



**DETERMINANTES DO AUMENTO DA ESCOLARIDADE E SEUS EFEITOS
SOBRE O DESENVOLVIMENTO ECONÔMICO PARANAENSE**

CURITIBA

2022

GOVERNO DO ESTADO DO PARANÁ

CARLOS MASSA RATINHO JÚNIOR - *Governador*

SECRETARIA DO PLANEJAMENTO E PROJETOS ESTRUTURANTES

LOUISE DA COSTA E SILVA GARNICA - *Secretária*

INSTITUTO PARANAENSE DE DESENVOLVIMENTO ECONÔMICO E SOCIAL

DANIEL NOJIMA - *Diretor-Presidente*

FRANCISCO CARLOS ROGÉRIO - *Diretor Administrativo-Financeiro*

JULIO TAKESHI SUZUKI JÚNIOR - *Diretor do Centro de Pesquisa*

GUSTAVO NUNES MOURÃO - *Diretor do Centro Estadual de Estatística*

NÚCLEO DE ESTUDOS POPULACIONAIS E SOCIAIS

Leonildo Pereira de Souza - *Coordenador*

Paulo Roberto Delgado

RESPONSÁVEL TÉCNICO

Prof. Dr. Walcir Soares da Silva Junior - *Consultor*

EDITORAÇÃO

MARCELO ANTONIO - *Coordenação*

MARIA LAURA ZOCCOLOTTI - *Supervisão editorial*

DIOGO AUGUSTO COTOVICZ - *Web Designer/Criação*

SUMÁRIO

1 INTRODUÇÃO	3
2 PREVISÃO PARA ANOS MÉDIOS DE ESCOLARIDADE	7
3 ESTATÍSTICAS DESCRITIVAS: VARIÁVEIS UTILIZADAS NO ESTUDO	14
4 MODELOS DE REGRESSÃO: RESULTADOS PRELIMINARES	25
5 MENSURANDO A INFLUÊNCIA DE INDICADORES EDUCACIONAIS: IDEB E A TAXA DE APROVAÇÃO	29
5.1 RESULTADOS DOS MODELOS FINAIS: IDEB.....	29
5.2 RESULTADOS DOS MODELOS FINAIS: TAXA DE APROVAÇÃO.....	32
6 MENSURANDO A INFLUÊNCIA DE INDICADORES EDUCACIONAIS: DESEMPENHO NO SAEB	35
6.1 RESULTADOS DOS MODELOS FINAIS: DESEMPENHO EM PORTUGUÊS E MATEMÁTICA.....	38
7 POLÍTICAS EDUCACIONAIS, ANOS MÉDIOS DE ESCOLARIDADE E PRODUTIVIDADE TOTAL DOS FATORES	42
7.1 EVOLUÇÃO TEMPORAL DAS VARIÁVEIS EDUCACIONAIS PARANAENSES....	42
7.2 A QUESTÃO DA QUALIDADE EDUCACIONAL.....	48
7.3 CORRELAÇÃO ENTRE A PRODUTIVIDADE TOTAL DOS FATORES (PTF) E ANOS MÉDIOS DE ESTUDOS PARA O PARANÁ.....	49
8 CONSIDERAÇÕES FINAIS	57
REFERÊNCIAS	60
APÊNDICE A - AJUSTE DAS SÉRIES DOS INDICADORES ANOS DE ESTUDO E TAXAS DE CONCLUSÃO DO ENSINO FUNDAMENTAL E MÉDIO	63
APÊNDICE B - TESTES DE MULTICOLINEARIDADE, HETEROCEDASTICIDADE E NORMALIDADE PARA OS MODELOS IDEB E TAXA DE APROVAÇÃO	67
APÊNDICE C - TESTES DE MULTICOLINEARIDADE, HETEROCEDASTICIDADE E NORMALIDADE PARA OS MODELOS DE DESEMPENHO	71

1 INTRODUÇÃO

Qualidade educacional é um fenômeno multidimensional. Avaliar a educação, acompanhar indicadores, traçar metas e trabalhar para atingi-las deve ser o propósito de qualquer gestor educacional, seja no nível escolar, municipal, estadual ou federal. Indicadores permitem dimensionar a magnitude de um fenômeno e avaliar sua melhora ou piora em determinado horizonte de tempo. Essa pesquisa pretende apresentar relações importantes entre indicadores educacionais disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e dar uma medida do quanto a política pública pode interceder na qualidade da educação paranaense, em especial, para o ensino fundamental (anos iniciais e finais) e o ensino médio.

De acordo com Leithwood *et al.* (2004), entre os principais fatores intraescolares que afetam o aprendizado dos alunos está a ação dos professores. Em segundo lugar está a gestão escolar que, no Brasil, se concentra na figura do diretor. Dada a evidência apresentada por Robinson, Lloyd e Rowe (2008) da importância dos gestores não só na gestão pedagógica, mas também no desenvolvimento profissional dos professores, essas duas dimensões, professores e diretores, têm uma importância ímpar quando se trata de melhorar os indicadores e, conseqüentemente, o processo de ensino-aprendizagem.

De acordo com dados do Prova Brasil 2019, compilados a partir de QEDu (2021), no 5.º ano do ensino fundamental, 71% e 65% é a proporção de alunos que aprenderam o adequado em português e matemática, respectivamente, nas escolas públicas paranaenses. Esse percentual é de 43% e 25% para português e matemática, respectivamente, no 9.º ano e 41% e 9% no terceiro ano do ensino médio em português e matemática, respectivamente. Segundo a meta do movimento Todos Pela Educação¹, até 2022, essas proporções deveriam atingir 70%.

O pior dentre esses resultados, matemática no ensino médio, com percentual de 9%, é ainda superior ao percentual para o Brasil, de 7%. Para os estados do sul, Santa Catarina possui um desempenho igual ao Paraná, enquanto Rio Grande do Sul tem uma proporção de 10%. O único resultado melhor que este no país é para o Espírito Santo, onde essa proporção chega a 15%. Para português, o resultado paranaense está entre os melhores, ainda que 41% esteja longe de poder ser considerado satisfatório. O estado perde apenas para Goiás (42%), Rio Grande do Sul (42%) e Espírito Santo (48%).

Esses dados dão uma ideia da importância de mudanças efetivas no aprendizado do país, em especial, paranaense, de modo a garantir que as metas estabelecidas sejam

¹ De acordo com QEDu (2021) o Todos Pela Educação é um movimento financiado exclusivamente pela iniciativa privada, que congrega sociedade civil organizada, educadores e gestores públicos que têm como objetivo contribuir para que o Brasil garanta a todas as crianças e jovens o direito à Educação Básica de qualidade.

atingidas e que os agentes envolvidos saibam quais indicadores precisam ser melhorados para contribuir com a melhora no aprendizado. Além do aspecto aprendizado, em que o pior nível educacional é o ensino médio, o problema da evasão escolar vinha se mostrando importante já em 2019 e se deteriorou ainda mais com as consequências da pandemia do vírus Covid-19. De acordo com dados da Secretaria da Educação e do Esporte (SEED), em 2019, 12.077 alunos desistiram da escola no ensino médio.

A literatura apresenta algumas evidências, corroboradas em parte pelos achados desta pesquisa. No que concerne aos anos iniciais do ensino fundamental, Costa e Carnoy (2015), avaliando o impacto da experiência cearense com o programa “Pacto pela Alfabetização na Idade Certa (PAIC)” ocorrido em 2007, encontraram que o programa, baseado em treinamento de diretores e professores, mobilização da rede municipal e provisão de livros para estudantes, teve efeitos positivos nas notas de Português e Matemática. O programa foi capaz de reduzir taxas de analfabetismo das crianças entre 6 e 8 anos. Esses resultados corroboram a evidência da importância da formação dos professores e da regularidade docente, apresentada neste relatório.

Menezes-Filho *et al.* (2008) encontraram evidência de efetividade do programa de progressão continuada como forma de aumentar as taxas de aprovação e reduzir taxas de abandono escolar, ainda que não haja evidência de impacto no desempenho na 4ª série do ensino fundamental (para a 8ª série a evidência é de redução do desempenho). Esse programa, como alternativa ao regime de séries, pressupõe o ensino por ciclos, em geral mais longos do que um ano ou uma série. Para os autores, utilizando um cálculo de retorno econômico, o aumento nos anos de estudo causados pelo programa mais do que compensou a redução do desempenho. Este resultado corrobora a evidência da relação negativa entre distorção idade-série e desempenho municipal em termos do IDEB, Taxa de Aprovação e notas em Português e Matemática, encontrada neste trabalho.

Para os anos finais do ensino fundamental, Bruns, Costa e Cunha (2018) apresentam uma avaliação de impacto de um programa que provê estratégias para aumentar a efetividade dos professores no estado do Ceará. O programa combinou feedbacks baseados na observação em sala de aula, mentoria de alta qualidade e disponibilidade de materiais de autoajuda para os professores na tentativa de melhorar e reduzir a variação das práticas dos professores em sala de aula. Este relatório apresenta evidência de que formação dos professores e regularidade docente estão associados a um melhor desempenho.

Bruns, Costa e Cunha (2018) encontraram que a prática dos professores em sala de aula é maleável e a melhora pode impactar positivamente o aprendizado. Os professores que participaram do programa aumentaram o uso de exercícios e atividades durante as aulas, consistente com os objetivos da mentoria de encorajar práticas de ensino mais interativas. Além disso, o programa manteve os alunos mais engajados, reduziu o tempo

em que grupos de alunos ficavam sem fazer tarefas, e ainda reduziu alguma variação nas práticas dos professores dentro das aulas, aumentando não só o desempenho nas notas de Português e Matemática nos testes padronizados, mas também o desempenho em Português no Exame Nacional do Ensino Médio (ENEM).

No entanto, no que concerne aos fatores que podem aumentar a efetividade dos professores em sala de aula, autores como Araújo *et al.* (2016) e Kane e Staiger (2008) encontraram que características observáveis como idade, educação, qualificações e regime de trabalho não explicam as diferenças de efetividade dos professores. A não ser o fato de serem menos efetivos nos primeiros três a cinco anos de docência, evidência do impacto da falta de experiência. Os resultados deste relatório corroboram essa evidência, não encontrando relação estatisticamente significativa para a remuneração dos professores, que é proxy para características observáveis como qualificação, regime de trabalho e idade.

Dada a falta de dados observacionais em sala de aula, que possam evidenciar os mecanismos desse aumento de efetividade, este estudo pretende avaliar de maneira exploratória como diversos indicadores educacionais disponibilizados pelo INEP se relacionam com variáveis resultado como o Índice de Desenvolvimento da Educação Básica (IDEB), Taxa de Aprovação e Desempenho dos alunos em Português e Matemática no Sistema de Avaliação da Educação Básica (Saeb).

A respeito do ensino médio, Gonçalves e Lourenço (2019), avaliaram o Programa Ensino Médio Inovador (ProEMI) em Santa Catarina. O programa tinha o objetivo de reduzir as taxas de abandono escolar apoiando e fortalecendo o desenvolvimento de propostas curriculares inovadoras nas escolas de ensino médio estaduais. A adesão ao programa se dá de maneira voluntária pelas Secretarias de Educação Estaduais e Distritais. Estas escolhem unidades para receber apoio técnico e financeiro para implementação de Propostas de Redesenho Curricular (PRC) que contemplem não só o conteúdo base, mas também contemplando necessidades reais da escola e da realidade local.

