	67,16	80,22	61,04	96,56	172,66
Estatística F (df = 54; 480342)	6.011,00***	4.368,07***	3.739,12***	3.847,49***	5.264,84***	3.154,44***

Fonte: Elaboração própria com base nos Microdados do Enem 2020

Nota: *p<0,1; **p<0,05; ***p<0,01; Erros padrão entre parênteses

Pode-se notar em todas as provas que alunos mais velhos, em média, atingem resultados mais baixos. A base de comparação entre as idades foram alunos com até 16 anos, como todos os coeficientes relacionados às outras idades são negativos, conclui-se que esses alunos possuem os melhores desempenhos na prova. Nota-se também que as maiores diferenças ocorrem entre as idades de 18 e 19 anos, tal fato pode estar relacionado à idade em que se completa o ensino médio

no Brasil, entre 17 e 18 anos para alunos que não repetem séries. Visto que a amostra é composta apenas por indivíduos que completaram o Ensino Médio em 2020 aqueles que possuem 19 anos ou mais provavelmente atrasaram a sua educação básica, assim tem-se um indício de que tal atraso está relacionado a piores desempenhos no exame.

No que diz respeito à raça, alunos não brancos apresentaram um desempenho inferior no exame, com duas exceções, alunos declaradamente amarelos tiveram melhores notas em média nas provas de Ciências da Natureza e Matemática, superando os brancos em 1,53 e 3,68 pontos respectivamente nas duas provas. Alunos declarados pretos, pardos ou indígenas obtiveram resultados piores que os brancos e amarelos em todas as provas, evidenciando um possível déficit educacional histórico no país.

Acerca das características escolares participantes de escolas públicas com administração federal apresentaram os melhores desempenhos na média geral e nas cinco provas. Alunos de escolas privadas tiveram melhores rendimentos em comparação com aqueles de escolas estaduais e municipais, sendo os últimos aqueles com os piores resultados. Quanto à localização os alunos de escolas em área urbana se saíram melhor no exame. A Tabela 6 mostra as estimativas relacionadas às características escolares

TABELA 6 - RESULTADOS CARACTERISTICAS ESCOLARES MODELO MQO *DUMMIES*

(continua)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
Escola Federal	26,92*** (0,47)	19,17*** (0,42)	27,04*** (0,51)	24,46*** (0,39)	37,76*** (0,61)	26,15*** (1,09)
Escola Estadual	-48,93*** (0,29)	-39,21*** (0,27)	-36,97*** (0,32)	-23,34*** (0,24)	-47,99*** (0,39)	-97,12*** (0,69)
Escola Municipal	-37,16*** (1,23)	-28,98*** (1,12)	-25,54*** (1,34)	-14,58*** (1,02)	-32,27*** (1,62)	-84,46*** (2,89)

(conclusão)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
Escola em área urbana	17,06*** (0,65)	9,40*** (0,59)	13,45*** (0,71)	15,50*** (0,54)	16,94*** (0,85)	30,02*** (1,52)
Constante	516,56*** (1,20)	500,23*** (1,09)	510,23*** (1,30)	507,77*** (0,99)	549,82*** (1,57)	514,74*** (2,80)
Observações	480.397	480.397	480.397	480.397	480.397	480.397
R ²	0,40	0,33	0,30	0,30	0,37	0,26
R ² Ajustado	0,40	0,33	0,30	0,30	0,37	0,26
Erro pad. residual (df = 480342)	73,79	67,16	80,22	61,04	96,56	172,66
Estatística F (df = 54; 480342)	6.011,00***	4.368,07***	3.739,12***	3.847,49***	5.264,84***	3.154,44***

Fonte: Elaboração própria com base nos Microdados do Enem 2020

Nota: *p<0,1; **p<0,05; ***p<0,01; Erros padrão entre parênteses

A respeito das características socioeconômicas familiares ilustradas na Tabela 7 observa-se uma progressão não linear dos coeficientes relacionados à escolaridade dos pais. A base de comparação foram os alunos com pais que não concluíram o ensino fundamental. De maneira geral, nota-se uma correlação positiva entre a escolaridade dos progenitores e o desempenho do indivíduo, com diferenças progressivamente maiores à medida que a escolaridade dos pais aumenta até uma estabilização da diferença entre pais com graduação completa e pós-graduação. Ou seja, a diferença entre alunos de pais com ensino fundamental completo e ensino médio completo é menor que a diferença entre alunos com pais com ensino médio completo e pais com graduação completa, que por sua vez é maior que a diferença entre os alunos de pais com graduação completa e pais com pós-graduação. Por exemplo, a discrepância entre os estratos de mães com ensino médio completo e graduação completa é de 7,69 pontos na média geral, enquanto a discrepância entre os estrados de mãe com graduação e mãe com pós-graduação é de -0,33 pontos.

Percebe-se também uma maior influência do nível de escolaridade dos pais em relação ao nível das mães.

TABELA 7 – RESULTADOS CARACTERÍSTICAS FAMILIARES MODELO MQO *DUMMIES*

(continua)

Variável dependente: Nota						
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
Pai: EF completo	3,81*** (0,38)	1,73*** (0,34)	2,40*** (0,41)	4,75*** (0,31)	2,61*** (0,49)	7,56*** (0,88)
Pai: EM completo	8,63*** (0,31)	5,47*** (0,29)	7,19*** (0,34)	8,78*** (0,26)	7,69*** (0,41)	14,01*** (0,73)
Pai: Graduação	20,03*** (0,43)	15,46*** (0,39)	18,17*** (0,46)	15,62*** (0,35)	22,77*** (0,56)	28,10*** (1,00)
Pai: Pós-graduação	21,93*** (0,50)	17,01*** (0,45)	19,77*** (0,54)	15,97*** (0,41)	24,10*** (0,65)	32,79*** (1,16)
Mãe: EF completo	2,80*** (0,42)	0,71* (0,38)	1,40*** (0,45)	4,35*** (0,35)	2,16*** (0,55)	5,36*** (0,98)
Mãe: EM completo	7,84*** (0,35)	4,27*** (0,32)	5,38*** (0,38)	8,37*** (0,29)	7,56*** (0,45)	13,61*** (0,81)
Mãe: Graduação	15,53*** (0,43)	10,50*** (0,39)	12,35*** (0,47)	12,35*** (0,36)	18,57*** (0,57)	23,87*** (1,02)
Mãe: Pós-graduação	15,20*** (0,46)	10,23*** (0,42)	12,27*** (0,50)	11,40*** (0,38)	16,73*** (0,60)	25,36*** (1,07)
Renda familiar: 1 a 3 salários mínimos	14,71*** (0,31)	9,13*** (0,28)	11,73*** (0,34)	12,03*** (0,26)	16,49*** (0,41)	24,16*** (0,73)
Renda familiar: 3 a 5 salários mínimos	27,23*** (0,44)	19,26*** (0,40)	22,42*** (0,48)	18,75*** (0,37)	34,72*** (0,58)	41,01*** (1,04)

