

Os impactos de melhoras na infraestrutura sobre a qualidade da educação básica municipal

César Freitas

Gabriel Lucas

João Pedro Dias

Júlia Michalski

Vinícius Nery

Maio de 2022

Resumo

Esse artigo tem por objetivo determinar, usando métodos de painel, os impactos de melhoras na infraestrutura de colégios da rede municipal sobre o desempenho acadêmico dos estudantes duas etapas do Ensino Fundamental, sendo, para isso, utilizado o Índice de Desenvolvimento da Educação Básica (IDEB), dados do Censo Escolar e outras características socioeconômicas municipais e escolares.

Os resultados revelam a importância do fornecimento de Internet para o desempenho acadêmico dos estudantes, o que é mais pronunciado no Nordeste. Grande parte das demais variáveis de infraestrutura se mostra estatisticamente insignificante após a adição de controles e a remoção de efeitos não-observáveis. Por fim, a análise regional reitera as desigualdades que permeiam o território brasileiro, bem como a ascensão relativa da qualidade da educação do Nordeste na última década.

Palavras-Chave: Educação, IDEB, Infraestrutura, Internet, Painel

Sumário

1	Introdução	2
2	O IDEB	3
3	Dados	4
4	Análises Descritivas	6
4.1	Infraestrutura	7
4.2	Região	13
4.3	Conclusão	15
5	Modelos Econométricos	16
5.1	Breve Teoria de Modelos de Painel	16
5.2	Resultados	17
6	Conclusão	24
7	Apêndice	25

1 Introdução

Até meados do século XX, a educação básica no Brasil foi um privilégio restrito às camadas sociais economicamente mais beneficiadas, o que colocava o país entre os territórios com os piores indicadores tanto em termos de nível quanto de desigualdade escolar (KOMATSU et al., 2019) (LÓPEZ-CALVA; LUSTIG, 2010). A partir de 1970, porém, houve um grande salto quantitativo, elevando a dispersão e a média educacional do país (LEEUVEN; LEEUVEN-LI, 2015), o que causou, por exemplo, um ganho de produtividade e uma redução da desigualdade de renda (BARROS et al., 2010).

Nesse sentido, a figura abaixo mostra a trajetória temporal da escolaridade média do Brasil e sua grande evolução ao longo das últimas décadas. Os dados são de Leeuwen e Leeuwen-Li (2015) e as linhas cinzas representam os diversos outros países da amostra:

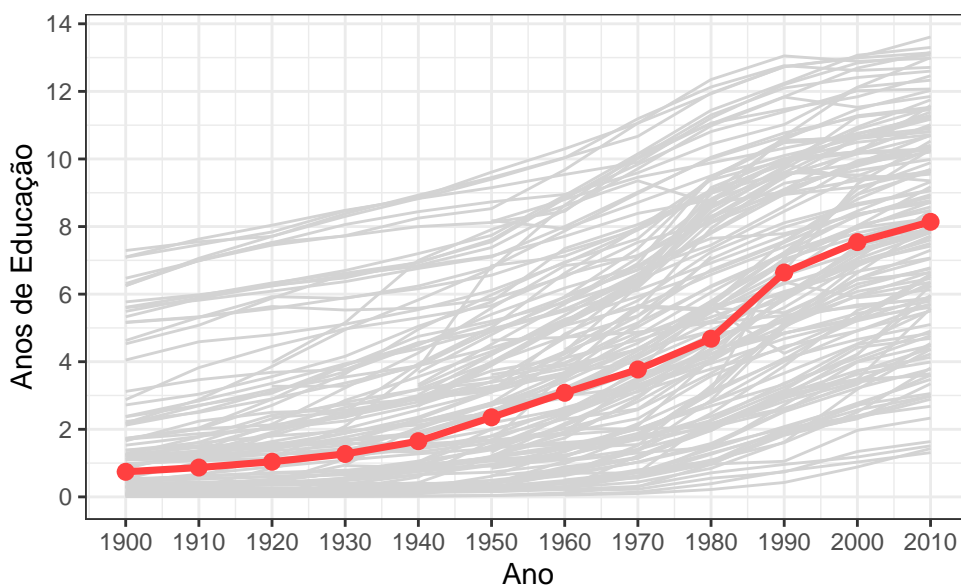


Figura 1: Escolaridade Média de Diversos Países entre 1900 e 2010

O século XXI trouxe consigo o desafio de aliar esse ganho quantitativo a melhoras qualitativas na educação, especialmente nas etapas mais básicas. Assim, indicadores sintéticos de performance escolar tornaram-se fundamentais para avaliar a qualidade do ensino brasileiro e melhor direcionar as políticas públicas nesta área, além de permitirem comparações mais acuradas entre diferentes períodos e localidades (MAIA; BUENO; SATO, 2021).

Nesse sentido, criou-se, em 2007, o Índice de Desenvolvimento da Educação Básica (IDEB). Ele reúne em um indicador duas medidas de qualidade da educação: o fluxo escolar – medido pela taxa de aprovação dos alunos – e as médias de desempenho em avaliações nacionais – que são uma *proxy* para o grau de aprendizagem. No ano de sua concepção, a meta era alcançar uma média nacional de 6 pontos (em uma escala que vai de 0 a 10) no IDEB até 2021, o que, apesar de pequenos avanços no passado recente, parece um sonho distante (OCDE, 2021).

Vários fatores estão correlacionados com a performance acadêmica e contribuem para o desempenho subótimo do país, como os altos níveis de pobreza e desigualdade (DUARTE, 2013); os gastos em educação relativamente pequenos em relação ao PIB (GRISA, 2019); e a baixa qualidade do valor investido por aluno, especialmente em casos como o do Brasil em que tal despesa já não é tão baixa (DUQUE, 2019).

Ressalta-se, porém, que não há consenso no que tange ao impacto desses fatores, sendo que há inclusive uma heterogeneidade no que tange às variáveis mais importantes para a predição do IDEB a depender do etapa educacional (Ensino Fundamental I, Fundamental II ou Médio) (MAIA; BUENO; SATO, 2021).

Faltam, porém, estudos acerca do papel de características de infraestrutura e de tecnologia dos colégios sobre o desempenho medido pelo IDEB que, apesar de correlacionadas com o nível socioeconômico da escola (BECSKEHÁZY, 2021), ainda são variáveis fundamentais de serem analisadas e importantes para o aprendizado. Além disso, Grisa (2019) mostra que, mesmo em anos recentes, a boa infraestrutura escolar está longe de ser universalizada no Brasil, especialmente no que tange às redes municipais e estaduais, as quais concentram a maior parte das matrículas do ensino básico brasileiro. Ademais, o autor também mostra que, em 2015, 69% das redes de ensino investiam, por aluno, um valor inferior à média nacional, o que se traduz em grandes deficiências estruturais.

Nesse sentido, o presente trabalho tem por objetivo analisar justamente o efeito das variáveis de infraestrutura sobre o desempenho acadêmico e a aprendizagem da rede mais numerosa e deficitária do país: a municipal. Em especial, o foco da pesquisa será o desempenho nas duas etapas da educação básica, ou seja, o IDEB do Ensino Fundamental I (“Anos Iniciais”) e II (“Anos Finais”). Para isso, foram extraídos dados de diversas fontes educacionais e socioeconômicas entre 2009 e 2019 ao nível de escola e de município, visando aferir, por meio de análises descritivas, regressões e modelos de painel, o impacto de melhoras na infraestrutura escolar sobre a performance educacional dos estudantes medida/estimada pelo IDEB.

2 O IDEB

Criado em 2007 para estimar a situação educacional do país, o IDEB é calculado bianualmente, haja vista a periodicidade das provas que o compõem. Além disso, o índice faz parte de uma das principais metas do Plano de Desenvolvimento da Educação (PDE), sendo que o objetivo nacional era alcançar uma média de 6 pontos – o que seria compatível com o patamar de países considerados desenvolvidos – no IDEB 2021, a ser divulgado em 2022, ano do bicentenário da Independência (FIGUEIREDO et al., 2018).

Como brevemente comentado na **Introdução**, o índice é construído a partir da multiplicação de duas medidas:

- O *indicador de rendimento* (R), construído com base na taxa de aprovação dos alunos da escola. Dessa forma, essa métrica mede o fluxo dos estudantes a partir de uma avaliação *interna*, ou seja, feita pela própria instituição;
- A *média do desempenho em provas* (A) de matemática e português¹. Assim, esse índice busca aproximar o desempenho acadêmico a partir de uma avaliação *externa*, ou seja, montada pelo Inep e aplicada em larga escala.

Simplificadamente, o IDEB pode ser calculado pela seguinte expressão²:

$$\text{IDEB} = R \cdot A \quad (1)$$

A construção do índice a partir da combinação entre “fluxo” e “aprendizagem” tem por objetivo equilibrar a avaliação do indicador, desincentivando, por exemplo, os gestores a reprovarem seus estudantes para que

¹A nível nacional e estadual, é aplicada a prova do Sistema de Avaliação da Educação Básica (Saeb). Para os municípios, a métrica é construída com base nos resultados da Prova Brasil.

²Mais detalhes podem ser vistos na nota técnica de concepção do IDEB, disponível [aqui](#).

alcançam resultados melhores nas provas posteriormente. Analogamente, uma alta taxa de aprovação sem aprendizado levaria a resultados piores nas avaliações externas, o que também seria captado pela métrica (INEP, 2022).

Contudo, o índice possui também muitas críticas, as quais foram compiladas e estendidas por [Figueiredo et al. \(2018\)](#). No texto, os autores pontuam que, uma vez que o índice só é anonimizado para os alunos e a abordagem da divulgação de seus resultados – que, em termos de transparência, é considerada um ponto positivo – está centrada no ranqueamento das escolas, as instituições possuem incentivos a burlarem alguns dos pressupostos da avaliação. Dessa forma, colégios preparam ativamente seus estudantes com questões de provas antigas e simulados, bem como deixam de aplicar a avaliação para aqueles considerados menos bem preparados.

A não-aplicação da prova pode ser combinada com uma política de alta taxa interna de aprovação, o que eleva artificialmente o índice da escola e pode representar um erro de medida na variável dependente dos modelos usados. Ademais, os autores também encontram que, em alguns estados, não há correspondência entre as notas do IDEB nas diferentes etapas do Ensino Fundamental, bem como entre a métrica do Ensino Médio e as notas do ENEM.

Por fim, por mais que uma avaliação externa de larga escala traga consigo benefícios relacionados à possibilidade de comparação entre períodos e regiões, a prova desconsidera questões qualitativas, socioeconômicas e conjunturais que podem vir a afetar o desempenho dos estudantes. É precisamente esse aspecto que o presente trabalho busca analisar, visando, a partir dos dados descritos na seção a seguir, estimar o impacto de melhoras na estrutura escolar sobre o desempenho acadêmico dos alunos.

3 Dados

Como dito anteriormente, este trabalho tem como objetivo principal analisar o desempenho das escolas municipais nas duas etapas que compõe a educação básica, o qual pode ser medido através do IDEB e dos índices que o compõe. Contudo, usou-se também uma série de outras características escolares e municipais a fim de se determinar os impactos *ceteris paribus* das variáveis de infraestrutura sobre a performance acadêmica das instituições de ensino. Apesar de serem de diferentes fontes, todos os dados foram extraídas do *datalake* da Base dos Dados ([BASE DOS DADOS, 2022](#)) usando o *Google Big Query*³.

No que tange ao [IDEB](#), foram usados os dados disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) a nível escolar e municipal. Assim, foram excluídas as escolas municipais brasileiras que não possuíam IDEB registrado, o que pode ocorrer por um número insuficiente de alunos realizando a Prova Brasil ou por falta de prestação de informação por parte da escola ([INEP, 2012](#)). Como há pouquíssimos colégios municipais de Ensino Médio e os microdados escolares do IDEB a cerca dessa etapa só passaram a ser divulgados em 2017, a análise será centrada no Ensino Fundamental.

Para as variáveis de infraestrutura, tecnologia, número de matrículas e quantidade professores, usou-se o [Censo Escolar](#), também disponibilizado pelo Inep. As características de infraestrutura e tecnologia foram obtidas ao nível do colégio; o número de alunos e docentes, por sua vez, foi obtido agrupando as respectivas tabelas por `id_escola` e `ano`. Por fim, usou-se os dados dos [Indicadores Educacionais](#) do Inep para obter a média de alunos por turma, a quantidade de horas-aula e o percentual de professores com ensino superior em cada uma das etapas de ensino (Fundamental I e II) em cada escola.

Além disso, o [Tesouro Nacional \(2019\)](#) chama a atenção para a correlação entre o desempenho no IDEB e

³As *queries* usadas estão disponíveis no arquivo `.Dados/Queries_LabEconometria.sql`.

as despesas municipais em educação por aluno; mais especificamente, uma *proxy* para o valor gasto por escola. Visando captar essa relação, foram coletados também dados do Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro (SICONFI) a fim de se obter o valor das despesas com educação⁴ de cada município.