Utilizando uma técnica quasi-experimental de pareamento por escore de propensão, os autores encontraram efeitos positivos do programa na redução do número de faltas nas disciplinas de Português e Matemática. Biondi e Felício (2007), utilizando um painel de dados do Saeb e Censo Escolar para estimar regressões por Mínimos Quadrados Ordinários e Efeitos Fixos, encontraram que a regularidade docente (ausência de rotatividade ao longo do ano letivo e experiência média dos professores superior a dois anos) e a existência na escola de conexão com a Internet afetam positivamente o resultado médio. Este trabalho apresenta evidência de que a regularidade docente está associada a um melhor desempenho no ensino médio.

Portanto, este relatório tem o objetivo de coletar evidências empíricas exploratórias que possam embasar a tomada de decisão quanto às políticas públicas associadas às medidas de desempenho escolar municipal e confrontá-las com as evidências encontradas

na literatura. Reitera-se a qualidade exploratória deste trabalho, principalmente devido à agregação dos dados, no caso dos modelos cross-section, e dos limitados períodos de tempo para as estimações em séries temporais.

Esse relatório está dividido em oito seções a contar com esta introdução. A segunda seção apresenta uma previsão para a variável “anos médios de educação” para o Paraná, de 2020 a 2035, utilizando um modelo ARIMA (p, d, q), em duas versões. A primeira a partir da Subnational Human Development Database, uma base de dados que compila diversas informações para 161 países para compor o Subnational Human Development Index (SHDI), do qual foi utilizada a estimativa “mean years of schooling”. A segunda, a partir de uma série ajustada da Pesquisa Nacional por Amostra de Domicílios (PNAD), fornecida pelo Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES), que compatibiliza os dados antigos desta série com a recente atualização da metodologia conhecida como PNAD Contínua.

A terceira seção apresenta as variáveis utilizadas neste estudo, dependentes (Taxa de Aprovação, Índice de Desenvolvimento da Educação Básica, IDEB e Desempenho no Sistema de Avaliação da Educação Básica, Saeb, em Português e Matemática) e independentes (percentual de docentes sem ensino superior, esforço docente, quantidade de alunos em sala de aula, Indicador de Nível Socioeconômico (INSE), regularidade docente, remuneração do docente, taxa de distorção idade-série e disponibilidade de internet e computadores nas escolas). Ambos os conjuntos de variáveis, quando referentes às diferentes etapas, apresentam dados para o ensino fundamental nos anos iniciais e finais e também para o ensino médio, segundo a disponibilidade do Inep.

A quarta seção apresenta resultados preliminares para seis modelos de regressão para as variáveis dependentes IDEB e Taxa de Aprovação, analisando o processo de escolha das variáveis estatisticamente significantes que serão utilizadas nos modelos finais. A quinta seção apresenta os modelos finais mensurando a influência dos indicadores educacionais sobre o IDEB e a Taxa de Aprovação. A sexta seção mensura essa influência sobre os desempenhos em Português e Matemática no Saeb. Por fim, a seção 7 apresenta um panorama descritivo e exploratório da evolução temporal do investimento agregado em educação e das variáveis Taxa de Abandono, Reprovação e Aprovação e distorção idade-série apresentadas na seção 3, para o Paraná, complementando a evolução temporal das variáveis de desempenho apresentadas na seção 6. Ainda na seção 7, discute-se a questão da qualidade educacional e algumas evidências da literatura brasileira e internacional a esse respeito. Por fim, é apresentada uma correlação exploratória entre a produtividade total dos fatores estimada para o Paraná e os anos médios de estudo. A oitava seção apresenta as principais conclusões deste trabalho.

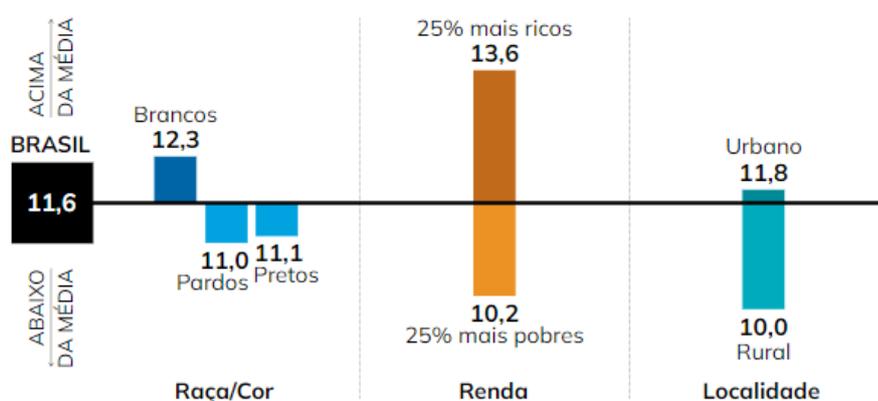
2 PREVISÃO PARA ANOS MÉDIOS DE ESCOLARIDADE

De acordo com dados da PNAD Contínua (2019), o Paraná tem atualmente uma média de anos de estudo de 9,54 anos, considerando pessoas de 25 anos ou mais de idade e o ensino fundamental com duração de 9 anos. Considerando a população de 18 a 29 anos, a escolaridade média paranaense em 2019 foi de 11,9 anos. Para o Brasil, essa média é de 11,6. A meta de nº 8 do Plano Nacional de Educação (PNE) consiste em:

Elevar a escolaridade média da população de 18 (dezoito) a 29 (vinte e nove) anos, de modo a alcançar, no mínimo, 12 (doze) anos de estudo no último ano de vigência deste Plano, para as populações do campo, da região de menor escolaridade no País e dos 25% (vinte e cinco por cento) mais pobres, e igualar a escolaridade média entre negros e não negros declarados à Fundação Instituto Brasileiro de Geografia e Estatística - IBGE. (LEI N.º 13.005/2014).

Outro aspecto muito importante quando se trata dos anos médios de escolaridade é a desigualdade. No Brasil, a escolaridade média considerando os 25% mais ricos, é de 13,6 anos. Para a população rural essa média cai para 10 anos de estudo, e essa defasagem se verifica também quando analisados os 25% mais pobres e as diferenças entre brancos, pretos e pardos, como mostra a figura 1. Portanto, prever a evolução dos anos médios de escolaridade é de extrema importância, não só para cumprir metas estabelecidas, verificar quais os horizontes tendenciais caso as políticas do presente permaneçam, mas também compreender como superar essas tendências.

FIGURA 1 - MÉDIA DE ANOS DE ESTUDO DA POPULAÇÃO DE 18 A 29 ANOS



FONTE: IBGE/PNAD Contínua. Elaboração

NOTA: Todos Pela Educação.

As previsões apresentadas nesta seção foram feitas a partir de modelos de série-temporal ARIMA (p, d, q). Conhecida também como metodologia Box-Jenkins (BJ), a ênfase

destes métodos está na análise probabilística das propriedades da própria série temporal econômica, levando em conta também particularidades como sazonalidade e tratamento de tendência temporal (autocorrelação serial).

De acordo com Gujarati (2011) modelar um modelo de série temporal ARIMA (p, d, q) é identificar se a série segue um processo AR puro (componente autorregressivo, identificado pelo parâmetro p); se segue um processo MA puro (componente de médias móveis, identificado pelo parâmetro q); ou a combinação desses processos. Esses modelos requerem também que a série tenha média e variância constantes, propriedade conhecida como estacionariedade. Caso não o seja, a série precisará passar por d diferenciações, número esse identificado pelo parâmetro d .

Segundo Gujarati (2011) a metodologia Box-Jenkins possui quatro etapas principais: i) identificação, em que os parâmetros p, d e q são identificados; ii) estimação, fase em que os parâmetros dos termos autorregressivos e de médias móveis do modelo são estimados; iii) verificação de diagnóstico, em que é analisada a qualidade do ajuste do modelo aos dados e; iv) previsão, em que a série é extrapolada com base no modelo ajustado. Esses foram os passos seguidos nesta estimação/previsão.

Optou-se, neste trabalho, por realizar a estimativa considerando duas séries históricas referentes à variável anos médios de estudo: a primeira, da *Subnational Human Development Index* (SHDI), que segue o padrão utilizado nos relatórios internacionais do Índice de Desenvolvimento Humano (IDH), onde a variável é calculada a partir de ciclos do processo educativo padronizados internacionalmente; e a segunda com base em dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), em que a variável pontua o último ano/série escolar concluído pelo indivíduo. Para essa segunda série, foi realizado um processo de ajuste que será descrito após a apresentação do segundo modelo.

O primeiro modelo estimado, a partir dos dados *Subnational Human Development Index* (SHDI) da média de anos de estudo considerando pessoas de 25 anos ou mais de idade, foi identificado utilizando os critérios de significância para todas as defasagens envolvidas e menores valores dos critérios de Akaike, Schwartz e Hannan-Quinn. Como a série é não estacionária, precisou passar por uma diferenciação, resultando em um modelo ARIMA (2, 1, 2). Os resultados podem ser vistos no Modelo 1.

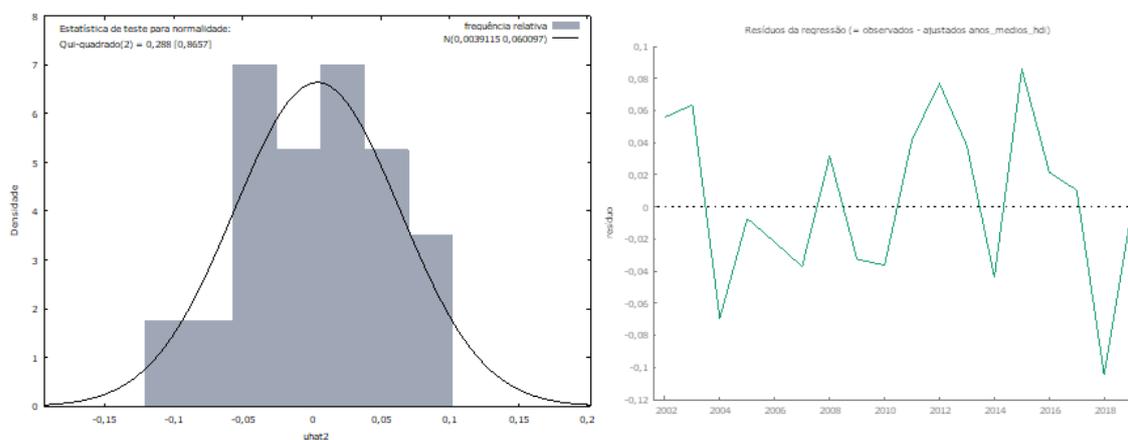
MODELO 1 - ARIMA (2, 1, 2) PARA ANOS MÉDIOS DE ESCOLARIDADE PARA A POPULAÇÃO DE 25 ANOS OU MAIS (SHDI)

ERROS PADRÃO BASEADOS NA HESSIANA					
	<i>Coefficiente</i>	<i>Erro Padrão</i>	<i>z</i>	<i>p-valor</i>	
const	0,126503	0,00212175	59,62	<0,0001	***
phi_1	1,19490	0,215303	5,550	<0,0001	***
phi_2	-0,559138	0,228616	-2,446	0,0145	**
theta_1	-1,97542	0,283735	-6,962	<0,0001	***
theta_2	0,999999	0,283521	3,527	0,0004	***
Média var. dependente	0,125000	D.P. var. dependente		0,074329	
Média de inovações	0,003912	D.P. das inovações		0,051222	
R-quadrado	0,993682	R-quadrado ajustado		0,992329	
Log da verossimilhança	25,02642	Critério de Akaike		-38,05284	
Critério de Schwarz	-32,71061	Critério Hannan-Quinn		-37,31622	
	<i>Real</i>	<i>Imaginária</i>	<i>Módulo</i>	<i>Frequência</i>	
AR					
Raiz 1	1,0685	-0,8042	1,3373	-0,1027	
Raiz 2	1,0685	0,8042	1,3373	0,1027	
MA					
Raiz 1	0,9877	-0,1563	1,0000	-0,0250	
Raiz 2	0,9877	0,1563	1,0000	0,0250	

FONTE: *Software* GRETL

NOTA: Elaboração do autor.