(conclusão)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
Renda familiar: 5 a 10 salários mínimos	38,71*** (0,49)	28,56*** (0,45)	32,27*** (0,54)	24,53*** (0,41)	51,25*** (0,64)	56,92*** (1,15)
Renda familiar: 10 a 20 salários mínimos	54,94*** (0,62)	41,41*** (0,57)	46,96*** (0,68)	33,12*** (0,51)	76,03*** (0,81)	77,19*** (1,45)
Renda familiar: mais de 20 salários mínimos	72,54*** (0,83)	54,09*** (0,76)	63,80*** (0,90)	43,08*** (0,69)	106,69*** (1,09)	95,02*** (1,95)
Possui computador	18,67*** (0,28)	14,14*** (0,26)	17,04*** (0,31)	12,18*** (0,24)	18,19*** (0,37)	31,81*** (0,67)
Acesso à internet	9,57*** (0,38)	3,64*** (0,35)	6,68*** (0,42)	9,72*** (0,32)	5,60*** (0,50)	22,22*** (0,90)
Constante	516,56*** (1,20)	500,23*** (1,09)	510,23*** (1,30)	507,77*** (0,99)	549,82*** (1,57)	514,74*** (2,80)
Observações	480.397	480.397	480.397	480.397	480.397	480.397
R ²	0,40	0,33	0,30	0,30	0,37	0,26
R ² Ajustado	0,40	0,33	0,30	0,30	0,37	0,26
Erro pad. residual (df = 480342)	73,79	67,16	80,22	61,04	96,56	172,66
Estatística F (df = 54; 480342)	6.011,00***	4.368,07***	3.739,12***	3.847,49***	5.264,84***	3.154,44***

Fonte: Elaboração própria com base nos Microdados do Enem 2020

Nota: *p<0,1; **p<0,05; ***p<0,01; Erros padrão entre parênteses

O nível de renda familiar apresentou uma correlação positiva com as notas obtidas no exame. A diferença entre os estratos foi mais acentuada na prova de Redação com exceção do último, renda família acima de 20 salários mínimos, em que a maior coeficiente observado foi na prova de matemática sendo esse o maior coeficiente observado entre as 6 regressões, atingindo uma diferença de 106,69

pontos em relação a indivíduos com renda familiar abaixo de um salário mínimo. A posse de computador e o acesso à internet também resultaram em efeitos positivos para o desempenho, sendo o primeiro mais expressivo.

De maneira geral as estimativas mostram-se significantes a partir dos testes t e F , com poucas exceções para o primeiro, estas acontecendo em cenários em que o valor estimado foi próximo de zero. Dentre as seis regressões aquela que apresentou maiores diferenças entre os grupos em geral foi a prova de Redação. Em geral homens brancos com alta renda familiar e alto nível de escolaridade dos pais foram aqueles que performaram melhor no exame.

4.2 RESULTADOS MODELO KLEIN & VELLA

Para o Modelo Klein & Vella os resultados são apresentados em duas tabelas, a Tabela 8 expõe os resultados para média geral, Ciências da Natureza e Ciências Humanas, a Tabela 9 para Linguagens e Códigos, Matemática e Redação. São mostradas em cada tabela estimativas obtidas utilizando o método de mínimos quadrados ordinários (MQO) e o método proposto por Klein e Vella (2010) e Farré, Klein e Vella (2013) (KV). As variáveis ρ mãe e ρ pai captam o efeito indireto da escolaridade dos pais sobre o desempenho escolar dos indivíduos, representando uma propagação de habilidades entre gerações. A diferença entre as estimativas MQO e KV refletem o viés das primeiras devido à endogeneidade presente.

Os efeitos indiretos da escolaridade dos pais não se mostraram significativos para a média geral em níveis de significância inferiores a 10%, porém os efeitos diretos foram ligeiramente menores no método que controla para a endogeneidade. Ao se comparar os coeficientes obtidos para os anos de escolaridade da mãe e do pai, nota-se um valor ligeiramente maior para o primeiro, a diferença foi de 0,08 pontos por ano de estudo pelo método de Klein e Vella (2010). Em termos do efeito indireto não houve diferença entre os efeitos da mãe e do pai. De modo geral alunos homens, brancos, de escola privada urbana com alta renda familiar e alta escolaridade dos pais obtiveram melhores resultados no exame.

TABELA 8 - RESULTADOS MODELO KLEIN & VELLA I

	Variável dependente: Nota					
	Média Geral		Ciências da Natureza		Ciências Humanas	
	MQO	KV	MQO	KV	MQO	KV
Escola privada	42,77*** (0,29)	43,80*** (0,31)	33,34*** (0,26)	35,83*** (0,40)	30,68*** (0,31)	31,82*** (0,33)
Escola Urbana	3,33*** (0,67)	3,72*** (0,68)	-1,33** (0,60)	-0,11 (0,62)	2,35*** (0,72)	2,79*** (0,72)
Raça: Branco	17,60*** (0,24)	17,87*** (0,24)	14,00*** (0,21)	14,79*** (0,24)	18,06*** (0,26)	18,37*** (0,26)
Sexo feminino	-2,20*** (0,23)	-2,41*** (0,23)	-18,80*** (0,20)	-19,22*** (0,21)	-11,48*** (0,24)	-11,69*** (0,24)
Renda familiar	18,49*** (0,13)	19,63*** (0,15)	14,83*** (0,11)	17,27*** (0,31)	17,14*** (0,13)	18,43*** (0,17)
Anos escolaridade mãe	1,76*** (0,03)	1,58*** (0,03)	1,12*** (0,02)	0,60*** (0,08)	1,38*** (0,03)	1,16*** (0,04)
Anos escolaridade pai	1,64*** (0,02)	1,50*** (0,03)	1,18*** (0,02)	0,73*** (0,07)	1,46*** (0,03)	1,31*** (0,04)
ρ mãe		0,02* (0,01)		0,30*** (0,07)		0,04** (0,02)
ρ pai		0,02* (0,01)		0,27*** (0,07)		0,03** (0,01)
Constante	458,39*** (0,71)	460,02*** (0,74)	451,13*** (0,63)	456,79*** (1,10)	458,18*** (0,75)	460,13*** (0,81)
Observações	489.408	489.408	489.408	489.408	489.408	489.408
R ²	0,33	0,33	0,28	0,28	0,24	0,24
R ² Ajustado	0,33	0,33	0,28	0,28	0,24	0,24
Estatística F (df = 7; 489400)	34.359,97***	34.359,97***	27.062,19***	27.062,19***	22.684,15***	22.684,15***

Fonte: Elaboração própria com base nos Microdados do Enem 2020

Nota: *p<0,1; **p<0,05; ***p<0,01; Erros padrão entre parênteses

Para a prova de Ciências da Natureza os resultados obtidos, para a localização da escola não se mostraram significativos ao se controlar para a endogeneidade. Foi observado também um efeito indireto positivo significativo, ao nível de 1%, para a escolaridade da mãe e do pai do aluno. Os efeitos diretos se mostraram notavelmente menores nas estimativas com o método de Klein e Vella (2010) sendo o valor relacionado ao pai maior que o da mãe.