Assim, sendo $desp_{educ}^{mun}$ a despesa do município com educação, $alunos^{mun}$ o número de matrículas do município – obtido por meio da agregação do número de matrículas do Censo Escolar ao nível municipal – e o superescrito ^{esc} indicando suas contrapartidas ao nível escolar, a *proxy* para o valor gasto por cada escola usada pelo Tesouro Nacional pode ser obtido pela seguinte expressão:

$$desp_{educ}^{esc} = \frac{alunos^{esc}}{alunos^{mun}} \cdot desp_{educ}^{mun} \quad (2)$$

Também foram utilizados os dados disponibilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) para a obtenção de informações acerca do PIB e da população municipal. Assim, foi possível criar variáveis de despesas com educação frente ao PIB do município – o que indica a “disposição” da cidade a investir em educação – e também em termos per capita.

Ademais, as variáveis de gastos municipais com educação e PIB foram todas trazidas para Reais do início de 2019, a fim de permitir melhores comparações nas análises descritivas. Para a primeira, foi construída uma tabela de índice de preços com base nos valores de produção do produto “Educação Pública” nas Tabelas de Recursos e Usos (TRUs) retropoladas. Já a segunda, por sua vez, foi deflacionada usando o deflator implícito do PIB nacional.

Na base de dados também foram incluídos os dados acerca das transferências do Governo Federal aos municípios pelo Fundo Nacional da Educação Básica (FUNDEB). O uso dessa variável tem por objetivo avaliar e controlar para a razão entre tais transferências e os gastos totais com educação de cada município, visando analisar a dependência do orçamento educacional das cidades frente aos recursos federais.

Por fim, usou-se a Relação Anual de Informações Sociais (RAIS) para se obter o número de trabalhadores formais e o salário médio (em reais correntes e em relação ao salário mínimo do respectivo ano) de cada município. Além disso, coletou-se também esses dados discriminados para os professores de cada etapa educacional relevante: Fundamental I (família 2312) e II (2313), fazendo-se a razão desses rendimentos frente ao salário médio do município. Ressalta-se, porém, que essas informações detalhadas da RAIS são esparsas nos anos iniciais da base e podem apresentar inconsistências com, por exemplo, o número de professores registrado no Censo Escolar.

Com a junção de todas as bases – o que foi feito a partir dos dados do IDEB e por meio de *inner joins*⁵ a fim de maximizar a quantidade de informações –, obteve-se uma tabela com 271.621 observações e 99 variáveis distribuídas bianualmente entre 2009 e 2019, uma vez que as avaliações acontecem de dois em dois anos.

Contudo, verificou-se alguns valores extremos no que tange às despesas com educação em alguns municípios, o que afeta também o cálculo da *proxy* proposta para os colégios. Dessa forma, considerou-se como *outliers* as observações com um gasto em educação per capita ou com uma *proxy* para a despesa das escolas que estavam nos 0,1% do topo ou da cauda da respectiva distribuição, o que resultou em uma eliminação de pouco menos de 1.000 observações do universo inicial, chegando a 270.660 colégios.

A distribuição das informações por ano e etapa escolar pode ser vista na Tabela 1 abaixo, na qual se pode ver a consistência numérica das informações do Ensino Fundamental ao longo do tempo. Por fim, ressalta-se que todas as escolas com IDEB registrado possuem informações acerca de sua estrutura e tecnologia no Censo Escolar desde 2009, o que será útil ao presente trabalho.

⁴Para o cálculo, considerou-se as Despesas Empenhadas e ambas as categorias de Inscritões de Resto a Pagar.

⁵As variáveis usadas para os *joins* foram as de *ano*, *id_escola*, *id_municipio* e *uf*, a depender das bases a serem fundidas.

Tabela 1: Distribuição das Observações por Ano e Etapa Escolar

Ano	Fundamental I	Fundamental II	Total
2009	29494	12458	41952
2011	29084	13243	42327
2013	29323	13932	43255
2015	29885	13906	43791
2017	33961	14290	48251
2019	34879	16205	51084
Total	186626	84034	270660

4 Análises Descritivas

Feita a introdução sobre os dados, é importante a realização de uma análise descritiva a fim de ver as principais relações entre o IDEB e as características de infraestrutura, além de eventuais diferenças regionais no desempenho acadêmico. Assim, serão analisados diversos gráficos acerca da evolução temporal do IDEB, bem como de sua distribuição condicional em outras variáveis.

A análise de infraestrutura irá focar nos efeitos do provimento de energia, água, internet, esgoto sobre a performance acadêmica dos estudantes das duas etapas do Ensino Fundamental, além de outras questões estruturais. Além disso, o tópico regional avaliará os índices de desempenho escolar dos alunos para cada uma das cinco regiões do Brasil, com o intuito de captar diferenças sistemáticas entre as regiões, as quais parecem ser persistentes no tempo (KOMATSU et al., 2019).

Antes de começar as análises de infraestrutura, é importante observar como se deu a evolução do IDEB e de seus componentes ao longo do tempo, separando a visualização por etapa escolar a fim de se verificar eventuais diferenças na evolução e no nível de aprendizado de estudantes de diferentes idades. Nele, vê-se que a evolução de ambas as etapas é similar, mas o IDEB do Ensino Fundamental I é superior àquele visto nos Anos Finais. Além disso, vê-se que o *gap* entre Anos Iniciais e Finais é maior no que tange à aprendizagem – aproximada pelo SAEB –, o que é compensado por uma menor diferença no Índice de Rendimento. De fato, parece haver um *trade-off* entre proficiência e aprovação (ZOGHBI et al., 2011).

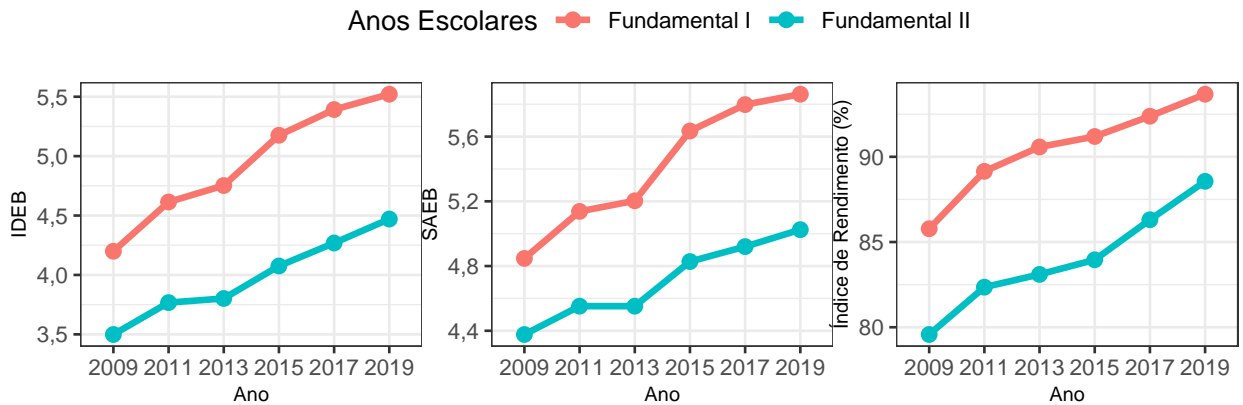


Figura 2: Evolução do IDEB e dos seus Componentes por Etapa Escolar

4.1 Infraestrutura

Felizmente, as características de infraestrutura têm caminhado rumo à universalização ao longo do tempo, mesmo entre as escolas municipais, as quais, geralmente, costumam receber menos recursos. A porcentagem de colégios que possuem as principais características de infraestrutura pode ser vista na tabela abaixo⁶, na qual se vê a ubiquidade do acesso à água, energia e esgoto, bem como a crescente disponibilidade de internet⁷:

Tabela 2: Médias de Infraestrutura por Ano

Ano	Água	Água RP	Energia	Esgoto	Internet	Lab. Informática	Biblioteca
2009	99,9%	80,3%	99,6%	99,4%	55,1%	46,6%	46,8%
2011	98,7%	82,4%	99,8%	99,4%	74,0%	72,5%	49,5%
2013	98,1%	82,4%	99,8%	99,2%	78,6%	78,5%	50,3%
2015	98,3%	83,0%	99,9%	99,2%	83,8%	75,4%	51,7%
2017	98,3%	77,0%	99,8%	98,8%	80,4%	61,1%	46,2%
2019	98,8%	77,8%	99,7%	97,5%	87,5%	48,1%	42,6%

4.1.1 Energia

No que diz respeito aos efeitos do provimento de energia para as escolas sobre o IDEB, é possível inferir dos boxplots e das densidades condicionais abaixo que os estudantes de instituições que desfrutam de energia elétrica tendem a ter um desempenho superior àqueles que frequentam colégios sem energia. Tal conclusão é baseada no fato de que a distribuição das escolas com energia (curva em azul) está à direita daquela dos colégios que não tem eletricidade (em vermelho). No gráfico de boxplot, por sua vez, vê-se que a distribuição das escolas com energia é superior àquela de suas contrapartidas, além de possuir uma maior variância.

Por fim, o gráfico de linhas mostra que há um grande hiato nas médias do IDEB entre as escolas com e sem energia. Mais ainda, vê-se que essa diferença vem aumentando ao longo do tempo – com os IDEBs de colégios com eletricidade crescendo e os de sem energia com um histórico volátil – e é maior no Ensino Fundamental I, o que indica que crianças em idades iniciais de alfabetização sofrem mais com a carência de recursos.

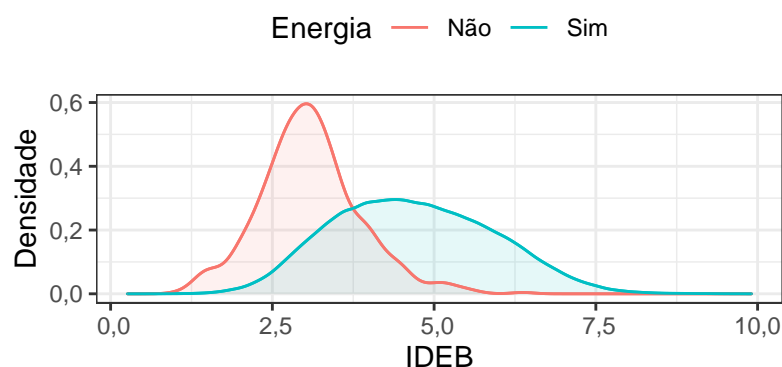


Figura 3: Densidade do IDEB por Energia

⁶Na tabela abaixo e ao longo do documento, “RP” significa que a infraestrutura é ligada à rede pública de energia/água.

⁷Algumas das variações contraintuitivas na tabela podem ser fruto de (1) mudança amostral (2) erros de declaração, uma vez que o Censo Escolar é autodeclarado.

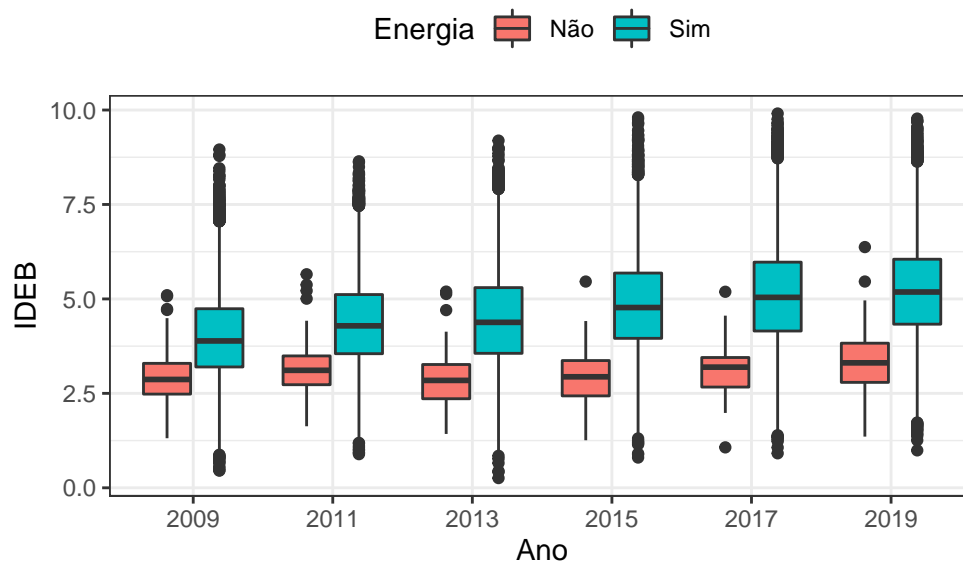


Figura 4: Boxplot do IDEB por Ano e Fornecimento de Energia

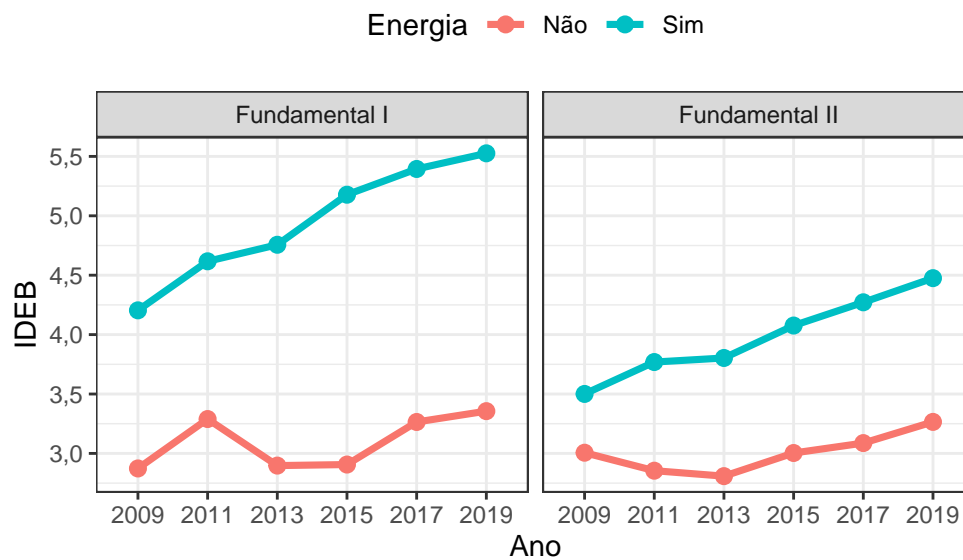


Figura 5: Evolução do IDEB por Ano e Fornecimento de Energia

4.1.2 Água e Água Fornecida pela Rede Pública

Em relação aos efeitos do provimento de água, é possível inferir do gráfico de violino abaixo – uma variação de um boxplot que mostra a densidade relativa dos dados – que a concentração de escolas com água se situa em valores de IDEB superiores àqueles observados para a concentração dos colégios sem água. A espessura dos violinos mostra também que as notas de instituições sem fornecimento de água são mais concentradas na cauda da distribuição, ao passo que a dispersão de colégios com água é mais uniforme.