FIGURA 2 - TESTE DE NORMALIDADE DOS RESÍDUOS, GRÁFICO DOS RESÍDUOS E AUTOCORRELAÇÃO



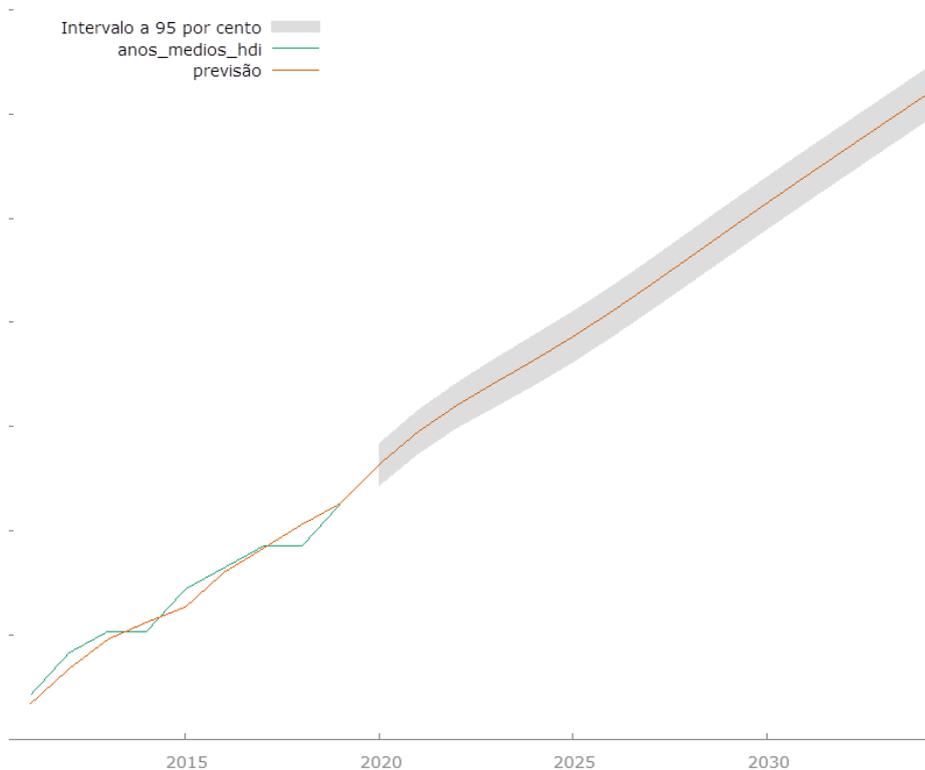
Teste para autocorrelação até a ordem 5

Ljung-Box $Q' = 1,57903$,
com p -valor = $P(\text{Qui-quadrado}(1) > 1,57903) = 0,2089$

FONTE: *Software* GRETL

NOTA: Elaboração do autor.

FIGURA 3 - PREVISÃO PARA OS ANOS MÉDIOS DE ESCOLARIDADE (SHDI) PARA 2020-2035



Para intervalos de confiança de 95%, $z(0,025) = 1,96$

Obs.	anos_medios_hdi	previsão	erro padrão	intervalo a 95%
2020	indefinido	8,31934	0,0512221	(8,21894, 8,41973)
2021	indefinido	8,47546	0,0524413	(8,37268, 8,57824)
2022	indefinido	8,60222	0,0542645	(8,49586, 8,70858)
2023	indefinido	8,71247	0,0584400	(8,59793, 8,82701)
2024	indefinido	8,81941	0,0608245	(8,70019, 8,93862)
2025	indefinido	8,93162	0,0611991	(8,81167, 9,05157)
2026	indefinido	9,05199	0,0612547	(8,93193, 9,17205)
2027	indefinido	9,17915	0,0617954	(9,05803, 9,30027)
2028	indefinido	9,30987	0,0625285	(9,18732, 9,43242)
2029	indefinido	9,44104	0,0630517	(9,31746, 9,56462)
2030	indefinido	9,57077	0,0633005	(9,44670, 9,69484)
2031	indefinido	9,69851	0,0633931	(9,57427, 9,82276)
2032	indefinido	9,82470	0,0634319	(9,70037, 9,94902)
2033	indefinido	9,95012	0,0634632	(9,82574, 10,0745)
2034	indefinido	10,0755	0,0635086	(9,95104, 10,2000)
2035	indefinido	10,2013	0,0635801	(10,0767, 10,3259)

FONTE: Software GRETL

NOTA: Elaboração do autor.

Para a etapa de verificação do diagnóstico, a figura 2 apresenta os resultados do teste de normalidade e de autocorrelação. A hipótese de normalidade dos resíduos não pode ser rejeitada, assim como a hipótese nula de que não há autocorrelação. Ainda que a amostra seja muito reduzida, os resíduos parecem seguir um comportamento de ruído-branco.

A figura 3 apresenta os resultados da previsão feita a partir do modelo para os anos médios de escolaridade considerando pessoas de 25 anos ou mais de idade, entre 2020 e 2035. Seguindo a tendência atual, para 2035 o estado do Paraná sairá de 8,13 anos médios de escolaridade (padrão internacional) para 10,20, considerando um erro-padrão de 0,063, o que resulta em um intervalo entre 10,07 e 10,32. Uma diferença de 2,07 anos em 16 anos de evolução.

Considerando-se os dados de 2019 para 186 países, o avanço estimado para o Paraná seria equivalente a avançar do quinto ao sétimo decil em termos de realização escolar; neste ano, a Alemanha apresentava a melhor performance neste indicador, com média de 14,2 anos de estudo para sua população de 25 e mais anos de idade. O valor esperado para o Paraná em 2035 ainda o situaria, na América do Sul, abaixo da média apresentada por Chile, Argentina e Venezuela em 2019.

O modelo 2 apresentado são os resultados de um modelo ARIMA (2, 1, 2) para uma série ajustada dos anos médios de escolaridade considerando pessoas de 25 anos ou mais de idade, da Pesquisa Nacional por Amostra de Domicílios (PNAD). Esta série, fornecida pelo Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES), compatibiliza os dados da série antiga da PNAD com a recente atualização da metodologia da base de dados, conhecida como PNAD Contínua². A série utilizada no modelo 2 foi também ajustada para o ano de 2010, pois a PNAD não era realizada em anos censitários. Para tanto, através de um ajuste de modelo de regressão, estimou-se o valor para a PNAD em 2010 com base nos dados SHDI.

Erros padrão baseados na Hessiana

MODELO 2 - ARIMA (2, 1, 2) PARA ANOS MÉDIOS DE ESCOLARIDADE PARA A POPULAÇÃO DE 25 ANOS OU MAIS (PNAD AJUSTADA)

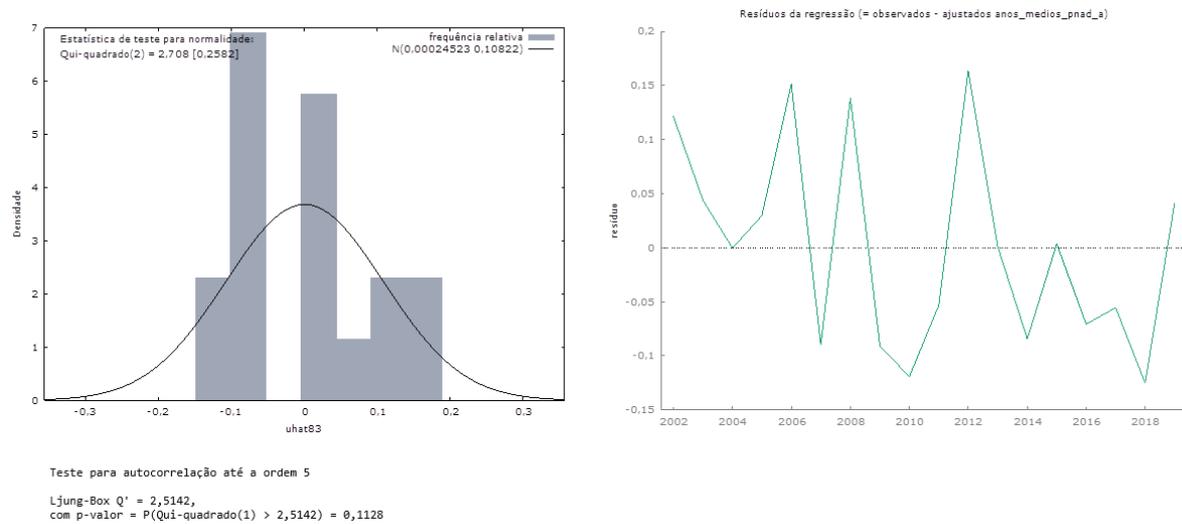
	<i>Coefficiente</i>	<i>Erro Padrão</i>	<i>z</i>	<i>p-valor</i>	
const	0,139395	0,0214882	6,487	<0,0001	***
phi_1	0,861700	0,123550	6,975	<0,0001	***
phi_2	-0,999113	0,00781278	-127,9	<0,0001	***
theta_1	-0,894443	0,264561	-3,381	0,0007	***
theta_2	1,00000	0,315858	3,166	0,0015	***
Média var. dependente	0,145671		D.P. var. dependente	0,124862	
Média de inovações	0,000245		D.P. das inovações	0,091968	
R-quadrado	0,985631		R-quadrado ajustado	0,982552	
Log da verossimilhança	15,38783		Critério de Akaike	-18,77566	
Critério de Schwarz	-13,43343		Critério Hannan-Quinn	-18,03904	
		<i>Real</i>	<i>Imaginária</i>	<i>Módulo</i>	<i>Frequência</i>
AR					
	Raiz 1	0,4312	-0,9027	1,0004	-0,1791
	Raiz 2	0,4312	0,9027	1,0004	0,1791
MA					
	Raiz 1	0,4472	-0,8944	1,0000	-0,1762
	Raiz 2	0,4472	0,8944	1,0000	0,1762

FONTE: Software GRETL

NOTA: Elaboração do autor.

² Mais detalhes no Apêndice A.