Acerca da prova de Ciências Humanas os efeitos indiretos da escolaridade dos pais foram significativos ao nível de 5%. Para a escolaridade do pai os efeitos diretos foram cerca de 43 vezes maiores que os indiretos, enquanto para a mãe a razão foi de 29 vezes.

Nos resultados para a prova de Linguagens e Códigos, nota-se que o coeficiente relacionado a escola privada é expressivamente menor ao comparado com as outras provas, indicando um melhor desempenho de alunos de escolas públicas nesta prova. Além disso o coeficiente para sexo feminino indica que as mulheres se saem ligeiramente melhores que os homens nesta competência. Os efeitos indiretos da escolaridade apresentaram sinais negativos tanto para a mãe quanto para o pai, porém não foram significativos, enquanto os efeitos diretos foram positivos e 1,6 vez maior para a mãe em relação ao pai.

TABELA 9 - RESULTADOS MODELO KLEIN & VELLA II

(continua)

	Variável dependente: Nota					
	Linguagens e Códigos		Matemática		Redação	
	MQO	KV	MQO	KV	MQO	KV
Escola privada	18,99*** (0,24)	16,27*** (1,59)	40,05*** (0,37)	42,58*** (0,46)	90,78*** (0,66)	81,84*** (3,19)
Escola Urbana	8,96*** (0,55)	7,74*** (0,87)	1,28 (0,87)	2,39*** (0,88)	5,38*** (1,55)	1,11 (2,07)
Raça: Branco	15,08*** (0,20)	14,19*** (0,55)	24,41*** (0,31)	25,17*** (0,32)	16,45*** (0,55)	13,51*** (1,19)
Sexo feminino	1,33*** (0,18)	1,81*** (0,32)	-40,76*** (0,29)	-41,17*** (0,30)	58,71*** (0,52)	60,26*** (0,74)

(conclusão)

	Variável dependente: Nota					
	Linguagens e Códigos		Matemática		Redação	
	MQO	KV	MQO	KV	MQO	KV
Renda familiar	11,94*** (0,10)	9,00*** (1,57)	25,81*** (0,16)	28,40*** (0,29)	22,73*** (0,29)	13,13*** (3,13)
Anos escolaridade mãe	1,34*** (0,02)	2,53*** (0,66)	1,76*** (0,03)	1,22*** (0,08)	3,23*** (0,06)	6,66*** (1,13)
Anos escolaridade pai	1,40*** (0,02)	1,61*** (0,14)	1,74*** (0,03)	1,35*** (0,06)	2,44*** (0,06)	3,66*** (0,45)
ρ mãe		-0,93 (0,70)		0,19*** (0,06)		-2,79** (1,21)
ρ pai		-0,18 (0,15)		0,14*** (0,04)		-1,05** (0,49)
Constante	467,65*** (0,58)	457,27*** (5,73)	467,56*** (0,91)	472,83*** (1,18)	447,44*** (1,62)	414,12*** (10,95)
Observações	489.408	489.408	489.408	489.408	489.408	489.408
R ²	0,24	0,24	0,31	0,31	0,20	0,20
R ² Ajustado	0,24	0,24	0,31	0,31	0,20	0,20
Estatística F (df = 7; 489400)	21.858,82***	21.858,82***	32.024,72***	32.024,72***	17.657,08***	17.657,08***

Fonte: Elaboração própria com base nos Microdados do Enem 2020

Nota: *p<0,1; **p<0,05; ***p<0,01; Erros padrão entre parênteses

Na prova de Matemática, ao se controlar para a endogeneidade, o efeito direto foi maior para a escolaridade do pai em comparação com a escolaridade da mãe, por 0,13 pontos por ano. Enquanto os efeitos indiretos foram ligeiramente maiores para as mães. Ambos os efeitos diretos e indiretos foram significativos ao nível de 1%, sendo os diretos 6,4 vezes maiores para as mães e 9,6 vezes para os pais.

A prova de Redação foi a que obteve os maiores coeficientes para escola privada e para o sexo feminino, 81,84 e 60,26 pontos respectivamente, para o método que controla para endogeneidade, sendo assim a única em que as mulheres

performam expressivamente melhor que os homens. Em relação aos efeitos diretos da escolaridade dos pais as estimativas obtidas pelo método de Klein e Vella (2010) foram consideravelmente maiores que as outras provas, isso se deu pois os efeitos indiretos resultaram em valores negativos indicando uma correlação negativa entre as habilidades dos pais e dos filhos. Os efeitos indiretos foram significativos ao nível de 5%. Ao se comparar os coeficientes dos anos de estudo das mães e dos pais, nota-se uma estimativa 1,8 vez maior para os efeitos das mães, apresentando o maior resultado entre todas as provas, de 6,66 pontos por ano.

Em todas as provas bem como na média geral os melhores resultados foram obtidos por alunos brancos de escola privada apresentando uma correlação positiva entre níveis mais altos de renda e escolaridade dos pais. Com exceção das provas de Linguagens e Códigos e Redação os homens atingiram maiores notas.

5 DISCUSSÃO

Ao se comparar os resultados com a literatura levantada no capítulo 2 é possível notar grandes semelhanças e algumas diferenças entre os trabalhos. Este capítulo se dedica a discutir tais comparações baseado nas variáveis utilizadas e características analisadas e caminhos para políticas a partir dos resultados.

5.1 COMPARAÇÃO DE RESULTADOS

Quanto ao resultado das regiões brasileiras nota-se uma diferença no ranqueamento das notas obtidas na Redação em relação ao trabalho de Viggiano e Matos (2013). No presente trabalho foi observado que no Enem 2020 as notas regionais mantiveram a ordem decrescente nas 5 provas e na média geral, esta ordem sendo Sudeste, Sul, Centro-Oeste, Nordeste e Norte, a mesma observada de maneira geral por Viggiano e Matos (2013) para o Enem 2010. Porém os autores observaram que na prova de Redação tal ordem foi alterada com as regiões Nordeste e Norte apresentando melhores notas que as regiões Centro-Oeste e Sul, o que não foi observado para o Enem 2020.