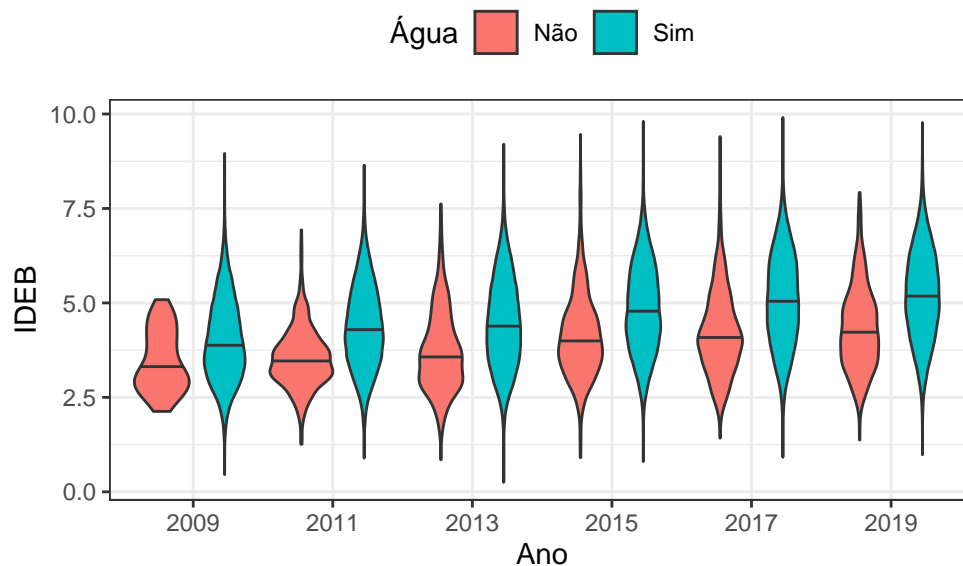


Figura 6: Distribuição do IDEB por Fornecimento de Água

A importância do provimento de água para as escolas é reforçada quando se analisa o provimento público de água, uma vez que alguns colégios podem ser abastecidos por caminhões pipa ou por poços artesanais, por exemplo. Segundo o gráfico apresentado a seguir, a média do IDEB das escolas com fornecimento público de água é maior em ambos os anos e tanto no Ensino Fundamental I quanto no II, sendo que a diferença é maior naquela etapa frente a esta – mesma conclusão que foi encontrada na análise de energia. Ao contrário da subseção anterior, porém, esse hiato parece estar relativamente constante ao longo do tempo.

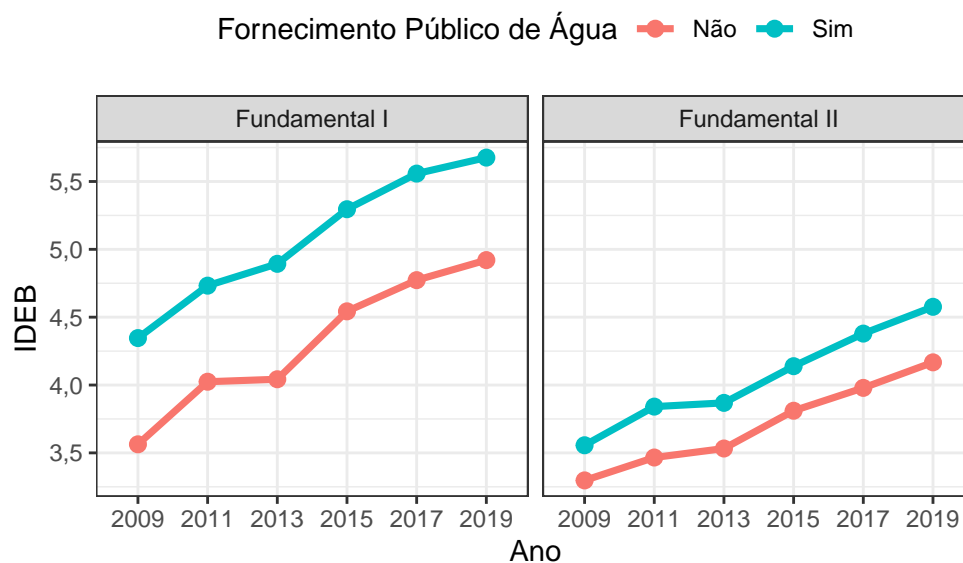


Figura 7: Evolução do IDEB por Ano e Fornecimento Público de Água

4.1.3 Esgoto

O sistema de esgoto é outro componente estrutural que também influi sobre o desempenho escolar dos estudantes, pois sua presença indica a qualidade do saneamento de uma escola. Tal fato é corroborado não só pelas informações providas pelo gráfico a seguir, mas também por seus efeitos positivos sobre o desempenho dos estudantes, o que já foi documentado por vários estudos a respeito do tema (SCRIPTORE; AZZONI; MENEZES-FILHO, 2018). Novamente, a diferença parece ser maior nos anos iniciais da alfabetização, sendo também relativamente constante ao longo do tempo:

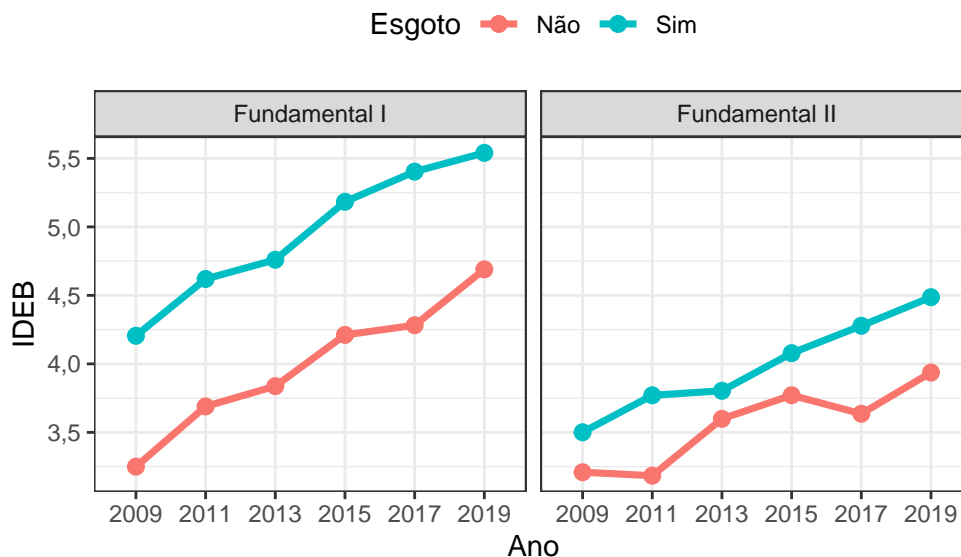


Figura 8: Evolução do IDEB por Ano e Fornecimento Público de Água

4.1.4 Internet

Como visto na tabela inicial dessa seção, uma parcela crescente das escolas municipais têm disponibilizado internet aos seus alunos, visando acompanhar a transformação digital que, inevitavelmente, afeta os métodos de aprendizado dos estudantes. Dessa forma, espera-se que a diferença de rendimento entre os estudantes com e sem acesso aos meios digitais cresça ao longo do tempo.

De fato, o gráfico de linhas abaixo mostra que isso é verdade em ambas as etapas escolares, sendo que o hiato aumentou cerca de 50% tanto no Ensino Fundamental I quanto no II. Novamente, tanto o nível quanto a diferença são maiores nos Anos Iniciais, o que corrobora a hipótese já desenvolvida nas subseções anteriores da maior importância de infraestrutura para a boa alfabetização. Tal fato é justificado pela importância atribuída a presença da internet para os estudantes, sobretudo como instrumento de pesquisa (OLIVEIRA; CASAGRANDE; GALERANI, 2016) tendo em vista a facilidade e praticidade para obter informações nos meios virtuais.

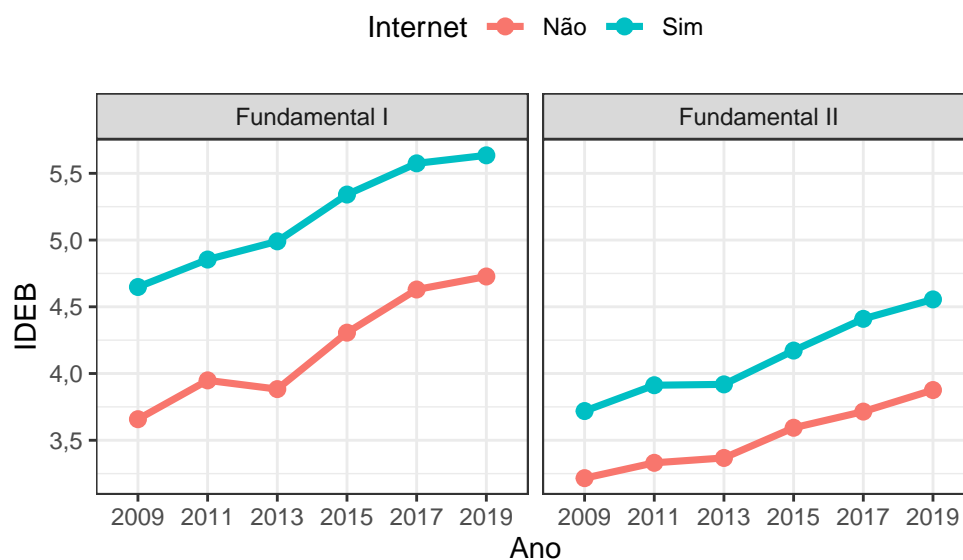


Figura 9: Evolução do IDEB por Ano e Acesso à Internet

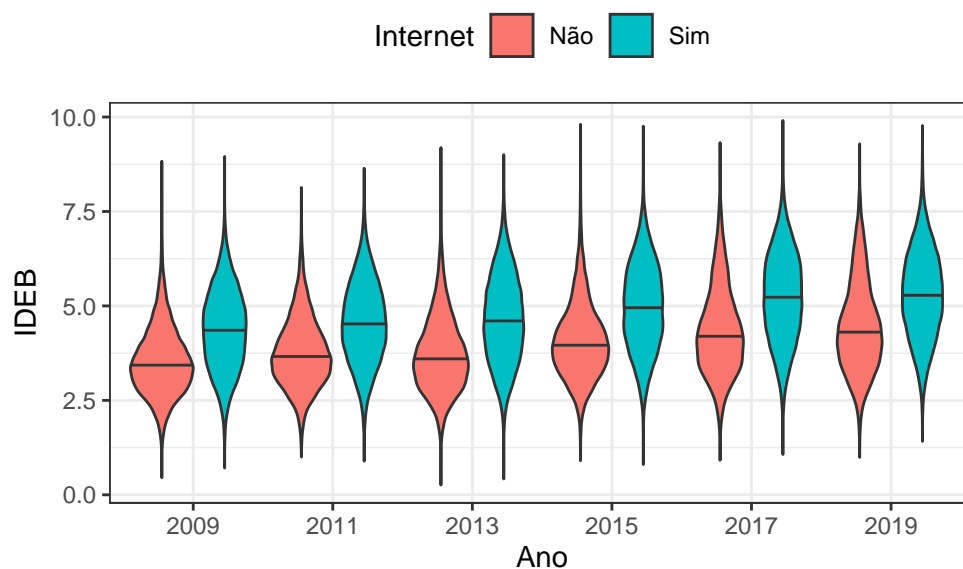


Figura 10: Distribuição do IDEB por Acesso à Internet

4.1.5 Quadras de Esporte

Outro fator estrutural que contribui para o bom do desempenho dos estudantes é a presença de quadras de esporte, uma vez que a prática de atividades desportivas contribui positivamente para o rendimento dos alunos (PESERICO; KRAVCHYCHYN; OLIVEIRA, 2015). A análise do gráfico corrobora o fato acima, uma vez que colégios com quadras possuem IDEBs sistematicamente maiores que suas contrapartidas; ressalta-se, contudo, um problema na declaração dessa variável em 2011 e 2013, anos nos quais não há nenhuma instituição com quadra reportada.