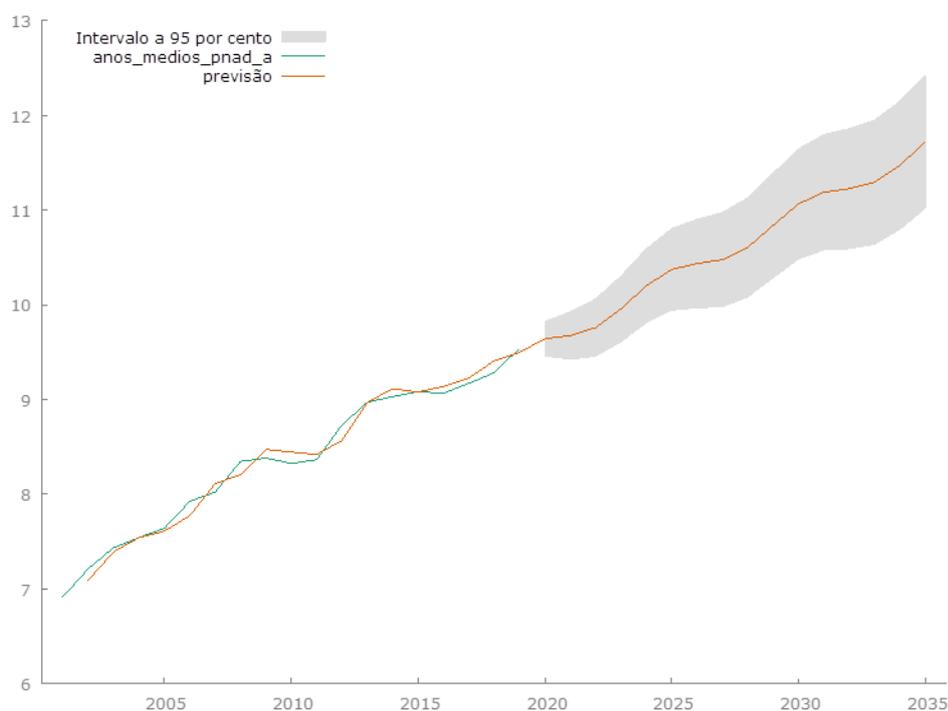
FIGURA 4 - TESTE DE NORMALIDADE DOS RESÍDUOS, GRÁFICO DOS RESÍDUOS E AUTOCORRELAÇÃO



FONTE: *Software* GRETL

NOTA: Elaboração do autor.

FIGURA 5 - PREVISÃO PARA OS ANOS MÉDIOS DE ESCOLARIDADE (PNAD AJUSTADA) PARA 2020 A 2035



Para intervalos de confiança de 95%, $z(0,025) = 1,96$

Obs.	anos_medios_pnad_a	previsão	erro padrão	intervalo a 95%
2020	indefinido	9,64276	0,0919680	(9,46250, 9,82301)
2021	indefinido	9,67661	0,127951	(9,42584, 9,92739)
2022	indefinido	9,76063	0,154415	(9,45798, 10,0633)
2023	indefinido	9,95774	0,177377	(9,61009, 10,3054)
2024	indefinido	10,2022	0,199141	(9,81190, 10,5925)
2025	indefinido	10,3745	0,219564	(9,94414, 10,8048)
2026	indefinido	10,4372	0,237642	(9,97145, 10,9030)
2027	indefinido	10,4777	0,253283	(9,98129, 10,9741)
2028	indefinido	10,6085	0,267598	(10,0840, 11,1330)
2029	indefinido	10,8392	0,281864	(10,2868, 11,3917)
2030	indefinido	11,0660	0,296402	(10,4851, 11,6469)
2031	indefinido	11,1894	0,310421	(10,5810, 11,7978)
2032	indefinido	11,2277	0,323077	(10,5945, 11,8609)
2033	indefinido	11,2960	0,334504	(10,6404, 11,9516)
2034	indefinido	11,4751	0,345610	(10,7977, 12,1525)
2035	indefinido	11,7198	0,357125	(11,0198, 12,4197)

FONTE: Software GRETL

NOTA: Elaboração do autor.

A figura 4 apresenta os testes de normalidade dos resíduos, gráfico dos resíduos e teste de autocorrelação. Assim como no modelo anterior, as hipóteses nulas de normalidade dos resíduos e de não existência de autocorrelação não foram rejeitadas. A figura 5 apresenta os resultados da projeção realizada para a PNAD ajustada. Seguindo a tendência atual, para 2035 o estado do Paraná sairá de 9,54 anos médios de escolaridade (padrão PNAD Contínua) para 11,72, considerando um erro-padrão de 0,36, o que resulta

em um intervalo entre 11,02 e 12,41. Uma diferença de 2,18 em 16 anos. Ressalte-se que, para esta população, em 2019, o Distrito Federal apresentava a maior média de anos de estudo (11,49) entre as unidades da federação, seguido pelo Rio de Janeiro (10,63).

3 ESTATÍSTICAS DESCRITIVAS: VARIÁVEIS UTILIZADAS NO ESTUDO

De acordo com Souza (2010), indicadores permitem estabelecer a dimensão de um fenômeno e analisar a sua evolução no tempo. Para tanto, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) divulga periodicamente microdados educacionais e diversos indicadores no intuito de promover, avaliar e auxiliar a manutenção da educação brasileira, seja ela pública ou privada.

Neste trabalho foram utilizados diversos indicadores, em nível municipal, para o estado do Paraná e somente para escolas públicas, no ano de 2019. O critério para escolha do corte temporal está relacionado ao período mais recente evitando-se os reflexos da pandemia de 2020. Essas variáveis utilizam os indicadores divulgados pelo INEP, mas não necessariamente na mesma escala ou formato. Alguns indicadores, como o Indicador de Nível Socioeconômico, o Índice de Regularidade Docente (IRD) e a remuneração do docente se referem a todas as etapas de ensino. Os demais indicadores estão divididos entre ensino fundamental anos iniciais, ensino fundamental anos finais e ensino médio. A tabela 1 apresenta as variáveis utilizadas neste estudo e as principais estatísticas destas variáveis.

TABELA 1 - ESTATÍSTICAS DAS VARIÁVEIS UTILIZADAS NO ESTUDO

VARIÁVEL	DESCRIÇÃO	MÉDIA	DESVIO-PADRÃO	MIN	MAX
<i>ldeb_fun1</i>	IDEB municipal – Ensino fundamental anos iniciais	6,38	0,64	4,8	8,8
<i>ldeb_fun2</i>	IDEB municipal – Ensino fundamental anos finais	5,46	0,39	4,33	6,57
<i>ldeb_med</i>	IDEB municipal - Ensino Médio	4,37	0,41	3,1	5,8
<i>Taxa_aprov_fun1</i>	Taxa de aprovação – Ensino fundamental anos iniciais	95,29	3,18	80,7	100
<i>Taxa_aprov_fun2</i>	Taxa de aprovação – Ensino fundamental anos finais	93,68	3,79	80,8	100
<i>Taxa_aprov_medio</i>	Taxa de aprovação – Ensino médio	90,49	4,49	78,1	100
<i>mt_ai</i>	Desempenho médio municipal no Saeb – Matemática – Ensino Fundamental Anos Iniciais	243,07	16,76	202,61	308,5
<i>pt_ai</i>	Desempenho médio municipal no Saeb – Língua Portuguesa – Ensino Fundamental Anos Iniciais	225,44	13,97	188,2	281,23
<i>mt_af</i>	Desempenho médio municipal no Saeb – Matemática – Ensino Fundamental Anos Finais	267,57	12,94	232,59	309,86
<i>pt_af</i>	Desempenho médio municipal no Saeb – Língua Portuguesa – Ensino Fundamental Anos Finais	260,58	11,41	223,19	291,82
<i>mt_em</i>	Desempenho médio municipal no Saeb – Matemática – Ensino Médio	279,50	15,46	235,80	325,88
<i>pt_em</i>	Desempenho médio municipal no Saeb – Língua Portuguesa – Ensino Médio	279,16	12,83	229,97	315,79
<i>doc_sem_sup_fun1_media</i>	Variável binária = 1 se o município possui % de docentes sem ensino superior acima de 10% (quartil 75) – Ensino fundamental anos iniciais	-	-	-	-

continua

TABELA 1 - ESTATÍSTICAS DAS VARIÁVEIS UTILIZADAS NO ESTUDO

VARIÁVEL	DESCRIÇÃO	MÉDIA	DESVIO-PADRÃO	conclusão	
				MIN	MAX
<i>doc_sem_sup_fun2_media</i>	Variável binária = 1 se o município possui % de docentes sem ensino superior acima de 3,8% (quartil 90) – Ensino fundamental anos finais	-	-	-	-
<i>doc_sem_sup_med_media</i>	Variável binária = 1 se o município possui % de docentes sem ensino superior acima de 2,9% (quartil 90) – Ensino médio	-	-	-	-
<i>esforco_doc_fun_456</i>	Percentual de professores no município, que lecionam no ensino fundamental, nos níveis de esforço 4, 5, 6 do indicador do INEP	50,33	8,02	22,7	76,9
<i>esforco_doc_med_456</i>	Percentual de professores no município, que lecionam no ensino médio, nos níveis de esforço 4, 5, 6 do indicador do INEP	82,74	10,15	28,5	100
<i>qtd_acima_media_fun1</i>	Variável binária = 1 se o município possui quantidade média de alunos em sala superior à média estadual, no ensino fundamental anos iniciais	-	-	-	-
<i>qtd_acima_media_fun2</i>	Variável binária = 1 se o município possui quantidade média de alunos em sala superior à média estadual, no ensino fundamental anos finais	-	-	-	-
<i>qtd_acima_media_med</i>	Variável binária = 1 se o município possui quantidade média de alunos em sala superior à média estadual, no ensino médio	-	-	-	-
<i>percen_INSE_12</i>	Percentual de alunos no município que estão nos níveis 1 e 2 do Indicador de Nível Socioeconômico do INEP.	7,56	4,84	0	28,87
<i>percen_INSE_1234</i>	Percentual de alunos no município que estão nos níveis 1, 2, 3 e 4 do Indicador de Nível Socioeconômico do INEP.	43,24	10,59	19,64	78,83
<i>perc_IRD_baixo_mediobaixo</i>	Percentual de professores com Índice de Regularidade Docente (IRD) do INEP nos níveis baixo e médio-baixo.	45,63	24,98	0	100
<i>remuneracao_doc_2017</i>	Remuneração média padronizada para 40 horas dos docentes em exercício na educação básica - 2017.	3.763,89	875,54	1522,04	7578,82
<i>remuneracao_doc_mediana</i>	Mediana da remuneração bruta dos docentes em exercício na educação básica - 2017.	2.247,14	453,78	1136,33	4501,31
<i>dist_is_fun1</i>	Taxa de distorção idade-série (proporção de alunos com mais de 2 anos de atraso escolar), ensino fundamental anos iniciais	8,01	4,48	0,5	23,8
<i>dist_is_fun2</i>	Taxa de distorção idade-série (proporção de alunos com mais de 2 anos de atraso escolar), ensino fundamental anos finais	18,64	6,01	5,1	40,7
<i>dist_is_med</i>	Taxa de distorção idade-série (proporção de alunos com mais de 2 anos de atraso escolar), ensino médio	19,63	6,34	5,8	38,8
<i>percent_sem_comp_fun</i>	Percentual municipal de escolas que não possuem microcomputadores, no Ensino Fundamental	2,19	8,26	0	57,14
<i>percent_sem_comp_med</i>	Percentual municipal de escolas que não possuem microcomputadores, no Ensino médio	0,18	2,65	0	50
<i>percent_sem_internet_fun</i>	Percentual municipal de escolas que não possuem acesso à internet, no Ensino Fundamental	3,77	10,59	0	71,43
<i>percent_sem_internet_med</i>	Percentual municipal de escolas que não possuem acesso à internet, no Ensino Médio	0,22	2,77	0	50