Sobre a diferença do desempenho entre os sexos, os trabalhos analisados e o presente apresentaram conclusões diversas. Enquanto Curi e Menezes Filho (2013), Henares, Menezes Filho e Komatsu (2021) e Jaloto e Primi (2021) observaram um melhor desempenho geral dos homens, o mesmo resultado observado aqui, Figuerêdo, Nogueiray e Santanaz (2014) observaram um melhor desempenho das mulheres no Enem. Ainda, Henares, Menezes Filho e Komatsu (2021) ao analisarem as provas separadamente concluem que as mulheres tiveram melhores notas em Linguagens e Códigos e Redação em 2015, o que condiz com os resultados encontrados para o Enem 2020, já Jaloto e Primi (2021), que não analisaram a prova de Redação, encontraram rendimentos superiores dos homens, em média, nas quatro provas objetivas no Enem 2018. Tais diferenças podem ser resultado de características não observadas como a motivação e o nível de esforço que podem ter tido diferenças entre os anos e os sexos.

Em relação à raça os resultados são claros. Mesmo controlado por renda e escolaridade dos pais, os indivíduos declarados negros apresentam piores desempenhos. Curi e Menezes Filho (2013), Henares, Menezes Filho e Komatsu

(2021), Jaloto e Primi (2021) e Figuerêdo, Nogueiray e Santanaz (2014) obtiveram esta conclusão para a prova do Enem e Gutierrez, Moreira e Marquetti (2021) apresenta a mesma conclusão para o Ensino Superior, baseado no Enade. No presente trabalho foram agrupados os indivíduos pretos, pardos e indígenas no mesmo grupo, este apresentou os piores resultados em relação à raça, frente aos brancos e aos amarelos, Tal resultado traz indícios de uma defasagem racial histórica na educação no Brasil.

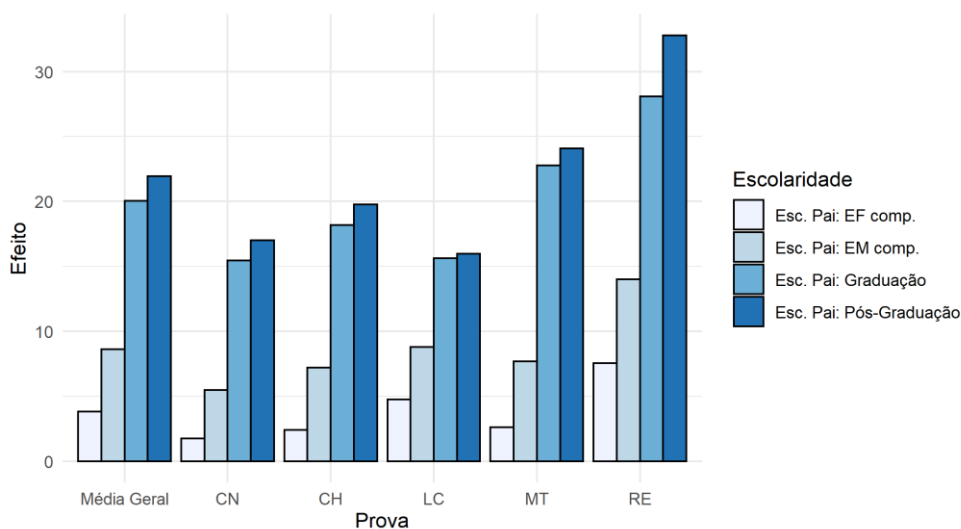
No tocante à idade, os trabalhos de Curi e Menezes Filho (2013) e Henares, Menezes Filho e Komatsu (2021) apresentaram uma relação indireta entre a idade e o desempenho no exame, ou seja, alunos mais velhos tiveram piores desempenhos, mesmo resultado obtido para o Enem 2020. Já Jaloto e Primi (2021) incluíram a idade dos participantes no modelo para observar o efeito do atraso na formação escolar sobre a nota, os autores encontraram também um coeficiente negativo associado à variável apontando que alunos que atrasaram a sua formação no Ensino Médio tiveram piores resultados na prova. Tal relação pode ser observada no presente trabalho pelo salto nos valores dos coeficientes a partir de 19 anos obtidos no primeiro modelo. Assim, o indício é de que alunos que concluem o Ensino Médio no tempo previsto apresentam melhores resultados no Enem.

Os resultados relacionados ao tipo de administração da escola apontam, em média uma superioridade das escolas privadas. O trabalho de Figuerêdo, Nogueiray e Santanaz (2014), bem como o segundo modelo deste trabalho compararam as escolas privadas com as públicas sem distinção entre escolas municipais, estaduais e federais, ambos observando desempenhos superiores das escolas privadas no Enem. Gutierrez, Moreira e Marquetti (2021) realizaram a mesma comparação para o Ensino Superior e obtiveram resultados semelhantes. Já Henares, Menezes Filho e Komatsu (2021), Jaloto e Primi (2021) e o primeiro modelo deste trabalho segregaram as escolas públicas pelas esferas administrativas, observando, para todas as provas analisadas, a seguinte ordem decrescente de notas no Enem: escolas federais, escolas privadas, escolas municipais, escolas estaduais. Logo, de forma geral as escolas públicas apresentam piores resultados nos testes, porém, especificamente as escolas federais superam as privadas.

A respeito da escolaridade dos pais, Curi e Menezes Filho (2013) incluíram apenas a escolaridade da mãe no modelo obtendo um efeito positivo. Henares, Menezes Filho e Komatsu (2021) incluíram ambas as escolaridades do pai e da mãe

obtendo um efeito de maior magnitude relacionado à educação dos pais e concluíram que a prova com maiores coeficientes foi a Redação, como pôde ser visto nos gráficos 1 e 2. Os resultados encontrados no primeiro modelo estão de acordo com os resultados de Henares, Menezes Filho e Komatsu (2021), como pode ser visto nos gráficos 5 e 6 os efeitos relacionados à escolaridade do pai superam aqueles relacionados à escolaridade da mãe e a prova com os maiores coeficientes foi a Redação.