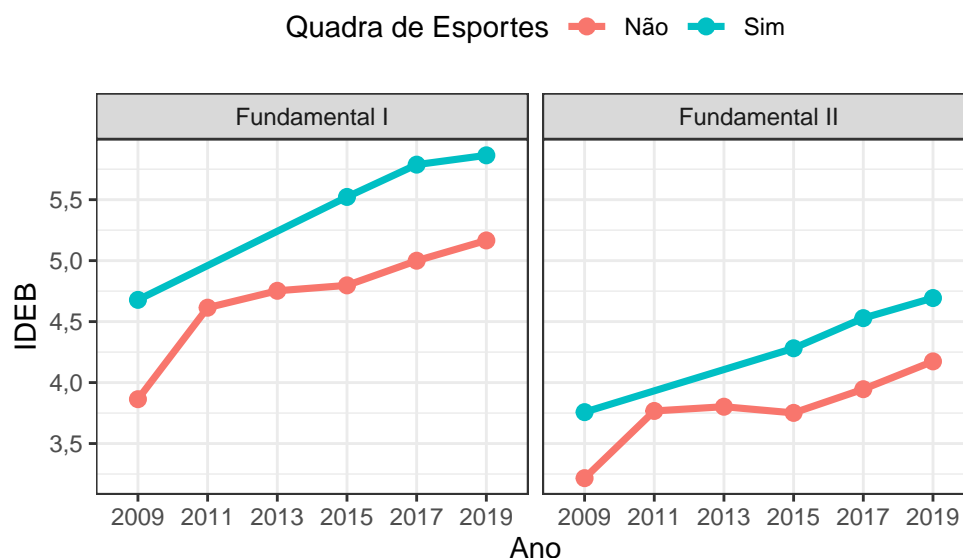


Figura 11: Evolução do IDEB por Ano e Existência de Quadras Esportivas

4.1.6 Laboratório de Informática

No que diz respeito ao laboratório de informática, pode-se dizer que a sua presença em instituições de ensino também é importante para o rendimento escolar dos estudantes, pois é uma estrutura que complementa as benesses providas pela internet. Tais fatos são corroborados pelas informações do gráfico a seguir, o qual aponta para uma correlação positiva entre a presença de laboratórios de informática e o desempenho da escola:

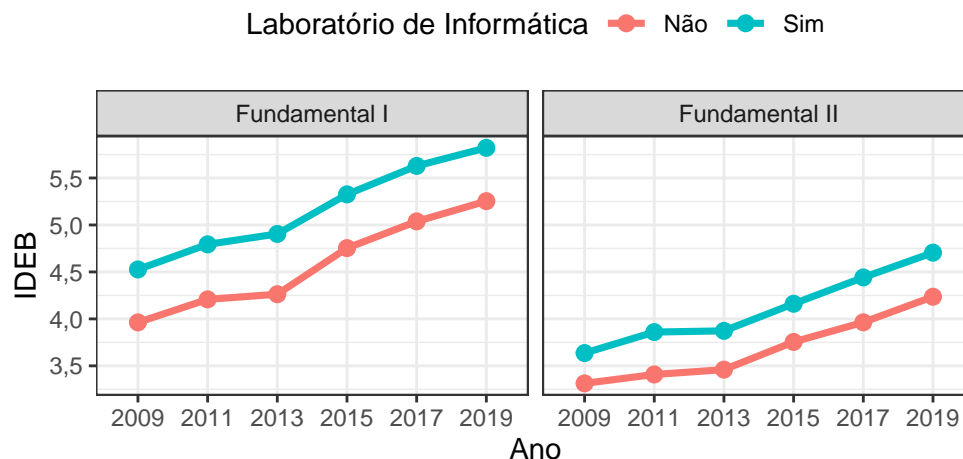


Figura 12: Evolução do IDEB por Ano e Acesso a Laboratório de Informática

4.1.7 Biblioteca

Por último, vale destacar a importância das bibliotecas no ambiente escolar como um componente estrutural que visa o incentivo ao estudo e à leitura. No gráfico a seguir, como era de esperar, nota-se a relação positiva entre a presença de bibliotecas e os valores de IDEB das instituições que contam com essa estrutura específica.

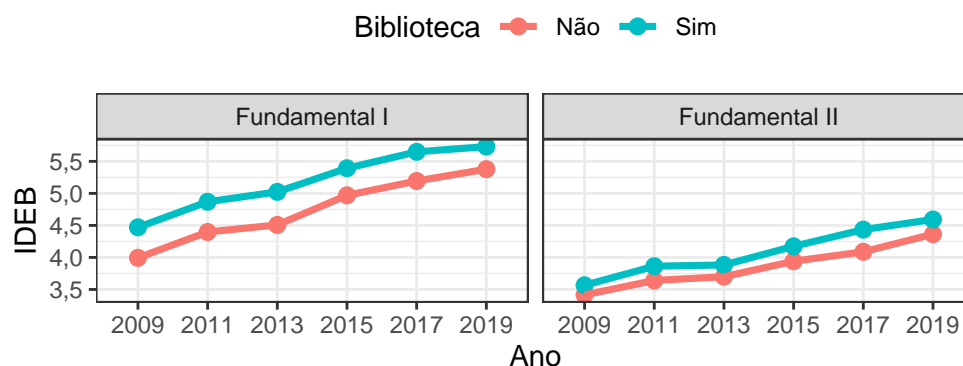


Figura 13: Evolução do IDEB por Ano e Acesso a Biblioteca

4.2 Região

Dada a heterogeneidade e a desigualdade regional brasileira, é interessante analisar o comportamento regional das variáveis de infraestrutura, bem como do desempenho acadêmico estimado pelo IDEB. Nesse sentido, vê-se que tanto as características estruturais como o grau de aprendizagem seguem o padrão de muitas outras estatísticas, sendo melhores no Sul/Sudeste e menores no Norte/Nordeste.

O gráfico abaixo mostra a evolução do IDEB por região e ano, evidenciando a desigualdade comentada acima e a melhora ubíqua do indicador ao longo do tempo. Contudo, vê-se uma lenta ascensão relativa da região Nordeste, a qual deixou o posto de região com menor média em 2015; nesse sentido, ressalta-se, por exemplo, os esforços e ganhos promovidos pelo Ceará na educação básica.

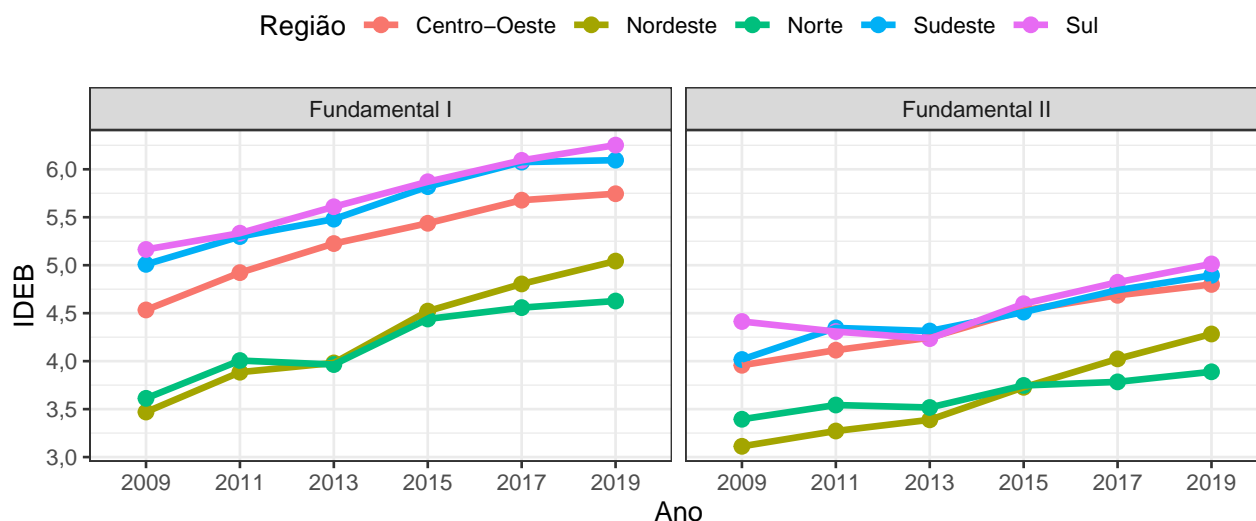


Figura 14: Evolução do IDEB por Região

Além disso, o boxplot abaixo mostra a distribuição das notas do IDEB, na qual fica ainda mais evidente o comportamento do Nordeste: apesar de possuir medianas inferiores às do Sul, Sudeste e Centro-Oeste, a região possui as maiores notas no IDEB:

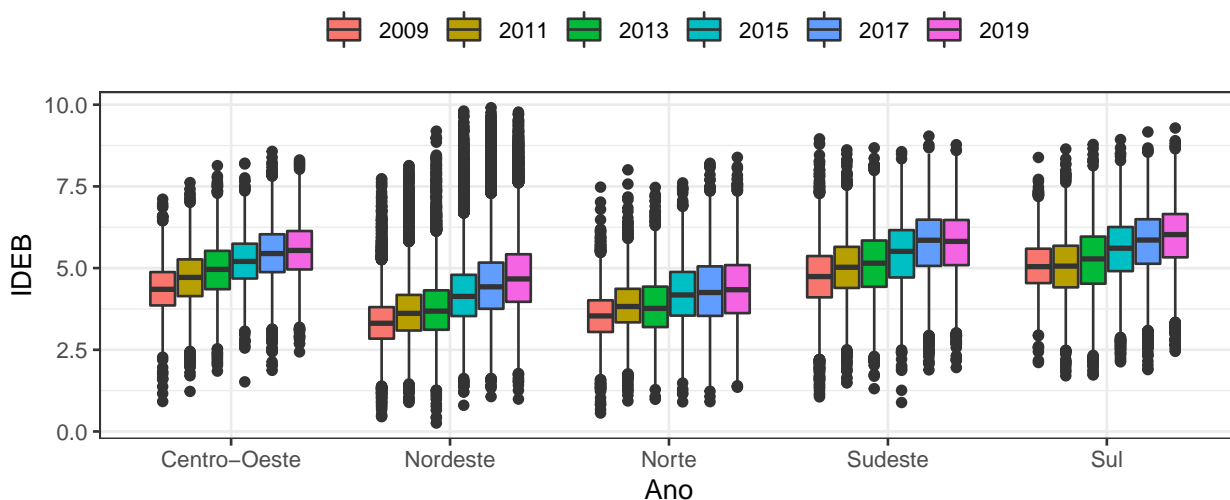


Figura 15: Evolução da Distribuição IDEB por Região

Uma outra forma de exemplificar essas diferenças regionais é através do mapa a seguir. Nele, é perceptível que, apesar de todos os estados terem avançado no IDEB no decorrer dos biênios, esse aumento não foi suficiente para romper as persistentes desigualdades regionais. Logo, ainda há diferenças consideráveis em relação a qualidade do ensino dependendo do local geográfico em que indivíduo habita.

Ressalta-se, novamente, o bom desempenho do Ceará, sendo um dos estados de maior média no IDEB. Além disso, não há dados para o Distrito Federal (por não possuir escolas municipais) e, em 2009, para o Rio Grande do Sul, sendo que, por isso, não se mostrou o mapa para esse ano.

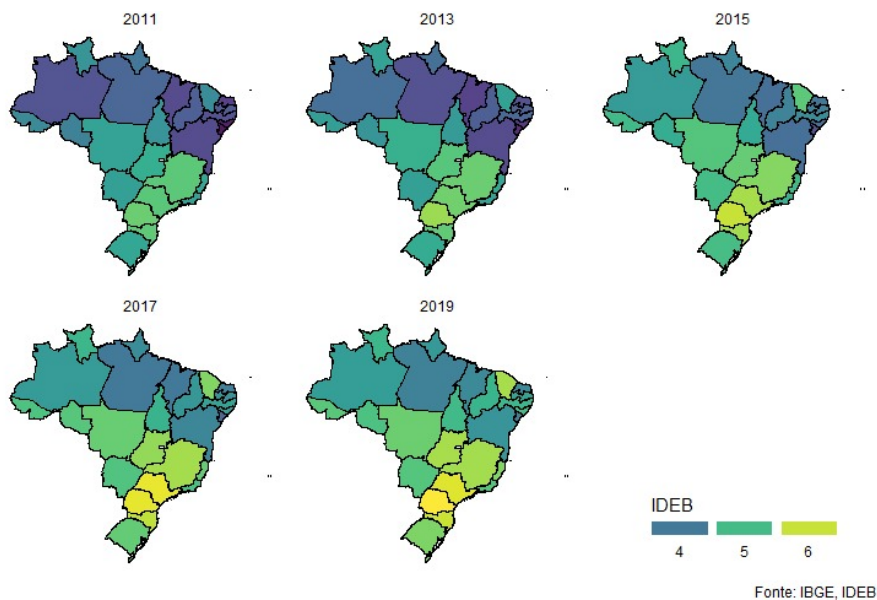


Figura 16: Mapa - Evolução Temporal do IDEB por UF

Se avaliarmos a evolução da infraestrutura dos colégios para cada uma das cinco regiões do Brasil, é possível observar que houve um aumento geral no número de escolas que passou a contar com cada componente

estrutural trabalhado até aqui, sendo que a única exceção diz respeito a diminuição do provimento de água por meio de instalações públicas. Contudo, tal tendência pode indicar que esse tipo de instalação foi substituído pelo mais convencional, como um sistema de água encanado já integrado ao da cidade.