FONTE: O autor (2021)

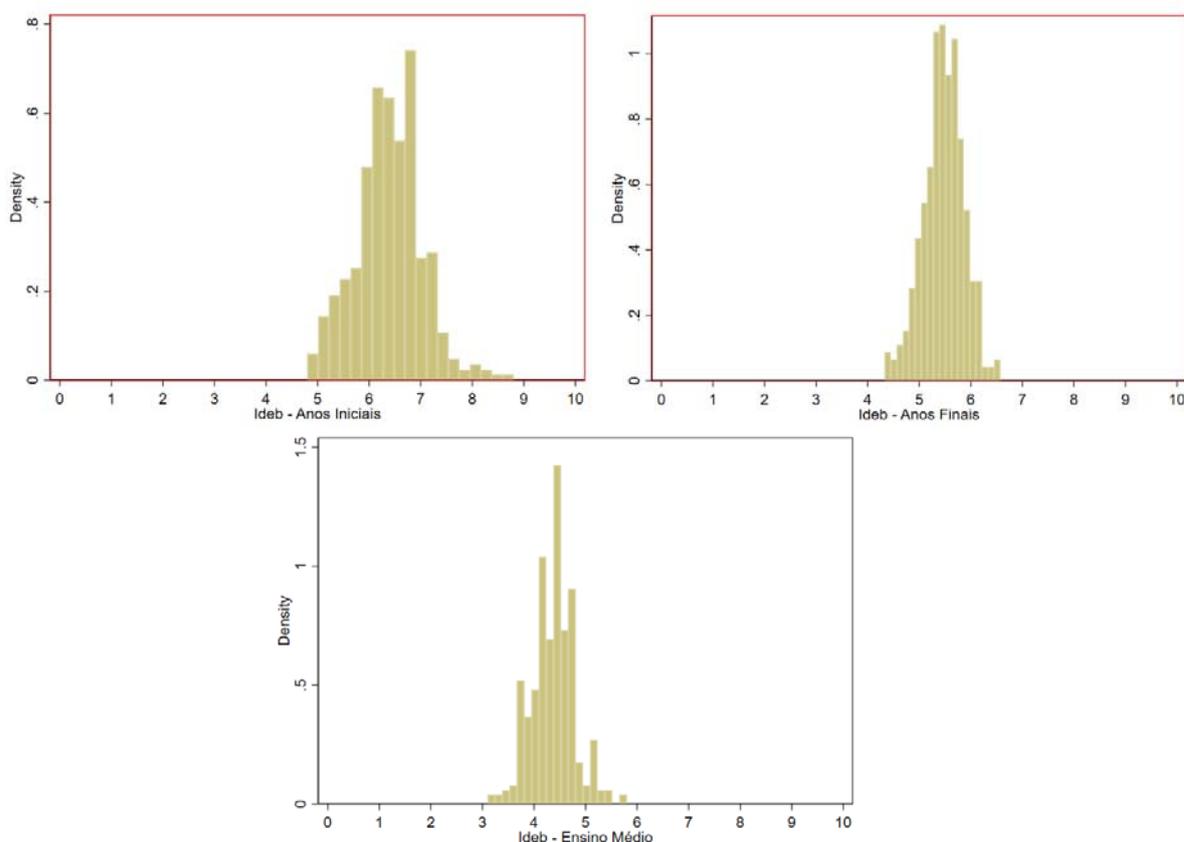
A variável IDEB (Índice de Desenvolvimento da Educação Básica) apresentada na tabela 1 é um índice criado em 2007 pelo Inep com objetivo de medir a qualidade do aprendizado nacional e estabelecer metas para sua melhoria. O IDEB é calculado a partir de dois componentes: a taxa de rendimento escolar (Taxa de Aprovação) e as médias de desempenho nos exames aplicados pelo Inep. Os índices de aprovação são obtidos a partir do Censo Escolar, realizado anualmente. Neste trabalho serão utilizados tanto o IDEB quanto a Taxa de Aprovação e as médias de desempenho, para mensurar a influência das variáveis educacionais.

As médias de desempenho utilizadas no IDEB são as da Prova Brasil e do Sistema de Avaliação da Educação Básica (Saeb), realizados a cada dois anos. O IDEB utiliza uma escala de 0 a 10, onde as metas estabelecidas são diferenciadas para cada escola e rede de ensino, com o objetivo único de alcançar 6 pontos até 2022, patamar de países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) como Estados Unidos, Canadá, Inglaterra e Suécia.

De acordo com as estatísticas apresentadas na tabela 1 é possível ver que o nível médio do IDEB é inferior quanto mais alto o nível escolar, sendo o menor no ensino médio. A média do IDEB para os anos iniciais do ensino fundamental é de 6,38. A média para os anos finais do ensino fundamental e para o ensino médio são de 5,46 e 4,37, respectivamente. Essa tendência se confirma também nas taxas de aprovação e na distorção idade-série. Essa evidência demonstra a importância da atenção ao ensino médio, não só em termos do desempenho, mas também da evasão escolar causada pelo alto custo de oportunidade, especialmente para aqueles alunos oriundos de uma situação socioeconômica fragilizada.

Apesar das diferenças em termos das médias do IDEB, a figura 6 mostra que a distribuição das notas é mais assimétrica à direita no ensino fundamental anos iniciais e mais assimétrica à esquerda no ensino fundamental anos finais. Ainda que com uma média maior, a variabilidade nos anos iniciais do ensino fundamental é a maior entre os três níveis educacionais, com um desvio-padrão de 0,64.

FIGURA 6 - HISTOGRAMA IDEB - ENSINO FUNDAMENTAL (ANOS INICIAIS E FINAIS) E MÉDIO



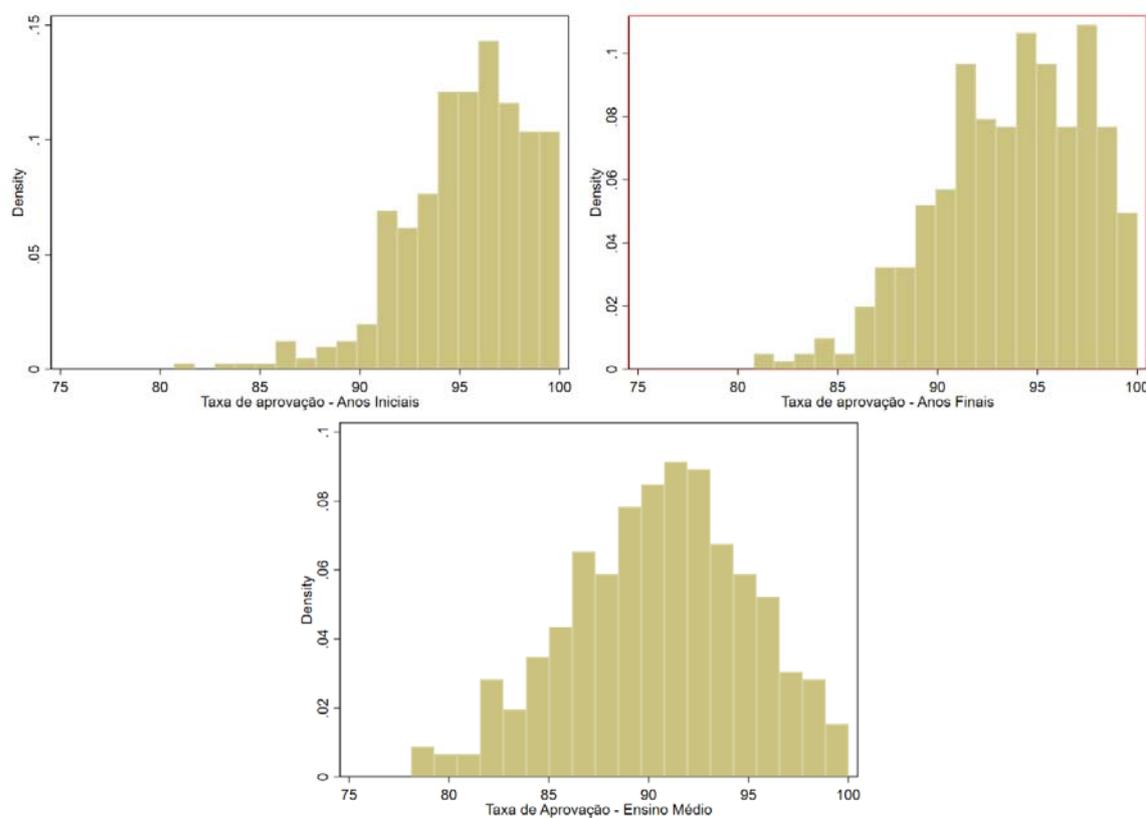
FONTE: *Software* STATA

NOTA: Elaboração do autor.

No que concerne às Taxas de Aprovação, na figura 7 é possível observar que a variabilidade maior se encontra no ensino médio, com uma distribuição fortemente simétrica. Para o ensino fundamental, tanto nos anos iniciais quanto finais, os histogramas apresentam formatos bastante assimétricos à esquerda, indicando que a maior parte dos municípios encontram-se em níveis mais altos de aprovação nestas etapas de ensino.

Ainda assim, os níveis mínimos de aprovação para as três etapas de ensino encontram-se em torno de 80%, com média de 93%. Sendo a taxa de aprovação um componente importante do IDEB e estando muito próxima de seu nível máximo, o desempenho no Saeb se destaca como importante aspecto a ser trabalhado para aumentos futuros do IDEB das escolas.