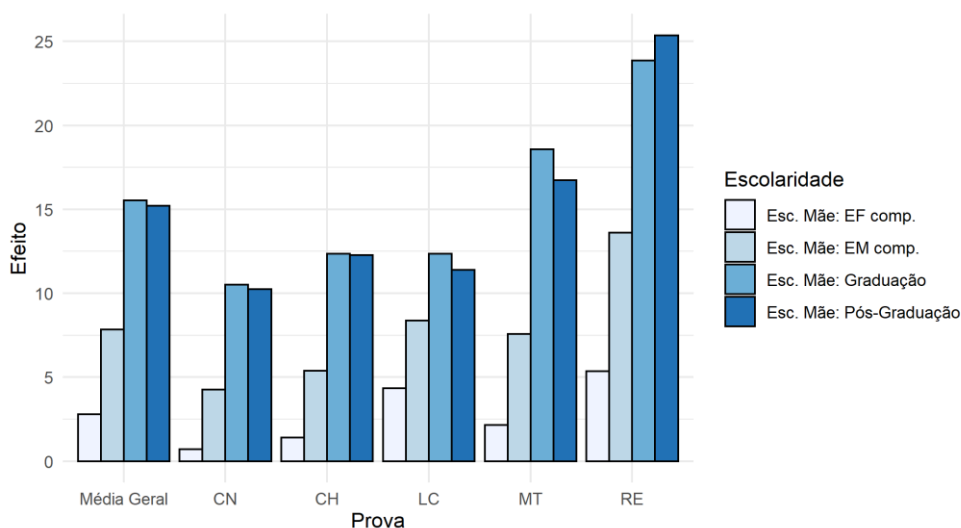
GRÁFICO 5 - INFLUÊNCIA DA ESCOLARIDADE DO PAI - ENEM 2020



Fonte: Elaboração própria com base nos Microdados do Enem 2020

Nota: Em comparação ao Ensino Fundamental incompleto

GRÁFICO 6 - INFLUÊNCIA DA ESCOLARIDADE DA MÃE - ENEM 2020



Fonte: Elaboração própria com base nos Microdados do Enem 2020

Nota: Em comparação ao Ensino Fundamental incompleto

Quanto ao segundo modelo estimado, baseado em Klein e Vella (2010), os resultados apontam também uma relação direta positiva entre as escolaridades dos pais e os resultados dos filhos no Enem, mesmo ao se controlar para a endogeneidade. Nota-se que o padrão observado no primeiro modelo, maior influência da educação do pai, não se mantém para todas as provas no segundo modelo, visto que o efeito direto relacionado à escolaridade da mãe foi maior na média geral e nas provas de Linguagens e Códigos e Redação.

Quanto aos efeitos indiretos, os resultados encontrados no presente trabalho e aqueles encontrados por Figuerêdo, Nogueiray e Santanaz (2014) e Gutierrez, Moreira e Marquetti (2021) foram diferentes. Figuerêdo, Nogueiray e Santanaz (2014) observaram efeitos indiretos para a educação da mãe maiores que os efeitos diretos e para a educação do pai os efeitos indiretos não foram significativos, enquanto Gutierrez, Moreira e Marquetti (2021) não obtiveram efeitos indiretos significativos nem para o pai nem para a mãe. No presente trabalho os efeitos indiretos foram significativos ao nível de 1% para as provas de Ciências da Natureza e Matemática, ao nível de 5% para Ciências Humanas e Redação, ao nível de 10% para a média geral e para a prova de Linguagens e Códigos não foram significativos, porém em nenhum dos casos os efeitos indiretos superaram os diretos. Nota-se também um coeficiente negativo para os efeitos indiretos tanto da escolaridade do pai quanto da mãe na prova de Redação, indicando uma possível inversão de habilidades entre gerações nesta prova. É importante ressaltar que tão parâmetro pode estar relacionando também a variáveis não observadas como a motivação e o esforço, logo uma investigação mais profunda, relacionada à medição de tais variáveis se faz necessária para estudos posteriores.

Sobre a renda familiar, os dois modelos estimados, bem como todos os trabalhos analisados indicam uma relação positiva entre o aumento da renda e o aumento do desempenho. Tal fato pode ser um indício de que à medida em que as necessidades básicas são supridas com maiores facilidades o investimento em capital humano através da educação ganha cada vez mais espaço dentro da família.

De forma geral, observa-se uma concordância entre o presente trabalho e os outros analisados e pode-se concluir que níveis socioeconômicos mais baixos estão relacionados a piores desempenhos escolares.

5.2 POLÍTICAS VOLTADAS À EDUCAÇÃO

Três fatores de grande importância para o desempenho acadêmico são a esfera de administração da escola, a raça do estudante e a renda familiar, os quais são abordados na Lei de Cotas (Brasil, 2012). Tal lei faz com que 50% das vagas das instituições federais de ensino superior sejam destinadas a estudantes que cursaram o ensino médio integralmente em escolas públicas, das quais 50% são destinadas a estudantes com renda familiar de no máximo 1,5 salário mínimo, além de ser reservado um percentual destas para pessoas autodeclaradas pretas, pardas ou indígenas iguais ao percentual destas na UF em questão. Esta política está de acordo com os resultados encontrados na medida em que os fatores abordados estão de fato relacionados a piores desempenhos escolares, podendo gerar também um efeito de propagação de ganhos visto que ela atua para um aumento da escolaridade de classes desfavorecidas e como a escolaridade dos pais é correlacionada positivamente com o desempenho acadêmico dos filhos o efeito se propaga entre gerações.

Programas de distribuição de renda como o Programa Bolsa-Família que garante um benefício monetário a famílias em situação de pobreza podem contribuir para uma melhoria da educação à medida que, com a garantia do suprimento das necessidades básicas o investimento em educação é viabilizado. Além de que o programa tinha como pré-requisito que as crianças e adolescente da família estejam matriculados e frequentem a escola, garantindo assim que os jovens tenham acesso a educação e que a família tenha melhores condições socioeconômicas. Um nível de renda familiar mais alto por si só já está correlacionado a um melhor desempenho acadêmico e possibilita acesso a instrumentos como computadores e a internet que de acordo com o primeiro modelo estimado possuem um efeito positivo no desempenho dos alunos.

Um ponto de grande importância é a melhoria das escolas públicas. De acordo com Albernaz, Ferreira e Franco (2002) fatores importantes para a diferença no desempenho dos alunos entre escolas são a infraestrutura e a saúde financeira da escola. Assim, para equiparar o desempenho das escolas públicas ao das privadas investimentos na infraestrutura escolar e a revisão das políticas de financiamento são caminhos a serem seguidos. Para entender melhor a situação um estudo das diferenças entre escolas federais, estaduais e municipais é um bom

ponto de partida, visto que as primeiras foram aquelas que apresentaram melhores resultados em média dentre todas as escolas participantes no Enem 2020.