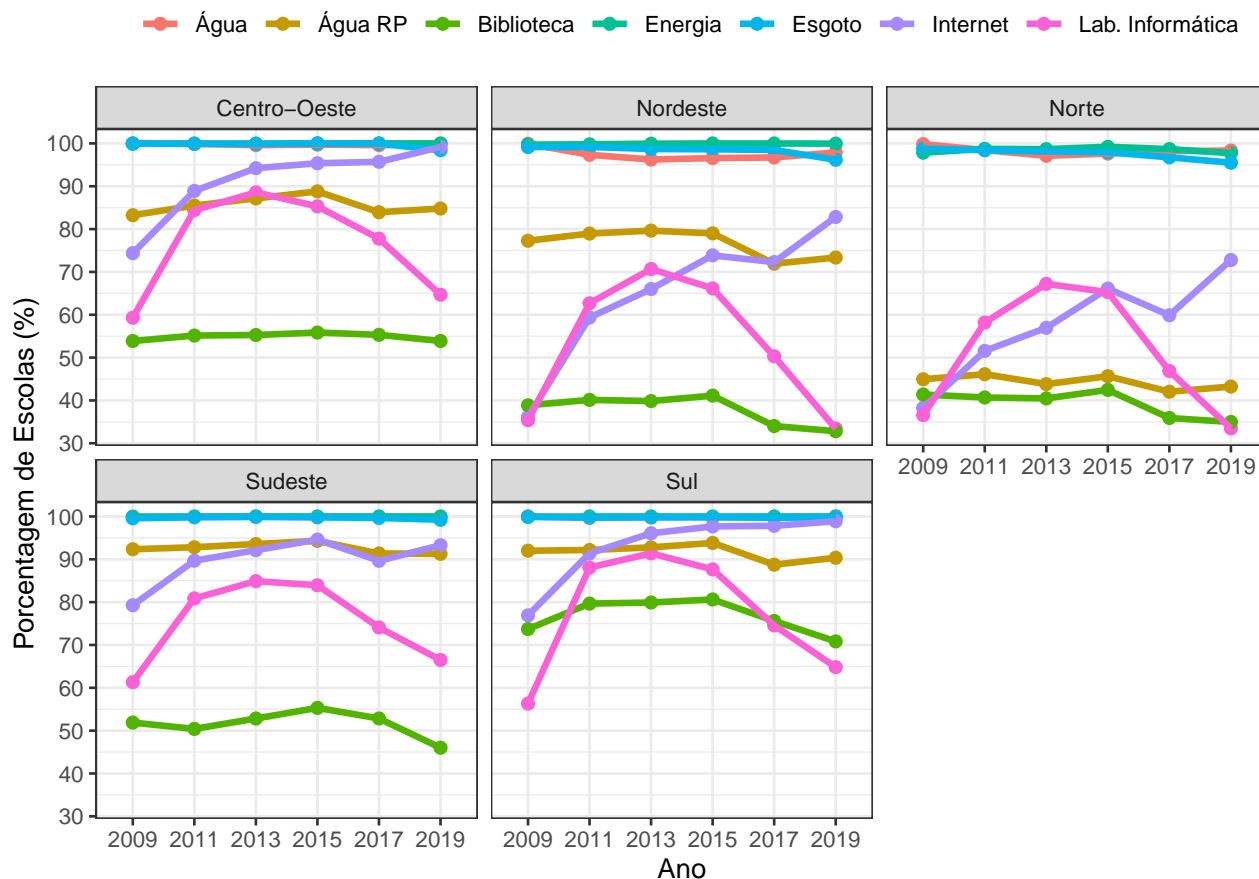


Figura 17: Evolução das Variáveis de Infraestrutura por Região

4.3 Conclusão

O conjunto das análises descritivas aqui realizadas corrobora a hipótese da importância da infraestrutura para o desempenho acadêmico dos estudantes medido pelo IDEB. Apesar de a distância entre as médias e as distribuições das escolas com ou sem infraestrutura variarem a depender da variável – com a presença de energia sendo “mais importante” que a de uma biblioteca –, vê-se que tal valor sempre é significativo.

Mais ainda, algumas variáveis – como Internet e Energia – possuem hiatos crescentes ao longo do tempo, o que pode ocorrer tanto pela maior importância dessas variáveis com a ascensão do ensino digital quanto pela existência de um efeito composição, o qual pode fazer com que apenas escolas em condições muito precárias – e, consequentemente, com menores notas no IDEB – fiquem sem tal característica de infraestrutura⁸.

Além disso, a análise regional evidencia que as desigualdades regionais brasileiras também estão presentes nas notas do IDEB, apesar de uma melhora relativa do Nordeste puxada, principalmente, pelo estado do Ceará. De modo geral, tais diferenças refletem as variações de infraestrutura entre as regiões, haja vista a maior precariedade do Norte e, em menor grau, do Nordeste e do Centro-Oeste.

⁸Isso pode explicar, por exemplo, as quebras de linha vistas em 2011 e 2013 em alguns gráficos.

5 Modelos Econométricos

Feitas as análises descritivas, o intuito desta seção é a estimação do efeito causal de algumas variáveis de infraestrutura sobre o desempenho acadêmico. Para isso, serão feitas diversas especificações de regressões e de modelos de painel, visando, na medida do possível, encontrar coeficientes não-viesados dos efeitos de uma melhor infraestrutura sobre o desempenho acadêmico medido pelo IDEB em cada uma das etapas escolares.

5.1 Breve Teoria de Modelos de Painel

Inicialmente, buscar-se-á estimar o efeito das variáveis de infraestrutura sobre o IDEB por meio de Mínimos Quadrados Ordinários Agrupados (MQOA) temporalmente. Ressalta-se que esse método tem algumas limitações e pode introduzir vieses nas estimativas se houver variáveis omitidas relevantes, mas é um bom ponto de partida.

Apesar de a adição de controles diminuir o viés das estimativas de MQOA, a endogeneidade ainda pode ocorrer devido à presença de variáveis não-observáveis que, por definição, não podem ser adicionadas ao modelo. Dessa maneira, a observação de um mesmo indivíduo por dois ou mais períodos de tempo permite a condução das estimativas econométricas por meio de **métodos de painel**, os quais visam reduzir – e, idealmente, eliminar – essa endogeneidade não observada.

Formalmente, a inconsistência decorrente da omissão de variáveis pode ser mostrada a partir da expressão para a esperança do estimador de mínimos quadrados:

$$E[\hat{\beta}] = \beta + (X'X)^{-1} \cdot E[X'u] \quad (3)$$

Assim, a presença de correlação entre alguma variável omitida – como, por exemplo, alguma característica individual c_i – relegada ao termo de erro u fará com que $E[X'u]$ seja diferente de 0, o que fará com que a esperança do estimador amostral não convirja para o valor populacional. Os modelos de painel visam justamente corrigir esse problema, assumindo as variáveis omitidas são fixas no tempo⁹.

Dessa forma, prefere-se as estimativas produzidas a partir de métodos de painel. Ademais, os outros métodos aprendidos nas aulas não são passíveis de serem implementados porque não há grupos de controle/tratamento formalmente definidos, uma vez que não se está avaliando um impacto de nenhuma política específica.

Além disso, cabe ressaltar o que já foi dito na seção acerca do IDEB: dada a manipulação da nota da prova por algumas escolas, é possível que haja um erro de medida na variável dependente. Contudo, como tal erro tende a ser sempre para cima – não estando, portanto, correlacionado com nenhum regressor –, as estimativas não sofrem viés por conta disso. Ademais, as análises serão segmentadas em dois bancos de dados correspondentes a cada etapa escolar, o que tem por objetivo captar as disparidades vistas na **Análise Descritiva** nos efeitos de infraestrutura sobre o desempenho acadêmico no Ensino Fundamental I ou II.

Os outros métodos de painel usados neste trabalho serão os de Efeitos Fixos (EF) e de Efeitos Aleatórios (EA). A abordagem de Efeitos Fixos permite que o efeito individual c_i seja correlacionado com o vetor de regressores X , o que é possível substituindo a Equação (5) de (4), resultando na Expressão (6):

$$y_{it} = \beta_1 x_{it} + c_i + u_{it} \quad (4)$$

$$\bar{y}_i = \beta_1 \bar{x}_i + c_i + \bar{u}_i \quad (5)$$

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it} + \ddot{u}_{it}, \quad (6)$$

⁹Uma hipótese forte, de modo que os métodos de painel apenas mitiguem o problema de viés por variáveis omitidas.

onde os dois pontos em cima da variável indicam que é preciso utilizar os dados centrados na média. Como a variável c_i é constante no tempo¹⁰, sua média é ela mesma, e quando subtraído de seu próprio valor, sua presença é eliminada.

A metodologia de Efeitos Aleatórios, por sua vez, assume que quaisquer características individuais c_i são independentes e não correlacionadas com X em todos os períodos da amostra. Trata-se de uma hipótese forte, que, se não for verdadeira, faz com que as transformações do estimador não gerem resultados consistentes.

Em termos matemáticos, este estimador se difere do MQO Agrupado pois define o erro como um termo composto $v_{it} = a_i + u_{it}$, onde o termo $v_{i,t}$ é serialmente correlacionado. Sob as hipóteses do modelo, tal autocorrelação é igual a:

$$\text{Corr}(v_{it}, v_{is}) = \sigma_c^2 / (\sigma_a^2 + \sigma_u^2), t \neq s, \quad (7)$$

onde σ_c^2 é a variância do termo c_i e σ_u^2 a variância do termo u_i . Realiza-se, por fim, uma transformação para eliminar a correlação serial, que consiste na equação

$$(y_{it} - \theta \bar{y}_i) = (X_{it} - \bar{X}_i)\beta + (a_i - \theta a_i) + (u_{it} - \theta \bar{u}_i), \quad (8)$$

sendo $\theta = 1 - \sqrt{\frac{\sigma_c^2}{\sigma_a^2 + \sigma_u^2}}$. A presença de θ faz com que o EA seja também lembrado por usar dados quase-centrados na média.

Antes de dar início as análises, é importante observar a estrutura do painel, sendo que o grau de balanceamento de cada banco pode ser visto abaixo; quanto mais próximo de 1 os valores de γ e ν , mais balanceado é o painel. Dessa forma, vê-se que ambos os *datasets* tem uma estrutura de balanceamento parecida, mas não 100% ideal. Ao fim do trabalho, serão feitos testes de robustez com o modelo preferido usando apenas a subamostra totalmente balanceada.

Tabela 3: Medidas de Ahrens e Pincus (1981) para Painéis Desbalanceados

	Ensino Fundamental I	Ensino Fundamental II
gamma	0.6710	0.6679
nu	0.8168	0.7858

5.2 Resultados

5.2.1 Modelos Básicos

Os primeiros modelos para o Ensino Fundamental I e II podem ser vistos abaixo, nos quais se considera apenas o ano da observação, sua região – sendo a base o Centro-Oeste – e as variáveis de infraestrutura vistas na seção de **Análise Descritiva**¹¹, sendo que a variável dependente é o logaritmo da nota do IDEB, o que dá aos coeficientes interpretações percentuais. Todos os erros-padrão são robustos à heteroscedasticidade e à autocorrelação serial¹².

¹⁰Apesar de mais fraca que a hipótese abaixo de Efeitos Aleatórios, assumir c_i constante no tempo também é algo forte, de forma que EF apenas reduz os problemas de vies advindo de variáveis não-observáveis.

¹¹Não se colocou *dummies* para a presença de quadras de esporte por os dados terem erros de declaração em 2011 e 2013. Além disso, omitiu-se a presença ou não de um laboratório de informática.

¹²Os erros-padrão foram obtidos usando a função `plm::vcovHC()` com `type = "HC3"` e `cluster = "group"`.