FIGURA 7 - HISTOGRAMA TAXA DE APROVAÇÃO – ENSINO FUNDAMENTAL (ANOS INICIAIS E FINAIS) E MÉDIO



FONTE: *Software STATA*

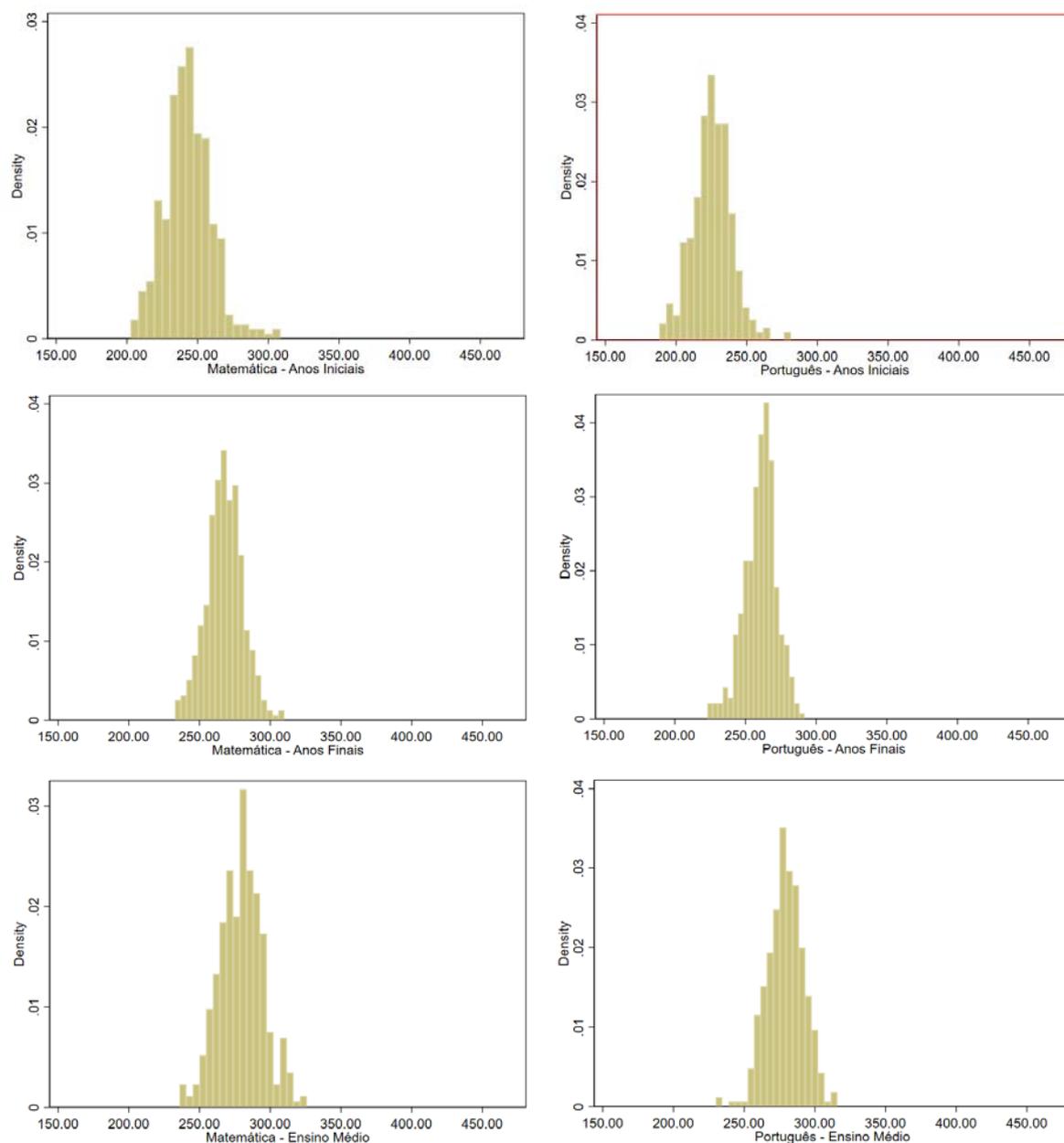
NOTA: Elaboração do autor.

As distribuições do desempenho municipal em Português e Matemática para as três etapas de ensino podem ser vistas na figura 8. As distribuições com maior variação são as distribuições de Matemática nos anos iniciais do ensino fundamental e Português no Ensino Médio. Seguindo os níveis de proficiência média e o padrão de desempenho apresentado no quadro 1, nos anos iniciais do ensino fundamental o desempenho em Português e Matemática encontra-se no nível Proficiente. Para os anos finais do ensino fundamental, o desempenho em ambas as disciplinas encontra-se no nível Básico.

Para o ensino médio, estando no nível 3 da escala Saeb³, ambas as disciplinas podem ser consideradas no nível insuficiente, entre 275 e 300 pontos.

³ Não apresentada no quadro.

FIGURA 8 - HISTOGRAMA DESEMPENHO SAEB – PORTUGUÊS E MATEMÁTICA, ENSINO FUNDAMENTAL (ANOS INICIAIS E FINAIS) E MÉDIO



FONTE: *Software* STATA

NOTA: Elaboração do autor.

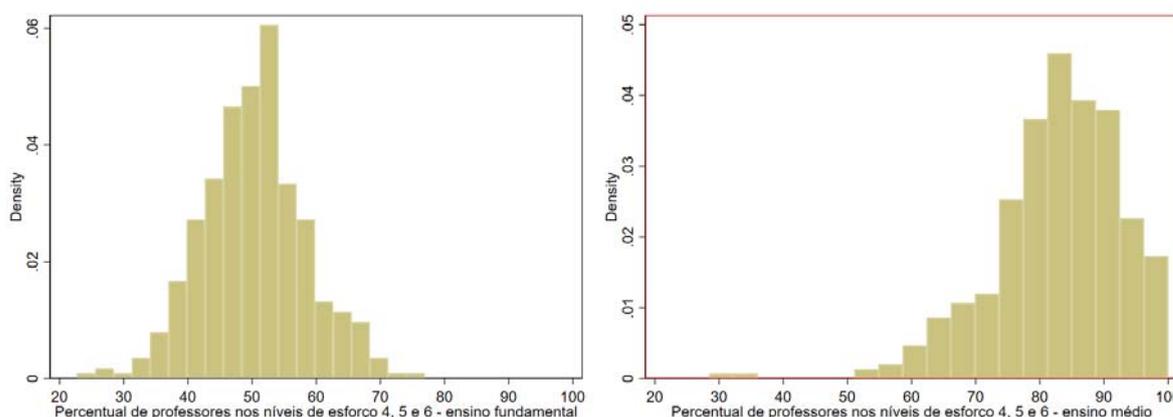
QUADRO 1 - NÍVEL DE PROFICIÊNCIA MÉDIA E PADRÃO DE DESEMPENHO NA ESCALA SAEB

NÍVEL DE APRENDIZADO	5.º ANO		9.º ANO	
	Português	Matemática	Português	Matemática
Insuficiente	Até 150	0 a 174	Até 200	Até 225
Básico	150 a 200	175 a 224	200 a 275	225 a 300
Proficiente	200 a 250	225 a 274	275 a 325	300 a 350
Avançado	Acima de 250	Igual ou maior que 275	Acima de 325	Acima de 350

FONTE: Foco Brasil, Inep, 2021.

A figura 9 apresenta a distribuição dos professores que estão no nível de esforço docente 4, 5 e 6 nos municípios. Nível de esforço aqui é uma *proxy* para sobrecarga dos docentes. Professores no nível 4 são aqueles que, em geral, têm entre 50 e 400 alunos e atuam em dois turnos, em uma ou duas escolas e em duas etapas. Professores no nível 5 têm mais de 300 alunos e atuam nos três turnos, em duas ou três escolas e em duas etapas ou três etapas. Já os professores no nível 6 têm mais de 400 alunos e atuam nos três turnos, em duas ou três escolas e em duas etapas ou três etapas. O histograma mostra que o nível de esforço é maior no ensino médio do que no ensino fundamental, tendo naquela etapa de ensino, uma distribuição fortemente assimétrica à esquerda.

FIGURA 9 -HISTOGRAMA PERCENTUAL DE PROFESSORES NO MUNICÍPIO, QUE LECIONAM NO ENSINO FUNDAMENTAL E MÉDIO, NOS NÍVEIS DE ESFORÇO 4, 5, 6 DO INDICADOR DO INEP



FONTE: *Software STATA*

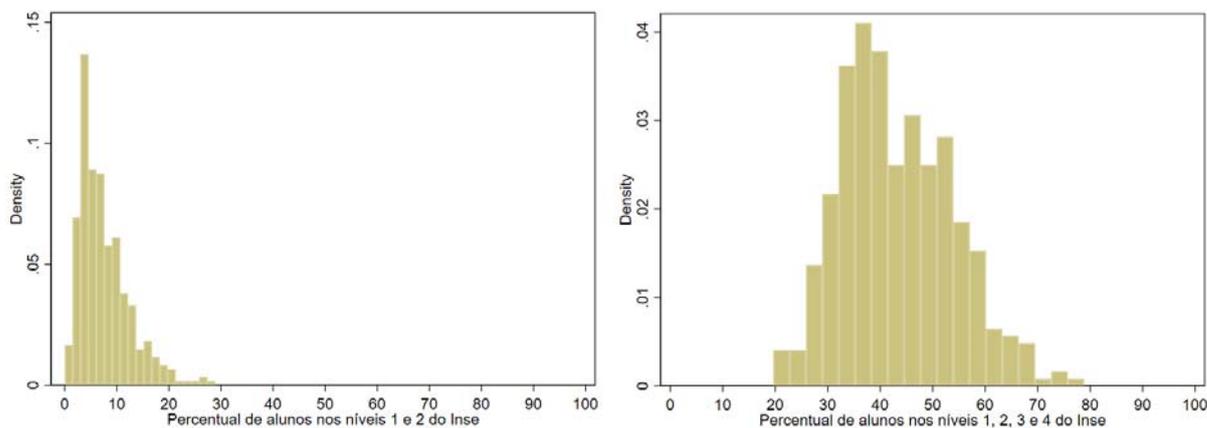
NOTA: Elaboração do autor.

O Indicador de Nível Socioeconômico (INSE) foi construído pela Diretoria de Avaliação da Educação Básica (DAEB) com o objetivo de contextualizar resultados obtidos em avaliações e exames aplicados por este instituto na educação básica. Em 2019 o questionário do INSE passou por revisão, aumentando o número de itens a serem utilizados para medir aspectos socioeconômicos além de ter sido psicometricamente testado. O indicador é a combinação dos elementos escolaridade dos pais e posse de bens e serviços e vale para todos os níveis de ensino de maneira agregada (BRASIL, 2021).

A figura 10 apresenta o histograma para o percentual de alunos nos níveis 1, 2 (primeiro histograma) e 1, 2, 3 e 4 (segundo histograma) no INSE, por município. O nível 1 do INSE é o nível inferior da escala, no qual os estudantes têm dois ou mais desvios-padrão abaixo da média nacional. Considerando a maioria dos estudantes, o pai/responsável não completou o 5º ano do ensino fundamental e a mãe/responsável tem o 5º ano do ensino

fundamental incompleto ou completo. A maioria dos estudantes deste nível possui uma geladeira, um ou dois quartos, uma televisão e um banheiro. Mas não possui muitos dos bens e serviços pesquisados (computador, carro, *wi-fi*, mesa para estudar, garagem, microondas, aspirador de pó, máquina de lavar roupa e freezer).

FIGURA 10 - HISTOGRAMA PERCENTUAL DE ALUNOS DOS MUNICÍPIOS QUE ESTÃO NOS NÍVEIS 1 E 2 E 1, 2, 3 E 4 DO INDICADOR DE NÍVEL SOCIOECONÔMICO DO INEP



FONTE: *Software STATA*

NOTA: Elaboração do autor.

No nível 2, os estudantes estão entre um e dois desvios-padrão abaixo da média nacional. Considerando a maioria dos estudantes, a mãe/responsável e/ou o pai/responsável tem o 5º ano do ensino fundamental incompleto ou completo. A maioria possui uma geladeira, um ou dois quartos, uma televisão e um banheiro. Mas não possui muitos dos bens e serviços pesquisados, exceto uma parte dos estudantes deste nível passa a ter freezer, máquina de lavar roupa e três ou mais quartos para dormir em sua casa.

No nível 3 os estudantes estão entre meio e um desvio-padrão abaixo da média nacional. Considerando a maioria dos estudantes, a mãe/responsável e o pai/responsável têm o ensino fundamental incompleto ou completo e/ou ensino médio completo. A maioria possui uma geladeira, um ou dois quartos, uma televisão, um banheiro, *wi-fi* e máquina de lavar roupas, mas não possui computador, carro, garagem e aspirador de pó. Parte dos estudantes passa a ter também freezer e forno de micro-ondas.