Em termos de políticas para a melhoria educacional do Brasil um conjunto de práticas com focos distintos seriam um bom caminho a ser seguido. Políticas como a Lei de Cotas (Brasil, 2012) são úteis para um efeito de médio prazo à medida que garantem o acesso à educação e aumento da escolaridade de parcelas da população com piores condições de investimento em educação. Por outro lado, políticas de distribuição de renda e melhoria das escolas públicas atuam para diminuir as desigualdades no acesso à educação colaborando para uma equalização do nível educacional a longo prazo.

6 CONCLUSÕES

O presente trabalho se dedicou a investigar como o ambiente socioeconômico em que o estudante está inserido influencia o seu desempenho acadêmico medido a partir das notas obtidas no Enem. Para isso foram estimados dois modelos econométricos, o primeiro se utilizou de variáveis *dummies* e o método dos Mínimos Quadrados Ordinários para analisar as diferenças de desempenhos a partir de características pessoais, escolares e familiares, o segundo se utilizou da metodologia proposta por Klein e Vella (2010) e Farré, Klein e Vella (2013) para controlar a endogeneidade presente na relação entre a escolaridade dos pais e o desempenho do aluno.

Os resultados apontaram que existe influência do *background* socioeconômico no desempenho acadêmico. A raça, a esfera administrativa da escola e a renda familiar foram características com grandes coeficientes associados apontando que as estas estão intimamente relacionadas com o desempenho acadêmico. É importante citar também que importantes variáveis influenciadoras dos resultados do Enem são não observáveis, como a motivação, o esforço e a habilidade dos indivíduos, o que pode levar a vieses nas estimativas. Ao se comparar os modelos aqui estimados com a literatura ressalta-se que os modelos e as variáveis são diferentes em todos os casos, fazendo com que não seja possível uma comparação de valores absolutos e que conclusões ligeiramente diferentes sejam tomadas, mesmo assim é possível notar uma convergência que aponta na direção de que piores condições socioeconômicas estão relacionadas a piores desempenhos educacionais.

Nota-se também que os efeitos não são iguais para todas as provas. Uma diferença notável é a superioridade das mulheres nas notas de Linguagens e Códigos e Redação no Enem 2020, enquanto os homens se saem melhores nas outras. Destaca-se também as provas de Matemática e a de Redação como aquelas que possuem coeficientes com valores mais altos apresentando maiores variações entre os grupos analisados.

Para que as diferenças educacionais sejam superadas políticas afirmativas se fazem necessárias. Ações com efeitos mais rápidos como a Lei de Cotas (Brasil, 2012) contribuem ao facilitar o acesso ao ensino superior e consequente ganho de oportunidades de indivíduos pertencentes a níveis socioeconômicos mais baixos e

pertencentes a grupos raciais historicamente marginalizados. Ações de distribuição de renda possibilitam a garantia do suprimento de necessidades básicas para famílias de baixa renda e viabilizam o investimento em educação. Ações para fortalecimento da infraestrutura de escolas públicas, principalmente estaduais e municipais se fazem necessárias para atenuar as diferenças entre o ensino público e o privado.

De forma geral conclui-se que as diferenças educacionais estão intimamente relacionadas às desigualdades socioeconômicas no Brasil. Devido à interdependência da esfera educacional e a esfera econômica, investimentos para a diminuição das desigualdades em ambas as esferas contribuem para a equidade de condições e oportunidades apontando para um desenvolvimento equilibrado e melhoria de indicadores sociais de forma geral no país.

REFERÊNCIAS

ALBERNAZ, Â.; FERREIRA, F. H. G.; FRANCO, C. **Qualidade e equidade no ensino fundamental brasileiro**. <http://ppe.ipea.gov.br>, 1 dez. 2002.

BARROS, R. P. DE; HENRIQUES, R.; MENDONÇA, R. S. P. DE. **Pelo fim das décadas perdidas: educação e desenvolvimento sustentado no Brasil**. www.ipea.gov.br, 1 jan. 2002.

BOLSA FAMÍLIA TRANSFERÊNCIA DE RENDA E APOIO À FAMÍLIA NO ACESSO À SAÚDE, À EDUCAÇÃO E À ASSISTÊNCIA SOCIAL. [s.l: s.n.]. Disponível em: <https://www.mds.gov.br/webarquivos/publicacao/bolsa_familia/Cartilhas/Cartilha_PB_F_2015.pdf>. Acesso em: 25 nov. 2022.

BRASIL. **Lei nº 12.711, de 29 de agosto de 2012**. Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências. Brasília, DF. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12711.htm. Acesso em: 25 nov. 2022.

CURI, A. Z.; MENEZES-FILHO, N. A. **Mensalidade escolar, background familiar e os resultados do Exame Nacional do Ensino Médio (ENEM)**. <http://ppe.ipea.gov.br>, 1 ago. 2013.

FARRÉ, L.; KLEIN, R.; VELLA, F. **A parametric control function approach to estimating the returns to schooling in the absence of exclusion restrictions: an application to the NLSY**. *Empirical Economics*, v. 44, n. 1, p. 111–133, 13 jun. 2010.

FIGUEIRÊDO, E.; NOGUEIRAY, L.; SANTANAZ, F. L. **Igualdade de Oportunidades: Analisando o Papel das Circunstâncias no Desempenho do ENEM**. *Revista Brasileira de Economia*, v. 68, n. 3, p. 373–392, set. 2014.

FERREIRA, S. G.; VELOSO, F. A. (2003). **Mobilidade intergeracional de educação no Brasil**. *Pesquisa e Planejamento Econômico*, v. 33, n. 3, p. 481–513, dez. 2003

GUTIERREZ, C. E. C.; MOREIRA, R. D. C.; MARQUETTI, A. A. **O efeito do background familiar sobre o desempenho dos estudantes do ensino superior**. *Revista Brasileira de Economia de Empresas*, v. 21, n. 2, 2021.

HENARES, L.; MENEZES FILHO, N.; KOMATSU, B. **Como as Desigualdades entre os Alunos se Refletem nas Notas dos Vários Componentes do ENEM?** 2021. Disponível em: <https://www.insper.edu.br/wp-content/uploads/2021/01/Policy_Paper_Enem_2015_v7.pdf>. Acesso em: 6 jul. 2022.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados do Enem 2020**. Brasília: Inep, 2020. Disponível em: <<https://www.gov.br/inep/ptbr/acesso-a-informacao/dados-abertos/microdados/enem>>. Acesso em: 6 jul. 2022.

JALOTO, A.; PRIMI, R. **Fatores socioeconômicos associados ao desempenho no Enem**. Em Aberto, v. 34, n. 112, 30 dez. 2021.