Tabela 4: Modelos Básicos em Paineis - Anos Iniciais

	MQO Agrupado	Efeitos Fixos	Efeitos Aleatórios
Água RP	0,048*** (0,002)	0,000 (0,003)	0,031*** (0,002)
Energia	0,213*** (0,017)	0,003 (0,019)	0,107*** (0,017)
Esgoto	0,047*** (0,007)	-0,011** (0,005)	0,017*** (0,005)
Internet	0,101*** (0,002)	0,028*** (0,001)	0,048*** (0,001)
Biblioteca	0,018*** (0,001)	-0,002** (0,001)	0,006*** (0,001)
Nordeste	-0,170*** (0,004)		-0,223*** (0,004)
Norte	-0,179*** (0,003)		-0,199*** (0,003)
Sudeste	0,066*** (0,003)		0,072*** (0,003)
Sul	0,076*** (0,003)		0,081*** (0,003)
2011	0,067*** (0,001)	0,077*** (0,001)	0,074*** (0,001)
2013	0,090*** (0,001)	0,103*** (0,001)	0,100*** (0,001)
2015	0,172*** (0,001)	0,190*** (0,001)	0,185*** (0,001)
2017	0,226*** (0,001)	0,238*** (0,001)	0,234*** (0,001)
2019	0,246*** (0,001)	0,265*** (0,001)	0,259*** (0,001)
Intercepto	1,108*** (0,018)		1,293*** (0,018)
Observations	170.094	170.094	170.094
R ²	0,471	0,465	0,598

Notas: *p<0,1; **p<0,05; ***p<0,01

Tabela 5: Modelos Básicos em Paineis - Anos Finais

	MQO Agrupado	Efeitos Fixos	Efeitos Aleatórios
Água RP	0,010*** (0,003)	-0,013*** (0,004)	0,005* (0,003)
Energia	0,174*** (0,025)	0,007 (0,021)	0,105*** (0,021)
Esgoto	0,029*** (0,010)	-0,006 (0,009)	0,016** (0,008)
Internet	0,080*** (0,003)	0,023*** (0,002)	0,044*** (0,002)
Biblioteca	0,003 (0,002)	-0,009*** (0,002)	-0,001 (0,002)
Nordeste	-0,140*** (0,007)		-0,167*** (0,007)
Norte	-0,167*** (0,006)		-0,169*** (0,006)
Sudeste	0,016*** (0,006)		0,023*** (0,006)
Sul	0,023*** (0,006)		0,025*** (0,007)
2011	0,051*** (0,002)	0,054*** (0,002)	0,053*** (0,002)
2013	0,056*** (0,003)	0,062*** (0,002)	0,060*** (0,002)
2015	0,126*** (0,003)	0,134*** (0,002)	0,130*** (0,002)
2017	0,192*** (0,003)	0,184*** (0,002)	0,183*** (0,002)
2019	0,233*** (0,003)	0,238*** (0,003)	0,235*** (0,002)
Intercepto	1,068*** (0,028)		1,172*** (0,024)
Observations	63.898	63.898	63.898

Tabela 5: Modelos Básicos em Paineis - Anos Finais

R^2	0,295	0,311	0,493
-------	-------	-------	-------

Notas: * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$

Inicialmente, os estimadores de MQO corroboram as análises feitas na seção anterior: as variáveis de infraestrutura parecem ser mais importantes para a aprendizagem no Ensino Fundamental I, o que mostra a importância da satisfação de necessidades básicas para a melhor alfabetização. Além disso, há indícios de que o fornecimento de energia contribui mais para a nota do IDEB – cerca de 21% no EF1 –, seguido da presença de internet e de uma rede de esgoto. Por fim, as desigualdades regionais na qualidade do ensino também ficam bastante aparentes.

Comparando os resultados de MQO com os encontrados via Efeitos Fixos e Efeitos Aleatórios, é possível observar alterações significativas nos coeficientes das variáveis estruturais. Em especial, a redução no módulo dos coeficientes vista principalmente em EF parece indicar um viés de habilidades não-observáveis, o que superestima os efeitos das variáveis estruturais em MQOA. De fato, o teste de Hausman produz um p-valor 0 até 16 casas decimais em ambas as etapas escolares: EA é viesado e prefere-se EF, o qual será analisado daqui em diante.

Nos Anos Iniciais, as variáveis de Água da Rede Pública e Energia – a mais significativa em MQOA – apresentaram coeficientes próximos de zero e não significantes. A presença de esgoto, por sua vez, possui um coeficiente negativo em ambas as etapas escolares, o que pode ser reflexo de que apenas escolas de menor condição socioeconômica – que costumam ter notas menores no IDEB – ainda não tinham acesso a esse item de saneamento na última década. De fato, a Tabela 2 mostra que, na maioria dos anos, mais de 99% dos colégios da amostra tinha esgoto devidamente provisionado.

A presença de uma biblioteca parece prejudicar a nota do IDEB em ambas as etapas, mas, apesar de estatisticamente significativa, seu coeficiente não é muito grande e pode estar sendo afetado por características não-observáveis que variam no tempo. No caso dos anos finais, a variável estrutural de água teve coeficiente negativo e significativo no modelo de EF, o que pode ocorrer devido a problemas amostrais. Há também uma tendência à privatização do fornecimento de água nos estados, de modo que o coeficiente pode estar captando o efeito em escolas mais “precárias”, assim como no caso do Esgoto.

Dada as propriedades de Efeitos Fixos, o valor das disparidades regionais não pode ser estimado – já que a região de cada escola é fixa no tempo –, mas é bem próximo em MQO e EA. A mesma proximidade é notada nas *dummies* temporais nos três modelos e em ambas as etapas.

Por fim, o gráfico abaixo mostra os coeficientes das variáveis de infraestrutura em todos os modelos. Nele, percebe-se que Energia realmente parece estar correlacionado com os efeitos não-observáveis c_i , bem como que a única característica estrutural consistentemente positiva e significativa é a Internet. Tal fato deve ter sido ainda mais exacerbado com a pandemia, tanto pela dependência do ensino digital quanto pela grande disparidade entre a conectividade de alunos de colégios mais ricos e a de crianças de instituições mais precárias:

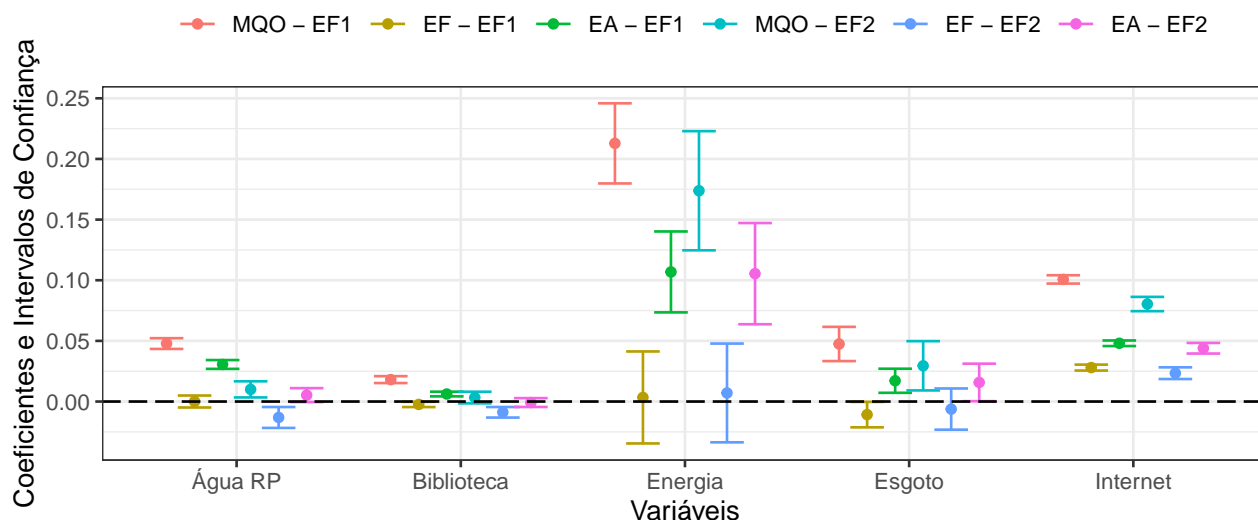


Figura 18: Comparação dos Coeficientes de Infraestrutura – Modelo Básico

5.2.2 Modelos com Controles

A robustez das análises pode ser verificada, inicialmente, com o acréscimo de variáveis de controle. Dessa forma, as tabelas abaixo mostram as estimativas com a adição de características como localização (rural ou urbano), média de alunos por turma, porcentagem de docentes com curso superior, média de horas-aula por dia¹³, percentual de alunas mulheres, o logaritmo da *proxy* para a massa de renda da escola¹⁴, a razão entre o rendimento médio dos professores daquela etapa escolar e o salário médio do município como um todo, o logaritmo da *proxy* para a despesa da escola descrita na seção de **Dados**, logaritmo do PIB per capita da cidade e, por fim, o logaritmo das transferências do FUNDEB recebidas por aquele município.

Tabela 6: Modelos com Controles - Anos Iniciais

	MQO Agrupado	Efeitos Fixos	Efeitos Aleatórios
Água RP	0,009*** (0,002)	−0,001 (0,003)	0,005** (0,002)
Energia	0,162*** (0,021)	−0,035 (0,023)	0,032 (0,021)
Esgoto	0,026*** (0,007)	−0,008 (0,005)	0,003 (0,005)
Internet	0,063*** (0,002)	0,019*** (0,001)	0,025*** (0,001)
Biblioteca	0,009*** (0,001)	−0,003** (0,001)	−0,000 (0,001)
Rural	−0,039*** (0,003)	−0,006 (0,008)	−0,059*** (0,003)
Média de Alunos por Turma	0,001*** (0,000)	−0,001*** (0,000)	−0,001*** (0,000)
Perc. de Docentes com Curso Superior	0,002*** (0,000)	0,000*** (0,000)	0,001*** (0,000)
Média de Horas-Aulas Diárias	0,007*** (0,001)	0,005*** (0,001)	0,006*** (0,001)
Perc. de Alunas Mulheres	0,504*** (0,019)	0,171*** (0,013)	0,214*** (0,013)
log(Massa de Renda da Escola)	0,004*** (0,002)	0,015*** (0,002)	0,016*** (0,001)
Razão com Salário Médio do Município	−0,012*** (0,002)	−0,010*** (0,001)	−0,012*** (0,001)

¹³Nota-se que essas informações, provenientes dos Indicadores Escolares do Inep, só existem a partir de 2011, o que diminui o horizonte temporal das análises e causa a redução do número de observações.

¹⁴Calculada a partir do salário médio dos professores de cada etapa escolar no município multiplicada pelo número de docentes do colégio.

Tabela 6: Modelos com Controles - Anos Iniciais

log(Proxy da Despesa da Escola)	−0,013*** (0,002)	−0,003*** (0,001)	−0,003*** (0,001)
log(PIB per Capita)	0,041*** (0,002)	0,021*** (0,003)	0,034*** (0,002)
log(FUNDEB)	−0,009*** (0,001)	0,037*** (0,005)	−0,004*** (0,001)
Nordeste	−0,125*** (0,004)		−0,190*** (0,004)
Norte	−0,097*** (0,003)		−0,134*** (0,004)
Sudeste	0,068*** (0,003)		0,062*** (0,003)
Sul	0,070*** (0,003)		0,073*** (0,003)
2013	−0,008*** (0,002)	0,007*** (0,002)	0,009*** (0,001)
2015	0,068*** (0,002)	0,080*** (0,002)	0,087*** (0,001)
2017	0,117*** (0,002)	0,119*** (0,003)	0,128*** (0,002)
2019	0,122*** (0,002)	0,135*** (0,003)	0,148*** (0,002)
Intercepto	0,768*** (0,029)		0,976*** (0,028)
Observations	143.642	143.642	143.642
R ²	0,477	0,391	0,527

Notas:

*p<0,1; **p<0,05; ***p<0,01

Tabela 7: Modelos com Controles - Anos Finais

	MQO Agrupado	Efeitos Fixos	Efeitos Aleatórios
Água RP	−0,003 (0,004)	−0,009* (0,005)	−0,008** (0,004)
Energia	0,180*** (0,031)	0,001 (0,030)	0,084*** (0,025)
Esgoto	0,022** (0,010)	−0,006 (0,010)	0,008 (0,009)
Internet	0,068*** (0,003)	0,019*** (0,003)	0,030*** (0,003)
Biblioteca	0,003 (0,003)	−0,007*** (0,002)	−0,004* (0,002)
Rural	−0,027*** (0,004)	−0,045*** (0,015)	−0,031*** (0,004)
Média de Alunos por Turma	−0,000 (0,000)	−0,001** (0,000)	−0,001*** (0,000)
Perc. de Docentes com Curso Superior	0,001*** (0,000)	0,000 (0,000)	0,000*** (0,000)
Média de Horas-Aulas Diárias	0,007*** (0,002)	0,012*** (0,002)	0,010*** (0,001)
Perc. de Alunas Mulheres	0,445*** (0,034)	0,218*** (0,029)	0,265*** (0,026)
log(Massa de Renda da Escola)	−0,004 (0,003)	0,014*** (0,003)	0,008*** (0,003)
Razão com Salário Médio do Município	0,030*** (0,004)	0,002 (0,003)	0,008*** (0,003)
log(Proxy da Despesa da Escola)	−0,030*** (0,003)	0,008*** (0,002)	−0,003 (0,002)
log(PIB per Capita)	0,057*** (0,003)	0,022*** (0,005)	0,043*** (0,003)
log(FUNDEB)	−0,000 (0,001)	0,007 (0,009)	−0,002* (0,001)
Nordeste	−0,126*** (0,007)		−0,142*** (0,007)
Norte	−0,111*** (0,006)		−0,110*** (0,007)
Sudeste	0,001 (0,006)		0,011* (0,006)
Sul	0,006 (0,006)		0,015** (0,007)
2013	−0,026*** (0,003)	0,003 (0,004)	−0,006** (0,003)
2015	0,043*** (0,004)	0,065*** (0,005)	0,057*** (0,003)
2017	0,105*** (0,004)	0,108*** (0,006)	0,103*** (0,003)

Tabela 7: Modelos com Controles - Anos Finais

2019	0,136*** (0,004)	0,156*** (0,007)	0,149*** (0,003)
Intercepto	0,682*** (0,049)		0,658*** (0,044)
Observations	54.715	54.715	54.715
R ²	0,297	0,283	0,453

Notas:

*p<0,1; **p<0,05; ***p<0,01

Como se pode ver, as variáveis de infraestrutura ainda são relevantes em MQO, mas tem sua importância ligeiramente diminuída com a adição de controles, o que sugere que a presença de tais características está positivamente correlacionada com as variáveis que estavam sendo omitidas. Mais ainda, as variáveis de energia e internet parecem ser agora igualmente relevantes para ambas as etapas, enquanto o fornecimento de água pela rede pública possui efeitos mistos – mas próximos de 0 – a depender da etapa escolar.