Por fim, no nível 4 os estudantes estão até meio desvio-padrão abaixo da média nacional. Considerando a maioria dos estudantes, a mãe/responsável e o pai/responsável têm o ensino fundamental incompleto ou completo e/ou ensino médio completo. A maioria possui uma geladeira, um ou dois quartos, um banheiro, *wi-fi*, máquina de lavar roupas e freezer, mas não possui aspirador de pó. Parte dos estudantes deste nível passa a ter também computador, carro, mesa de estudos, garagem, forno de micro-ondas e uma ou duas televisões.

A partir das descritivas apresentadas na tabela 1, 43,24% dos alunos em média estão na metade inferior na escala do INSE no Paraná. E 7,56% em média estão abaixo de dois desvios-padrão da média do INSE, sendo a condição socioeconômica uma importante restrição no ensino, não só em termos do acesso à educação de qualidade, mas no que concerne ao custo-de-oportunidade de continuar na escola, refletindo nas altas taxas de evasão, principalmente no ensino médio.

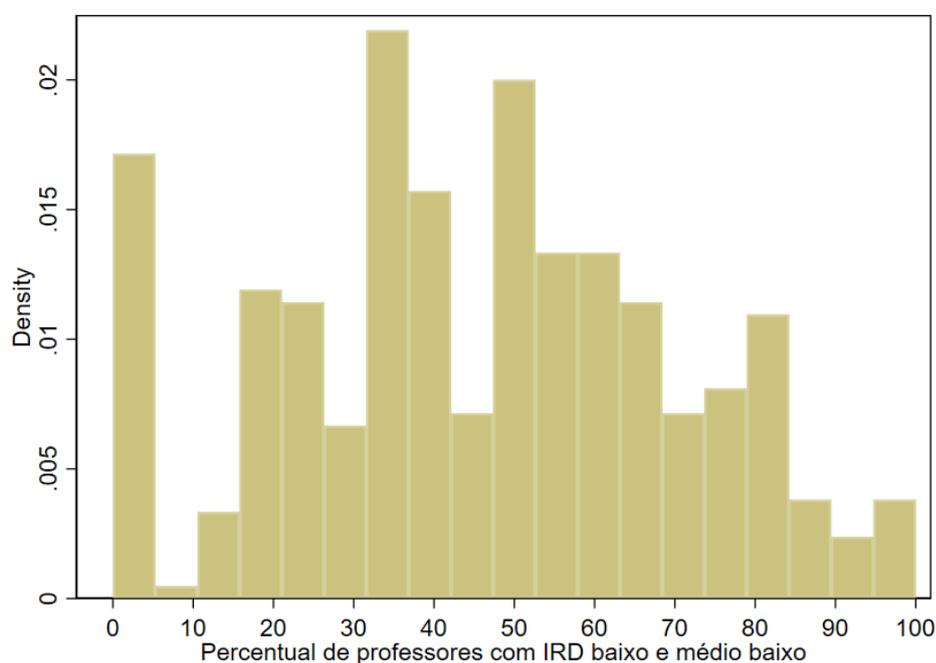
A distribuição, segundo a figura 10, tem uma grande variabilidade, principalmente quando se trata dos níveis 1, 2, 3 e 4 no INSE. Analisando por quartis, metade dos municípios da amostra tem até 41,74% dos alunos nos níveis 1, 2, 3 e 4. Analisando os níveis socioeconômicos mais baixos, 1 e 2, 90% dos municípios tem até 13,95% dos alunos nesse nível, ou seja, apenas 10% dos municípios têm um percentual entre 13,95% e 28,87% nos níveis socioeconômicos mais baixos.

A figura 11 apresenta o Indicador de Regularidade do Docente (IRD) através do Percentual de professores com IRD nos níveis baixo e médio-baixo no município. Esse indicador tem por objetivo avaliar a regularidade do corpo docente nas escolas a partir da permanência dos professores nas escolas nos últimos cinco anos (2015 a 2019).

O IRD leva em conta o total de anos em que o docente atuou na escola nos últimos 5 anos, a atuação do docente na escola em anos mais recentes e a atuação em anos consecutivos. Varia de 0 a 5, quanto mais próximo de 0, mais irregular é o vínculo do docente com a escola e quanto mais próximo de 5, mais regular é esse vínculo. Em média, 45,63% dos docentes nos municípios apresentam baixo nível de regularidade. Esse baixo nível pode ser reflexo de condições de trabalho e renda. O histograma 11, contudo, mostra que essa distribuição tem uma grande variabilidade, como pode ser visto no desvio-padrão de 24,98.

A distribuição da remuneração docente é apresentada na figura 12, tanto com relação à remuneração bruta padronizada para 40 horas como também para as medianas da remuneração dos professores. É possível perceber uma distribuição assimétrica à direita, indicando a concentração dos salários na parte inferior da escala. A remuneração média docente bruta, padronizada para 40 horas, no Paraná, é de R\$ 3.763,89. No entanto, dado o alto desvio-padrão, a média da mediana salarial é bem inferior, de R\$ 2.247,14.

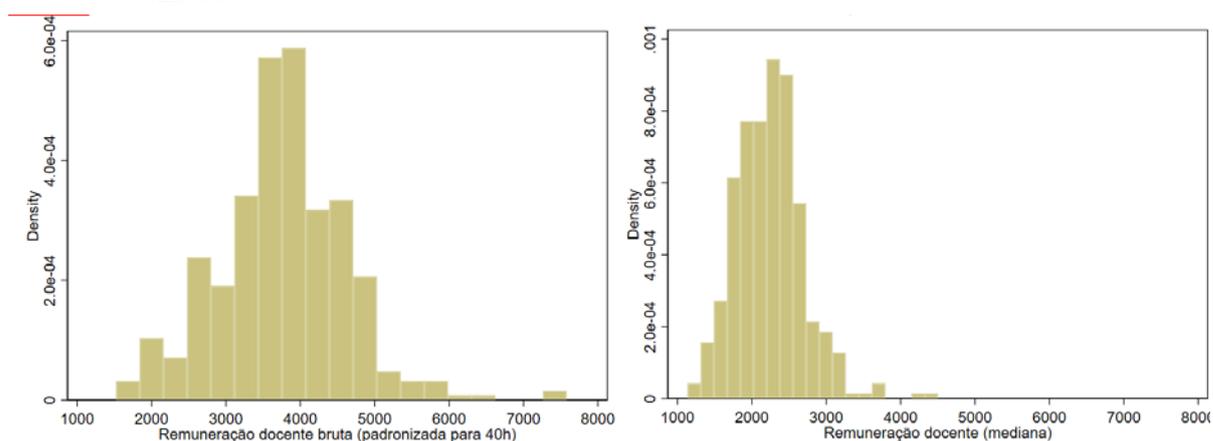
FIGURA 11 - HISTOGRAMA PERCENTUAL DE PROFESSORES COM ÍNDICE DE REGULARIDADE DOCENTE (IRD) DO INEP NOS NÍVEIS BAIXO E MÉDIO-BAIXO



FONTE: *Software STATA*

NOTA: Elaboração do autor.

FIGURA 12 - HISTOGRAMA REMUNERAÇÃO DOCENTE – BRUTA PADRONIZADA PARA 40 HORAS SEMANAIS E MEDIANA



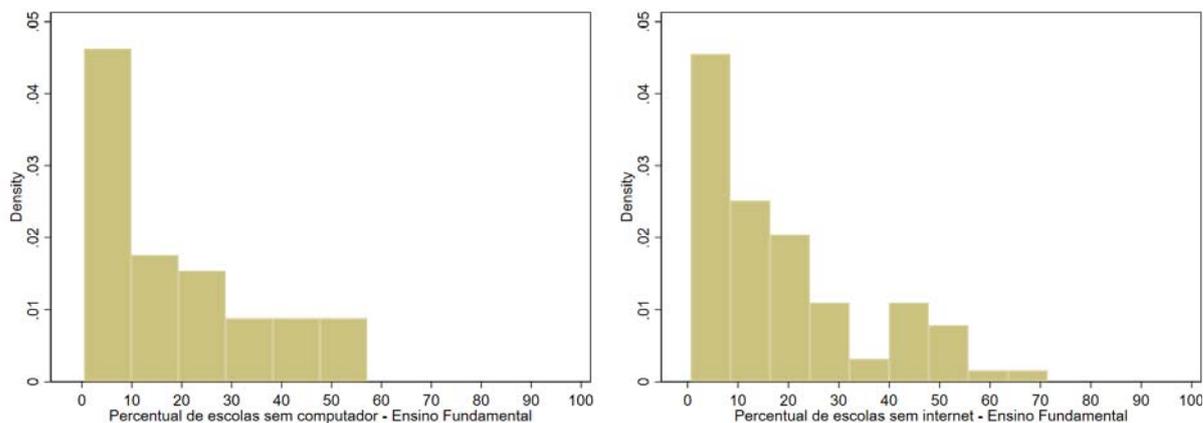
FONTE: *Software STATA*

NOTA: Elaboração do autor.

A figura 13 apresenta o histograma do percentual de escolas sem computador e sem acesso à internet. Para facilitar a visualização, dada a quantidade de municípios com valores zero atribuídos, esses foram excluídos do histograma para o ensino fundamental. É possível notar que, ainda que grande parte dos municípios tenham a totalidade de escolas

com acesso a computador e internet – 351 municípios para acesso a computador e 318 para acesso à internet – 10% dos municípios tem pelo menos 2,38% das escolas sem computador e 13,33% sem acesso à internet, no ensino fundamental.

FIGURA 13 - HISTOGRAMA PERCENTUAL DE ESCOLAS SEM COMPUTADOR E SEM ACESSO À INTERNET (ENSINO FUNDAMENTAL), EXCEPTUANDO OS ZEROS

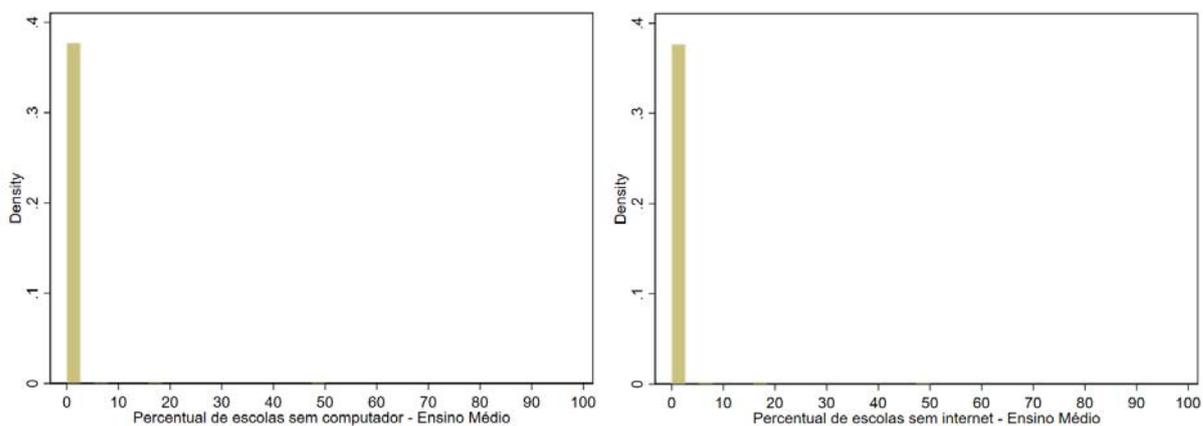


FORNTE: *Software STATA*

NOTA: Elaboração do autor.