KLEIN, R.; VELLA, F. **Estimating a class of triangular simultaneous equations models without exclusion restrictions**. Journal of Econometrics, v. 154, n. 2, p. 154–164, 2010.

LIMA, P. DA S. N. et al. **Análise de dados do Enade e Enem: uma revisão sistemática da literatura**. Avaliação: Revista da Avaliação da Educação Superior (Campinas), v. 24, n. 1, p. 89–107, maio 2019.

LOREL, B. **Assessing Brazilian educational inequalities**. Revista Brasileira de Economia, v. 62, n. 1, mar. 2008.

MAHLMEISTER, R. et al. **Revisitando a Mobilidade Intergeracional de Educação no Brasil**. Revista Brasileira de Economia, v. 73, n. 2, 2019.

MELO, L.; ARBIX, G. **Educação, qualificação, produtividade e crescimento econômico: a harmonia colocada em questão**. [s.l.: s.n.]. Disponível em: <<https://www.ipea.gov.br/code2011/chamada2011/pdf/area3/area3-artigo5.pdf>>.

SOLOW, R. M. **Technical Change and the Aggregate Production Function**. The Review of Economics and Statistics, v. 39, n. 3, p. 312, 1957.

VIANA, G.; LIMA, J. F. DE. **Capital humano e crescimento econômico**. Interações (Campo Grande), v. 11, n. 2, p. 137–148, dez. 2010.

VIGGIANO, E.; MATTOS, C. **O desempenho de estudantes no Enem 2010 em diferentes regiões brasileiras**. Revista brasileira de estudos pedagógicos (online). Brasília, v. 94, n. 237, 2013

APÊNDICE 1 - RESULTADO REGRESSÃO COMPLETA MODELO *DUMMIES* MQO

ESTIMATIVAS COMPLETAS PARA MODELO *DUMMIES* MQO

(continua)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
Sexo Feminino	-2,32*** (0,22)	-18,75*** (0,20)	-11,64*** (0,24)	1,13*** (0,18)	-41,24*** (0,28)	58,90*** (0,51)
17 anos	-3,76*** (0,83)	-0,12 (0,76)	-1,62* (0,91)	-3,10*** (0,69)	-8,70*** (1,09)	-5,25*** (1,95)
18 anos	-12,68*** (0,84)	-4,80*** (0,76)	-7,35*** (0,91)	-9,20*** (0,69)	-20,44*** (1,10)	-21,59*** (1,96)
19 anos	-40,90*** (0,92)	-20,17*** (0,84)	-29,42*** (1,00)	-28,85*** (0,76)	-48,51*** (1,21)	-77,57*** (2,16)
20 anos	-54,68*** (1,17)	-24,94*** (1,06)	-39,65*** (1,27)	-39,75*** (0,96)	-58,34*** (1,53)	-110,71*** (2,73)
21 anos	-63,57*** (1,69)	-29,16*** (1,54)	-42,52*** (1,84)	-47,45*** (1,40)	-61,14*** (2,21)	-137,58*** (3,96)
Maior que 21 anos	-70,21*** (1,29)	-25,81*** (1,18)	-42,08*** (1,41)	-50,24*** (1,07)	-63,13*** (1,69)	-169,82*** (3,03)
Preto, pardo ou indígena	-14,47*** (0,25)	-11,20*** (0,23)	-12,73*** (0,27)	-9,03*** (0,20)	-18,76*** (0,32)	-20,66*** (0,58)
Amarelo	-2,81*** (0,75)	1,53** (0,68)	-5,34*** (0,82)	-5,27*** (0,62)	3,68*** (0,98)	-8,67*** (1,76)
Escola Federal	26,92*** (0,47)	19,17*** (0,42)	27,04*** (0,51)	24,46*** (0,39)	37,76*** (0,61)	26,15*** (1,09)
Escola Estadual	-48,93*** (0,29)	-39,21*** (0,27)	-36,97*** (0,32)	-23,34*** (0,24)	-47,99*** (0,39)	-97,12*** (0,69)

(continuação)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
Escola Municipal	-37,16*** (1,23)	-28,98*** (1,12)	-25,54*** (1,34)	-14,58*** (1,02)	-32,27*** (1,62)	-84,46*** (2,89)
Escola em área urbana	17,06*** (0,65)	9,40*** (0,59)	13,45*** (0,71)	15,50*** (0,54)	16,94*** (0,85)	30,02*** (1,52)
Escolaridade do pai: EF completo	3,81*** (0,38)	1,73*** (0,34)	2,40*** (0,41)	4,75*** (0,31)	2,61*** (0,49)	7,56*** (0,88)
Escolaridade do pai: EM completo	8,63*** (0,31)	5,47*** (0,29)	7,19*** (0,34)	8,78*** (0,26)	7,69*** (0,41)	14,01*** (0,73)
Escolaridade do pai: Graduação	20,03*** (0,43)	15,46*** (0,39)	18,17*** (0,46)	15,62*** (0,35)	22,77*** (0,56)	28,10*** (1,00)
Escolaridade do pai: Pós-graduação	21,93*** (0,50)	17,01*** (0,45)	19,77*** (0,54)	15,97*** (0,41)	24,10*** (0,65)	32,79*** (1,16)
Escolaridade da mãe: EF completo	2,80*** (0,42)	0,71* (0,38)	1,40*** (0,45)	4,35*** (0,35)	2,16*** (0,55)	5,36*** (0,98)
Escolaridade da mãe: EM completo	7,84*** (0,35)	4,27*** (0,32)	5,38*** (0,38)	8,37*** (0,29)	7,56*** (0,45)	13,61*** (0,81)
Escolaridade da mãe: Graduação	15,53*** (0,43)	10,50*** (0,39)	12,35*** (0,47)	12,35*** (0,36)	18,57*** (0,57)	23,87*** (1,02)
Escolaridade da mãe: Pós-graduação	15,20***	10,23***	12,27***	11,40***	16,73***	25,36***