A comparação dos coeficientes das variáveis estruturais do MQO com os de EA e EF é semelhante aos modelos básicos, no sentido de que os coeficientes são bastante reduzidos. Novamente, o fornecimento de energia – característica mais relevante em MQO – se torna estatisticamente insignificante, indicando correlação com variáveis omitidas mesmo com a adição de controles. As variáveis de água e esgoto, como na seção anterior, podem estar captando o efeito de escolas mais precárias e com menor IDEB.

Novamente, a única variável positiva e significativa nas duas etapas e nos três modelos é a presença de Internet. De fato, essa característica se expandiu fortemente nas escolas brasileiras com o *Programa Banda Larga nas Escolas*, que estabeleceu que empresas de telefonia eram obrigadas a fornecer ao menos 2 megabytes de banda larga às instituições dos municípios onde começavam a prestar serviços.

O gráfico abaixo compara as estimativas dos três modelos para as variáveis estruturais, corroborando o que foi dito acima: muitas delas se tornam insignificantes com EF – que, com a adição de controles e pelo teste de Hausman, é o **modelo preferido** –, sendo que a única característica positiva é o fornecimento de Internet:

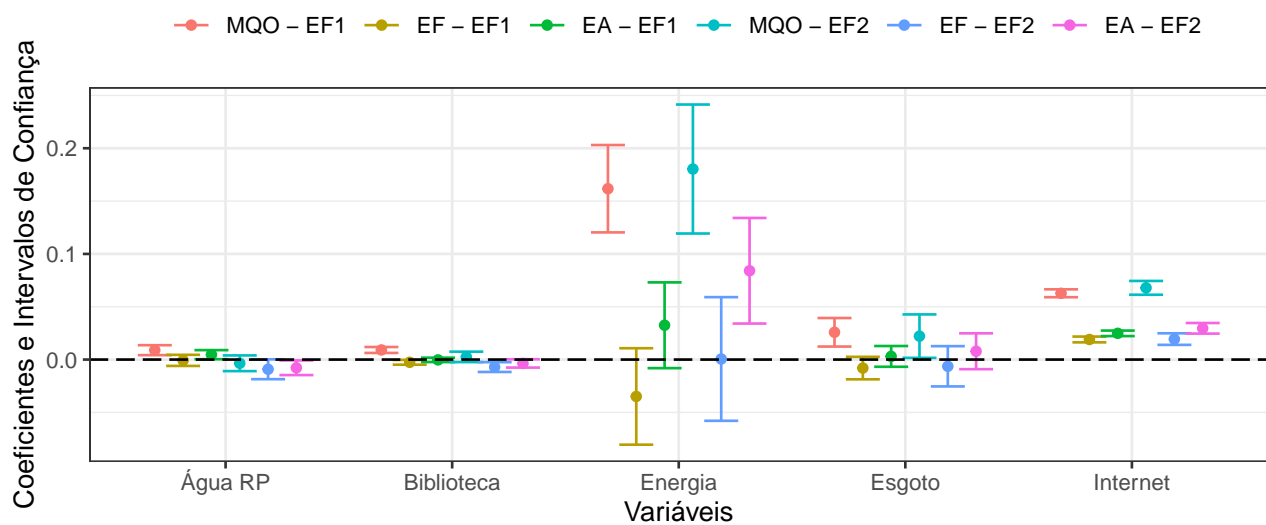


Figura 19: Comparação dos Coeficientes de Infraestrutura – Modelo com Controles

Partindo para a análise dos demais controles, a variável binária de Rural possui efeitos negativos e

significantes em todos os modelos. Dado que essa variável é teoricamente fixa no tempo, a modelagem de EF deveria eliminar seu efeito, de modo que o valor estimado diz respeito a escolas que tiveram sua classificação alterada por questões de registro ou outros fatores que inviabilizam a interpretação.

O PIB per Capita tem efeito positivo em ambas as etapas escolares e em todos os modelos. A razão dos salários de professores com a média do município e a massa de renda do colégio possuem efeitos mistos a depender da etapa escolar, o que pode ser um resultado da estrutura da carreira de professor do Brasil, a qual relaciona diretamente melhores salários a tempo de trabalho (e não qualidade do ensino).

Além disso, a despesa da escola apresenta valores negativos no EF1 e positivos ou insignificantes no EF2, o que corrobora a dificuldade de se estimar o efeito de variáveis monetárias sobre a qualidade do ensino. Ademais, o FUNDEB parece ter efeitos positivos sobre o IDEB do EF1, mas insignificante nos Anos Finais.

Por fim, chama a atenção o alto coeficiente atribuído à porcentagem de mulheres frente ao número total de alunos da escola, positivo e significativo nos três modelos. As desigualdades regionais vistas nos modelos anteriores de MQOA e EA são, em geral, mantidas (especialmente no EF1), mas agora em menor intensidade.

5.2.3 Interações Anuais e Regionais

Primeiramente, é interessante analisar como variam os efeitos das variáveis de infraestrutura ao decorrer do tempo, o que é possível de ser feito com interações de tais características com os anos (2011 a 2019). Os resultados para o modelo preferido (**Efeitos Fixos com controles**) podem ser vistos no apêndice.

A maioria das variáveis estruturais não apresenta mudança significativa no tempo, apesar de a inclusão das interações ter tornado as *dummies* de infraestrutura e de ano insignificantes. Vale mencionar, porém, que a Internet parece ter aumentado sua importância ao longo do tempo nos Anos Iniciais, sendo já muito relevante para os Anos Finais em 2011.

O oposto pode ser dito para a Biblioteca, cujos efeitos parecem ter se tornado mais negativos ao longo do tempo. Nesse sentido, especula-se que esses espaços são, cada vez mais, destinados a alunos que necessitam estudar mais por estarem ficando pra trás, o que não necessariamente se traduz em resultados melhores. Em termos regionais, chama atenção a ascensão do Nordeste em ambas as etapas de ensino, bem como a piora do desempenho do Sudeste e do Sul no IDEB dos Anos Finais.

É interessante também analisar como variam os efeitos das variáveis de infraestrutura a depender do local do país, o que é possível de ser feito com interações de tais características com as regiões. Os resultados podem ser vistos no apêndice.

Os efeitos de água parecem ser homogêneos regionalmente, enquanto o fornecimento de Energia produz resultados muito positivos nos Anos Iniciais no Centro-Oeste frente às demais localidades. Ademais, o Nordeste se beneficia extensivamente da Internet – o que mostra o poder da digitalização em uma área com menor acesso à banda larga –, enquanto ela e o Sudeste veem, estranhamente, efeitos negativos da presença de uma Biblioteca. Nos Anos Finais, não parece haver nenhuma disparidade significativa entre os efeitos das variáveis de infraestrutura nas regiões.

5.2.4 Testes de Robustez

A fim de se checar a robustez das análises, usou-se a especificação do modelo escolhido em duas subamostras de dados: aquela com colégios que ofertam exclusivamente uma das etapas escolares e outra apenas com escolas que aparecem em todos os anos. Os resultados podem ser vistos no apêndice.

Como se pode ver abaixo, as conclusões do modelo de Efeitos Fixos com controles usando toda a amostra se mantêm, o que reforça o papel positivo da Internet sobre o IDEB:

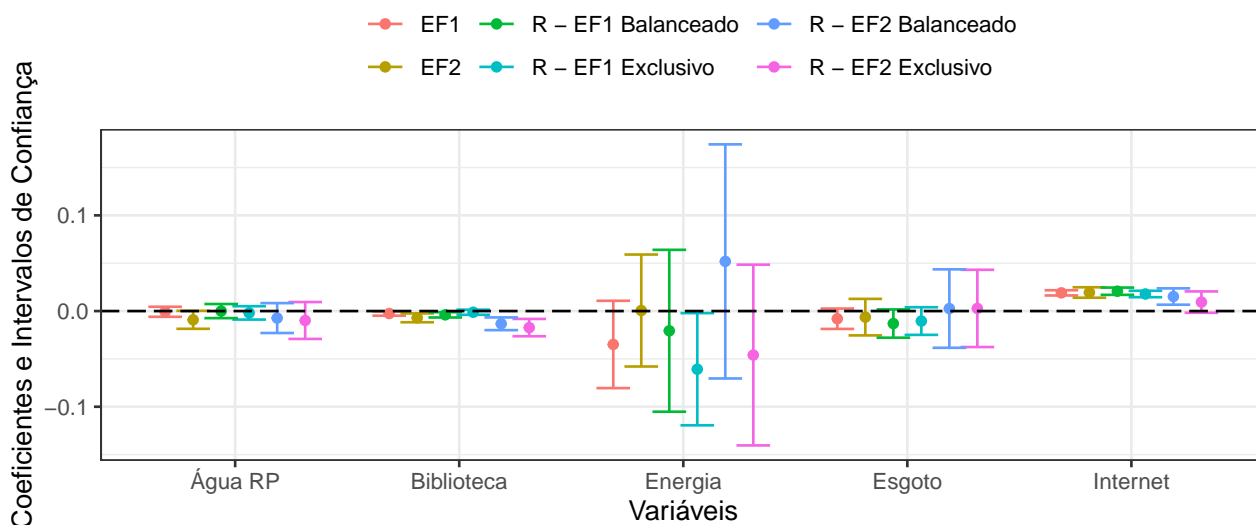


Figura 20: Comparação dos Coeficientes de Infraestrutura – Testes de Robustez

6 Conclusão

A partir de uma base de dados originada de múltiplas fontes, o presente estudo buscou contribuir para o entendimento do maior desafio da educação brasileira no começo do século XXI: a melhora de sua qualidade. Dessa forma, buscou-se estimar os efeitos de ganhos de infraestrutura sobre o IDEB, o que pode vir a servir para guiar políticas públicas e decisões de priorização de investimentos.

As análises descritivas mostram diferenças universais entre as escolas com e sem infraestrutura, sendo que tal disparidade vem crescendo ao longo da década. Os modelos econométricos de painel, porém, mostram que os efeitos dessas características se confundem e, com a adição de outros controles e de variáveis socioeconômicas, podem se tornar insignificantes.

Em termos de resultados significativos, a única característica que sobreviveu a todos os modelos e aos testes de robustez foi o fornecimento de Internet, o que mostra a importância da digitalização para a qualidade do ensino. Mais ainda, esse efeito é maior no Nordeste, área com menor acesso a banda larga e que vem obtendo melhoras frente ao resto do país na última década, sendo o Ceará o caso emblemático.

Portanto, fica clara a necessidade de expansão de investimentos para aumentar a conectividade dos alunos, o que se tornou ainda mais urgente e prioritário com a pandemia, a qual escancarou as disparidades entre escolas de melhor e pior condição socioeconômica. Assim, a melhora da qualidade da educação no Brasil passa, necessariamente, pelo maior fornecimento de Internet para os milhões de estudantes do país.

7 Apêndice

7.0.1 Interações – Ano

Tabela 8: Modelos com Interações Anuais

	Anos Iniciais	Anos Finais
Água RP	0,002 (0,004)	0,005 (0,006)
Energia	−0,084* (0,051)	0,064 (0,047)
Esgoto	−0,019 (0,018)	0,023 (0,025)
Internet	−0,000 (0,003)	0,011** (0,005)
Biblioteca	0,005*** (0,002)	−0,003 (0,004)
Rural	−0,012 (0,008)	−0,047*** (0,014)
Média de Alunos por Turma	−0,000*** (0,000)	−0,001** (0,000)
Perc. de Docentes com Curso Superior	0,000 (0,000)	−0,000*** (0,000)
Média de Horas-Aulas Diárias	0,003*** (0,001)	0,007*** (0,001)
Perc. de Alunas Mulheres	−0,009*** (0,001)	0,010*** (0,003)
log(Massa de Renda da Escola)	0,169*** (0,013)	0,249*** (0,028)
Razão com Salário Médio do Município	0,009*** (0,002)	−0,001 (0,003)
log(Proxy da Despesa da Escola)	−0,002 (0,001)	0,005* (0,002)
log(PIB per Capita)	−0,005* (0,003)	−0,020*** (0,006)
log(FUNDEB)	0,037*** (0,005)	−0,018* (0,010)
2013	−0,035 (0,067)	0,117 (0,071)
2015	−0,102 (0,098)	0,113 (0,083)
2017	0,056 (0,123)	0,392*** (0,112)
2019	0,054 (0,066)	0,369*** (0,072)
2013 Água RP	0,003 (0,004)	−0,023*** (0,006)
2015 Água RP	−0,008** (0,004)	−0,023*** (0,006)
2017 Água RP	−0,001 (0,004)	−0,013** (0,007)
2019 Água RP	−0,007* (0,004)	−0,010 (0,007)
2013 Energia	0,033 (0,065)	−0,035 (0,062)
2015 Energia	0,152 (0,095)	0,041 (0,077)
2017 Energia	0,029 (0,121)	−0,224** (0,107)
2019 Energia	0,066 (0,064)	−0,157** (0,065)
2013 Esgoto	0,023 (0,020)	−0,022 (0,037)
2015 Esgoto	0,031 (0,020)	−0,022 (0,028)
2017 Esgoto	0,041** (0,020)	−0,021 (0,031)
2019 Esgoto	0,013 (0,019)	−0,019 (0,028)
2013 Internet	0,017*** (0,003)	−0,001 (0,006)
2015 Internet	0,006* (0,003)	−0,010 (0,007)
2017 Internet	0,003 (0,003)	0,000 (0,007)
2019 Internet	0,007* (0,004)	−0,016** (0,008)
2013 Biblioteca	−0,000 (0,002)	−0,009** (0,004)
2015 Biblioteca	−0,010*** (0,002)	0,005 (0,004)