(continuação)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
	(0,46)	(0,42)	(0,50)	(0,38)	(0,60)	(1,07)
Faixa de renda familiar: 1 a 3 salários mínimos	14,71***	9,13***	11,73***	12,03***	16,49***	24,16***
	(0,31)	(0,28)	(0,34)	(0,26)	(0,41)	(0,73)
Faixa de renda familiar: 3 a 5 salários mínimos	27,23***	19,26***	22,42***	18,75***	34,72***	41,01***
	(0,44)	(0,40)	(0,48)	(0,37)	(0,58)	(1,04)
Faixa de renda familiar: 5 a 10 salários mínimos	38,71***	28,56***	32,27***	24,53***	51,25***	56,92***
	(0,49)	(0,45)	(0,54)	(0,41)	(0,64)	(1,15)
Faixa de renda familiar: 10 a 20 salários mínimos	54,94***	41,41***	46,96***	33,12***	76,03***	77,19***
	(0,62)	(0,57)	(0,68)	(0,51)	(0,81)	(1,45)
Faixa de renda familiar: mais de 20 salários mínimos	72,54***	54,09***	63,80***	43,08***	106,69***	95,02***
	(0,83)	(0,76)	(0,90)	(0,69)	(1,09)	(1,95)
Possui computador em casa	18,67***	14,14***	17,04***	12,18***	18,19***	31,81***
	(0,28)	(0,26)	(0,31)	(0,24)	(0,37)	(0,67)
Tem acesso à internet	9,57***	3,64***	6,68***	9,72***	5,60***	22,22***
	(0,38)	(0,35)	(0,42)	(0,32)	(0,50)	(0,90)
AC	-12,51***	-14,30***	-18,67***	-18,29***	-38,24***	26,94***
	(1,66)	(1,51)	(1,81)	(1,37)	(2,17)	(3,89)

(continuação)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
AL	-7,53*** (0,89)	-15,44*** (0,81)	-19,90*** (0,97)	-18,07*** (0,73)	-23,48*** (1,16)	39,21*** (2,08)
AM	-13,71*** (0,83)	-15,42*** (0,75)	-10,62*** (0,90)	-26,82*** (0,68)	-24,87*** (1,08)	9,17*** (1,93)
AP	-27,79*** (1,72)	-25,62*** (1,57)	-31,63*** (1,87)	-31,93*** (1,43)	-55,23*** (2,26)	5,46 (4,03)
BA	4,63*** (0,53)	-1,29*** (0,48)	-3,87*** (0,57)	-7,52*** (0,44)	-10,23*** (0,69)	46,05*** (1,23)
CE	11,64*** (0,47)	0,79* (0,43)	-1,72*** (0,51)	-8,54*** (0,39)	3,58*** (0,62)	64,07*** (1,11)
DF	-2,98*** (0,71)	-3,31*** (0,65)	-3,60*** (0,77)	-0,79 (0,59)	-15,81*** (0,93)	8,61*** (1,66)
ES	13,21*** (0,76)	7,93*** (0,69)	5,92*** (0,82)	-4,02*** (0,63)	10,01*** (0,99)	46,19*** (1,77)
GO	4,56*** (0,61)	-0,47 (0,56)	-4,53*** (0,67)	-7,53*** (0,51)	-9,16*** (0,80)	44,49*** (1,44)
MA	-12,71*** (0,69)	-15,67*** (0,63)	-20,87*** (0,75)	-24,71*** (0,57)	-32,52*** (0,90)	30,20*** (1,61)
MG	24,23*** (0,44)	13,37*** (0,40)	15,02*** (0,48)	5,99*** (0,36)	21,87*** (0,58)	64,87*** (1,03)
MS	-9,67*** (0,97)	-7,68*** (0,88)	-13,42*** (1,05)	-13,35*** (0,80)	-21,13*** (1,27)	7,22*** (2,27)
MT	-17,49*** (0,88)	-16,83*** (0,80)	-20,78*** (0,95)	-21,27*** (0,73)	-31,94*** (1,15)	3,39* (2,06)
PA	-2,90*** (0,62)	-9,95*** (0,56)	-13,45*** (0,67)	-17,70*** (0,51)	-27,39*** (0,81)	54,00*** (1,45)

(continuação)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
PB	6,43*** (0,73)	-6,17*** (0,66)	-7,00*** (0,79)	-12,07*** (0,60)	-12,54*** (0,95)	69,91*** (1,70)
PE	5,42*** (0,55)	-1,86*** (0,50)	-6,73*** (0,60)	-6,85*** (0,45)	-3,57*** (0,72)	46,12*** (1,28)
PI	6,46*** (0,81)	-4,52*** (0,74)	-8,84*** (0,88)	-14,79*** (0,67)	-9,30*** (1,06)	69,76*** (1,90)
PR	-6,36*** (0,54)	-1,91*** (0,49)	-0,98* (0,58)	-6,91*** (0,44)	-6,85*** (0,70)	-15,15*** (1,25)
RJ	6,97*** (0,47)	-2,59*** (0,43)	2,10*** (0,51)	1,39*** (0,39)	-3,80*** (0,61)	37,74*** (1,09)
RN	10,53*** (0,81)	-0,67 (0,73)	-2,29*** (0,88)	-4,07*** (0,67)	-3,14*** (1,06)	62,82*** (1,89)
RO	-6,35*** (1,24)	-7,24*** (1,13)	-14,20*** (1,35)	-17,34*** (1,03)	-20,53*** (1,62)	27,56*** (2,90)
RR	-26,69*** (2,21)	-20,81*** (2,01)	-24,70*** (2,40)	-23,35*** (1,83)	-41,08*** (2,89)	-23,54*** (5,17)
RS	7,17*** (0,57)	-2,14*** (0,52)	5,94*** (0,62)	3,47*** (0,47)	0,92 (0,75)	27,69*** (1,34)
SC	-2,98*** (0,67)	-1,37** (0,61)	0,39 (0,73)	-6,67*** (0,56)	-3,95*** (0,88)	-3,32** (1,57)
SE	13,93*** (1,01)	2,93*** (0,92)	-1,62 (1,09)	-8,38*** (0,83)	-7,97*** (1,32)	84,66*** (2,36)
TO	-14,67*** (1,26)	-14,44*** (1,15)	-20,41*** (1,37)	-23,91*** (1,05)	-28,42*** (1,65)	13,82*** (2,96)
Constante	516,56*** (1,20)	500,23*** (1,09)	510,23*** (1,30)	507,77*** (0,99)	549,82*** (1,57)	514,74*** (2,80)

(conclusão)

	Variável dependente: Nota					
	Média Geral	Ciências da Natureza	Ciências Humanas	Linguagens e Códigos	Matemática	Redação
	(1)	(2)	(3)	(4)	(5)	(6)
Observações	480.397	480.397	480.397	480.397	480.397	480.397
R ²	0,40	0,33	0,30	0,30	0,37	0,26
R ² Ajustado	0,40	0,33	0,30	0,30	0,37	0,26
Erro Pad. Residual (df = 480342)	73,79	67,16	80,22	61,04	96,56	172,66
Estatística F (df = 54; 480342)	6.011,00***	4.368,07***	3.739,12***	3.847,49***	5.264,84***	3.154,44***

Fonte: Elaboração própria com base nos Microdados do Enem 2020

Nota: *p<0,1; **p<0,05; ***p<0,01; Erros padrão entre parênteses