Tabela 8: Modelos com Interações Anuais

2017 Biblioteca	−0,014*** (0,002)	0,003 (0,004)
2019 Biblioteca	−0,017*** (0,002)	−0,004 (0,004)
Norte 2013	−0,045*** (0,005)	−0,014 (0,010)
Nordeste 2013	−0,022*** (0,004)	0,013* (0,008)
Sudeste 2013	−0,022*** (0,003)	−0,038*** (0,007)
Sul 2013	−0,004 (0,004)	−0,043*** (0,008)
Norte 2015	0,008 (0,005)	−0,000 (0,011)
Nordeste 2015	0,061*** (0,004)	0,040*** (0,008)
Sudeste 2015	0,003 (0,003)	−0,063*** (0,008)
Sul 2015	0,003 (0,004)	−0,033*** (0,008)
Norte 2017	0,007 (0,005)	0,020* (0,012)
Nordeste 2017	0,069*** (0,004)	0,077*** (0,009)
Sudeste 2017	−0,008** (0,004)	−0,053*** (0,009)
Sul 2017	−0,016*** (0,004)	−0,025** (0,010)
Norte 2019	0,016*** (0,005)	0,030** (0,012)
Nordeste 2019	0,110*** (0,004)	0,133*** (0,009)
Sudeste 2019	−0,013*** (0,004)	−0,042*** (0,009)
Sul 2019	−0,002 (0,004)	−0,009 (0,010)
Observations	143.642	54.715
R ²	0,428	0,322

Notas: *p<0,1; **p<0,05; ***p<0,01

7.0.2 Interações – Região

Tabela 9: Modelos com Interações Regionais

	Anos Iniciais	Anos Finais
Água RP	−0,009 (0,009)	−0,031 (0,027)
Energia	0,304*** (0,008)	0,073 (0,164)
Esgoto	−0,004 (0,023)	−0,070 (0,098)
Internet	−0,001 (0,007)	0,011 (0,015)
Biblioteca	0,006* (0,004)	−0,005 (0,009)
Rural	−0,008 (0,008)	−0,045*** (0,015)
Média de Alunos por Turma	−0,001*** (0,000)	−0,001** (0,000)
Perc. de Docentes com Curso Superior	0,000*** (0,000)	0,000 (0,000)
Média de Horas-Aulas Diárias	0,005*** (0,001)	0,011*** (0,002)
Perc. de Alunas Mulheres	0,171*** (0,013)	0,219*** (0,029)
log(Massa de Renda da Escola)	0,014*** (0,002)	0,013*** (0,003)
Razão com Salário Médio do Município	−0,009*** (0,001)	0,003 (0,003)
log(Proxy da Despesa da Escola)	−0,003*** (0,001)	0,007*** (0,002)
log(PIB per Capita)	0,019*** (0,003)	0,022*** (0,005)

Tabela 9: Modelos com Interações Regionais

log(FUNDEB)	0,040*** (0,005)	0,009 (0,009)
2013	0,007*** (0,002)	0,003 (0,004)
2015	0,080*** (0,002)	0,065*** (0,005)
2017	0,118*** (0,003)	0,107*** (0,006)
2019	0,134*** (0,003)	0,155*** (0,007)
Norte Água RP	0,014 (0,011)	0,029 (0,030)
Nordeste Água RP	0,007 (0,010)	0,019 (0,027)
Sudeste Água RP	0,013 (0,011)	0,017 (0,029)
Sul Água RP	−0,003 (0,012)	0,076** (0,031)
Norte Energia	−0,338*** (0,033)	−0,081 (0,166)
Nordeste Energia	−0,354*** (0,033)	
Sudeste Energia	−0,227*** (0,011)	
Norte Esgoto	0,011 (0,028)	0,052 (0,100)
Nordeste Esgoto	−0,009 (0,024)	0,064 (0,098)
Sudeste Esgoto	0,015 (0,026)	0,081 (0,104)
Sul Esgoto	0,001 (0,033)	0,046 (0,106)
Norte Internet	−0,001 (0,008)	−0,022 (0,017)
Nordeste Internet	0,039*** (0,007)	0,022 (0,016)
Sudeste Internet	−0,008 (0,007)	−0,016 (0,016)
Sul Internet	−0,004 (0,008)	0,008 (0,020)
Norte Biblioteca	−0,007 (0,005)	0,001 (0,012)
Nordeste Biblioteca	−0,011** (0,004)	−0,005 (0,010)
Sudeste Biblioteca	−0,012*** (0,004)	−0,004 (0,010)
Sul Biblioteca	−0,001 (0,004)	0,018 (0,012)
Observations	143.642	54.715
R ²	0,393	0,285

Notas: *p<0,1; **p<0,05; ***p<0,01

7.0.3 Etapas Exclusivas

Tabela 10: Modelos com Escolas Exclusivas

	Anos Iniciais	Anos Finais
Água RP	−0,002 (0,004)	−0,010 (0,010)
Energia	−0,061** (0,030)	−0,046 (0,048)
Esgoto	−0,010 (0,007)	0,003 (0,021)
Internet	0,018*** (0,002)	0,009 (0,006)
Biblioteca	−0,001 (0,001)	−0,017*** (0,005)
Rural	−0,003 (0,011)	−0,100** (0,046)
Média de Alunos por Turma	−0,001*** (0,000)	−0,001 (0,001)
Perc. de Docentes com Curso Superior	0,000*** (0,000)	0,000 (0,000)

Tabela 10: Modelos com Escolas Exclusivas

Média de Horas-Aulas Diárias	0,004*** (0,001)	0,008*** (0,002)
Perc. de Alunas Mulheres	0,166*** (0,016)	0,239*** (0,052)
log(Massa de Renda da Escola)	0,015*** (0,002)	0,007 (0,006)
Razão com Salário Médio do Município	−0,009*** (0,002)	0,005 (0,005)
log(Proxy da Despesa da Escola)	−0,004*** (0,001)	0,007 (0,005)
log(PIB per Capita)	0,016*** (0,003)	0,007 (0,010)
log(FUNDEB)	0,056*** (0,006)	0,070*** (0,020)
2013	0,003 (0,002)	0,003 (0,008)
2015	0,072*** (0,003)	0,059*** (0,010)
2017	0,111*** (0,003)	0,100*** (0,012)
2019	0,124*** (0,004)	0,150*** (0,014)
Observations	98.755	15.780
R ²	0,395	0,290

Notas: *p<0,1; **p<0,05; ***p<0,01

7.0.4 Painel Balanceado

Tabela 11: Modelos com Painel Balanceado

	Anos Iniciais	Anos Finais
Água RP	−0,000 (0,004)	−0,007 (0,008)
Energia	−0,021 (0,043)	0,052 (0,062)
Esgoto	−0,013* (0,008)	0,003 (0,021)
Internet	0,021*** (0,002)	0,015*** (0,004)
Biblioteca	−0,004*** (0,001)	−0,013*** (0,003)
Rural	−0,021** (0,011)	0,023 (0,018)
Média de Alunos por Turma	−0,001*** (0,000)	−0,001** (0,000)
Perc. de Docentes com Curso Superior	0,000*** (0,000)	0,000** (0,000)
Média de Horas-Aulas Diárias	0,005*** (0,001)	0,014*** (0,002)
Perc. de Alunas Mulheres	0,154*** (0,017)	0,262*** (0,047)
log(Massa de Renda da Escola)	0,013*** (0,002)	0,012** (0,005)
Razão com Salário Médio do Município	−0,006*** (0,002)	0,002 (0,004)
log(Proxy da Despesa da Escola)	−0,007*** (0,002)	0,009** (0,004)
log(PIB per Capita)	0,027*** (0,004)	0,021** (0,009)
log(FUNDEB)	0,050*** (0,006)	0,008 (0,015)
2013	0,005** (0,002)	0,010* (0,006)
2015	0,072*** (0,003)	0,078*** (0,007)
2017	0,109*** (0,003)	0,122*** (0,009)
2019	0,119*** (0,004)	0,167*** (0,011)
Observations	70.792	19.434

Tabela 11: Modelos com Paineis Balanceados

R^2	0,435	0,332
Notas:	*p<0,1; **p<0,05; ***p<0,01	

Referências

- BARROS, Ricardo et al. Markets, the State, and the Dynamics of Inequality in Brazil. **Declining inequality in Latin America: A decade of progress**, Brookings Institution e UNDP, pp. 134–174, Washington, DC, v. 169, 2010.
- BASE DOS DADOS. **Base dos Dados Mais**. [S.l.: s.n.], 2022. Disponível em: <<https://github.com/basedosdados/mais>>.
- BECSKEHÁZY, Ilona. Educação. In: SHIKIDA, Claudio; MONASTÉRIO, Leonardo; NERY, Pedro Fernando (Ed.). **Guia Brasileiro de Análise de Dados: Armadilhas e Soluções**. Brasília: Escola Nacional de Administração Pública, 2021. p. 98–123.
- DUARTE, Natalia. O impacto da pobreza no Ideb: um estudo multinível. **Revista Brasileira de Estudos Pedagógicos**, v. 94, n. 237, 2013.
- DUQUE, Daniel. Os investimentos na educação: o futuro do futuro do Brasil. **Nexo**, 21 mai. 2019. Disponível em: <<https://www.nexojournal.com.br/ensaio/2019/Os-investimentos-na-educa%C3%A7%C3%A3o-o-futuro-do-futuro-do-Brasil>>. Acesso em: 22 fev. 2022.
- FIGUEIREDO, Dalson et al. Os cavalos também caem: Tratado das inconsistências do IDEB. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 26, n. 100, p. 552–572, 2018.
- GRISA, Gregório. Financiamento da educação: mais gestão e mais recursos. **Nexo**, 4 jun. 2019. Disponível em: <<https://www.nexojournal.com.br/ensaio/2019/Financiamento-da-educa%C3%A7%C3%A3o-mais-gest%C3%A3o-e-mais-recursos>>. Acesso em: 22 fev. 2022.
- INEP. **Índice de Desenvolvimento da Educação Básica (Ideb)**. 2022. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/ideb>>. Acesso em: 22 fev. 2022.
- _____. **Nota Informativa do IDEB 2011**. Brasília: Inep, 2012. Disponível em: <https://download.inep.gov.br/educacao_basica/portal_ideb/o_que_e_o_ideb/nota_informativa_ideb_2011.pdf>.
- KOMATSU, Bruno et al. Novas Medidas de Educação e de Desigualdade Educacional para a Primeira Metade do Século XX no Brasil. **Estudos Econômicos**, v. 49, n. 4, p. 687–722, 2019.
- LEEUWEN, Bas van; LEEUWEN-LI, Jieli van. **Average Years of Education**. [S.l.]: IISH Dataverse, 2015. Disponível em: <<http://hdl.handle.net/10622/KCBMKI>>.
- LÓPEZ-CALVA, Luis Felipe; LUSTIG, Nora Claudia. **Declining inequality in Latin America: A decade of progress?** [S.l.]: Brookings Institution Press, 2010.
- MAIA, Joyce de Souza Zanirato; BUENO, Ana Paula Arantes; SATO, João Ricardo. Assessing the educational performance of different Brazilian school cycles using data science methods. **PloS one**, Public Library of Science San Francisco, CA USA, v. 16, n. 3, e0248525, 2021.

OCDE. **Education Policy Outlook Brasil**. [S.l.], 2021. Disponível em: <https://www.itausocial.org.br/wp-content/uploads/2021/06/Education-Policy-Outlook_Poli%CC%81ticas-Internacionais.pdf>.

OLIVEIRA, Janaina; CASAGRANDE, Natalia Maria; GALERANI, Leide Daiana de Jorge. A Evolução Tecnológica e sua Influência na Educação. **Revista Interface Tecnológica**, v. 13, n. 1, p. 23–38, dez. 2016. Disponível em: <<https://revista.fatectq.edu.br/index.php/interfacetecnologica/article/view/123>>.

PESERICO, Cecilia; KRAVCHYCHYN, Claudio; OLIVEIRA, Amauri. Análise da Relação entre Esporte e Desempenho Escolar: um estudo de caso. **Pensar a Prática**, v. 18, jun. 2015. DOI: [10.5216/rpp.v18i2.27965](https://doi.org/10.5216/rpp.v18i2.27965).

SCRIPTORE, Juliana Souza; AZZONI, Carlos Roberto; MENEZES-FILHO, Naércio Aquino. Os impactos do saneamento básico sobre a educação: usando a privatização como variável instrumental. In: ENCONTRO Nacional de Economia. [S.l.]: ANPEC, 2018.

TESOURO NACIONAL. **Análise sobre SICONFI: Despesas com educação x IDEB**. 2019. Disponível em: <<https://analise-siconfi-ideb.tesouro.gov.br>>. Acesso em: 22 fev. 2022.

ZOGHBI, Ana Carolina et al. Uma análise da eficiência nos gastos em educação fundamental para os municípios paulistas. **Planejamento e políticas públicas**, n. 36, 2011.