Para o ensino médio, figura 14, optou-se por não se excluir os zeros, uma vez que apenas três municípios não apresentaram 100% das escolas com computadores e apenas cinco não apresentaram 100% das escolas com acesso à internet.

FIGURA 14 - HISTOGRAMA PERCENTUAL DE ESCOLAS SEM COMPUTADOR E SEM ACESSO À INTERNET (ENSINO MÉDIO)



FORNTE: *Software STATA*

NOTA: Elaboração do autor.

Entre ensino fundamental e médio, existem ainda municípios com mais de 50% de escolas sem computadores (municípios de Boa Vista da Aparecida, Pitanga, Honório Serpa e Engenheiro Beltrão) e mais de 50% de escolas sem acesso à internet (municípios de Boa Vista da Aparecida, Rio Negro, Nossa Senhora das Graças, Quatiguá, Candói, Honório Serpa e Engenheiro Beltrão).

4 MODELOS DE REGRESSÃO: RESULTADOS PRELIMINARES

Modelos exploratórios têm a função de avaliar as relações entre variáveis independentes junto à variável dependente, sua significância estatística e seus sinais em comparação com os sinais esperados. O modelo 3 apresenta os resultados de um modelo de regressão exploratório para a variável dependente IDEB (ensino fundamental anos iniciais). Analisando-se os resultados é possível observar que há uma relação inversa e estatisticamente significativa entre o percentual de professores sem ensino superior e o IDEB. Maiores taxas de distorção idade-série estão associadas a um IDEB menor. Um maior percentual de escolas sem computadores está associado a um IDEB menor. As demais variáveis não são significantes a 10%.

MODELO 3 - MODELO DE REGRESSÃO EXPLORATÓRIO - IDEB - FUNDAMENTAL ANOS INICIAIS

Source	SS	df	MS	Number of obs	=	393
Model	23.4245086	9	2.60272317	F(9, 383)	=	7.17
Residual	139.074779	383	.363119527	Prob > F	=	0.0000
				R-squared	=	0.1442
				Adj R-squared	=	0.1240
Total	162.499288	392	.414538999	Root MSE	=	.60259

IDEb_fun1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
doc_sem_sup_fun1_media	-.1344046	.0706603	-1.90	0.058	-.2733353 .0045261
esforco_doc_fun_456	-.0026552	.0038883	-0.68	0.495	-.0103003 .0049899
qtd_acima_media_fun1	.079412	.0624176	1.27	0.204	-.0433121 .2021361
percen_INSE_1234	.0036038	.0030441	1.18	0.237	-.0023814 .009589
perc_IRD_baixo_mediobaixo	-.0018014	.0012381	-1.45	0.146	-.0042357 .0006329
remuneracao_doc_2017	.0000123	.000035	0.35	0.724	-.0000565 .0000811
dist_is_fun1	-.0484949	.0069333	-6.99	0.000	-.062127 -.0348629
percent_sem_comp_fun	-.01729	.0065459	-2.64	0.009	-.0301604 -.0044197
percent_sem_internet_fun	.0083421	.0052933	1.58	0.116	-.0020656 .0187497
_cons	6.794394	.2935636	23.14	0.000	6.217196 7.371592

FONTE: *Software* STATA

NOTA: Elaboração do autor.

Para o modelo 4, variável dependente IDEB (ensino fundamental anos finais), é possível observar uma associação negativa entre percentual de professores sem ensino superior e a nota do IDEB. Para os anos finais, escolas com índice de regularidade baixo estão associadas a maiores notas do IDEB. Maiores taxas de distorção idade-série estão associadas a um pior desempenho do IDEB. Municípios com maior percentual de escolas sem computadores estão associados a um IDEB menor.

MODELO 4 - MODELO DE REGRESSÃO EXPLORATÓRIO – IDEB – FUNDAMENTAL ANOS FINAIS

Source	SS	df	MS	Number of obs	=	383
Model	4.73897728	9	.526553032	F(9, 373)	=	3.71
Residual	52.9260081	373	.141892783	Prob > F	=	0.0002
				R-squared	=	0.0822
				Adj R-squared	=	0.0600
Total	57.6649854	382	.150955459	Root MSE	=	.37669

IDEb_fun2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
doc_sem_sup_fun2_media	-.2311719	.0676872	-3.42	0.001	-.3642681	-.0980756
esforco_doc_fun_456	-.003199	.0024935	-1.28	0.200	-.0081019	.001704
qtd_acima_media_fun2	.0446663	.0720297	0.62	0.536	-.0969688	.1863014
percen_INSE_1234	-.0011159	.0019322	-0.58	0.564	-.0049152	.0026834
perc_IRD_baixo_mediobaixo	.0019231	.0007779	2.47	0.014	.0003936	.0034527
remuneracao_doc_2017	.0000298	.0000222	1.34	0.181	-.0000139	.0000735
dist_is_fun2	-.0099011	.003371	-2.94	0.004	-.0165297	-.0032725
percent_sem_comp_fun	-.0068597	.0039554	-1.73	0.084	-.0146374	.000918
percent_sem_internet_fun	.0022412	.0031895	0.70	0.483	-.0040304	.0085127
_cons	5.693676	.1934792	29.43	0.000	5.31323	6.074123

FONTE: *Software* STATA

NOTA: Elaboração do autor.

Para o modelo 5, variável dependente IDEB (ensino médio), municípios com maior quantidade de escolas com quantidade de alunos acima da média por sala estão associados a um IDEB menor. Municípios com mais alunos nos níveis 1, 2, 3 e 4 do Índice de Nível Socioeconômico do INEP (INSE) estão associados a um pior desempenho no IDEB.

MODELO 5 - MODELO DE REGRESSÃO EXPLORATÓRIO - IDEB - ENSINO MÉDIO

Source	SS	df	MS	Number of obs	=	363
Model	6.04842709	9	.672047454	F(9, 353)	=	4.28
Residual	55.4481569	353	.157076932	Prob > F	=	0.0000
				R-squared	=	0.0984
				Adj R-squared	=	0.0754
Total	61.496584	362	.169880066	Root MSE	=	.39633

IDEb_med	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
doc_sem_sup_med_media	-.0250002	.0747922	-0.33	0.738	-.1720945	.1220941
esforco_doc_med_456	-.0006032	.0020975	-0.29	0.774	-.0047284	.003522
qtd_acima_media_med	-.0713535	.04299	-1.66	0.098	-.1559022	.0131952
percen_INSE_1234	-.0110888	.0020385	-5.44	0.000	-.015098	-.0070797
perc_IRD_baixo_mediobaixo	-.000376	.0008633	-0.44	0.663	-.0020739	.0013219
remuneracao_doc_2017	9.14e-06	.0000245	0.37	0.709	-.0000391	.0000573
dist_is_med	.0035414	.0033626	1.05	0.293	-.0030718	.0101547
percent_sem_comp_med	.0050375	.0228269	0.22	0.825	-.0398562	.0499313
percent_sem_internet_med	-.0103644	.0217908	-0.48	0.635	-.0532205	.0324917
_cons	4.85664	.2494164	19.47	0.000	4.366111	5.347169

FONTE: *Software* STATA

NOTA: Elaboração do autor.

Para o modelo 6, variável dependente taxa de aprovação (ensino fundamental anos iniciais), municípios com maior percentual de alunos nos níveis 1, 2, 3 e 4 do INSE estão associados a maiores taxas de aprovação. Municípios com maior percentual de professores com Índice de Regularidade do Docente (IRD) baixo e médio-baixo estão associados a menores taxas de aprovação. Municípios com maiores taxas de distorção idade-série estão associados a menores taxas de aprovação.

MODELO 6 - MODELO DE REGRESSÃO EXPLORATÓRIO - TAXA DE APROVAÇÃO - FUNDAMENTAL ANOS INICIAIS

Source	SS	df	MS	Number of obs	=	395
Model	1792.5164	9	199.168489	F(9, 385)	=	34.09
Residual	2249.01054	385	5.84158582	Prob > F	=	0.0000
				R-squared	=	0.4435
				Adj R-squared	=	0.4305
Total	4041.52694	394	10.2576826	Root MSE	=	2.4169

taxa_aprov_fun1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
doc_sem_sup_fun1_media	-.368999	.2829859	-1.30	0.193	-.9253903 .1873922
esforco_doc_fun_456	.0220063	.0155115	1.42	0.157	-.0084917 .0525042
qtd_acima_media_fun1	.2583707	.2496527	1.03	0.301	-.2324828 .7492241
percen_INSE_1234	.0243757	.0122051	2.00	0.047	.0003787 .0483728
perc_IRD_baixo_mediobaixo	-.0112115	.0049648	-2.26	0.024	-.020973 -.00145
remuneracao_doc_2017	-.0002027	.0001403	-1.44	0.149	-.0004785 .0000731
dist_is_fun1	-.4662101	.0276187	-16.88	0.000	-.5205123 -.4119078
percent_sem_comp_fun	.0191685	.0252632	0.76	0.448	-.0305027 .0688397
percent_sem_internet_fun	-.0148616	.0201535	-0.74	0.461	-.0544863 .0247631
_cons	98.1279	1.171709	83.75	0.000	95.82415 100.4316

FONTE: *Software STATA*

NOTA: Elaboração do autor.

O modelo 7 apresenta os resultados para a variável dependente taxa de aprovação (ensino fundamental anos finais). Municípios com maior percentual de professores nos níveis de esforço 4, 5 e 6 estão associados a maiores taxas de aprovação. Municípios com maior percentual de escolas com quantidade de alunos acima da média por sala estão associados a menores taxas de aprovação. Municípios com maior percentual de professores com Índice de Regularidade do Docente (IRD) baixo e médio-baixo estão associados a menores taxas de aprovação. Maiores taxas de distorção idade-série estão também associadas a menores taxas de aprovação.

Por fim, para o modelo 8, variável dependente taxa de aprovação (ensino médio), municípios com maior percentual de escolas com quantidade de alunos acima da média por sala estão associados a menores taxas de aprovação. Municípios com maiores taxas de distorção idade-série estão associados a menores taxas de aprovação.

MODELO 7 - MODELO DE REGRESSÃO EXPLORATÓRIO – TAXA DE APROVAÇÃO – FUNDAMENTAL ANOS FINAIS

Source	SS	df	MS	Number of obs	=	393
Model	1623.36725	9	180.374138	F(9, 383)	=	17.13
Residual	4032.98237	383	10.5299801	Prob > F	=	0.0000
				R-squared	=	0.2870
				Adj R-squared	=	0.2702
Total	5656.34962	392	14.4294633	Root MSE	=	3.245

taxa_aprov_fun2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
doc_sem_sup_fun2_media	.2878869	.5604959	0.51	0.608	-.8141474	1.389921
esforco_doc_fun_456	.0445886	.0211406	2.11	0.036	.0030225	.0861548
qtd_acima_media_fun2	-1.082155	.6198572	-1.75	0.082	-2.300904	.1365939
percen_INSE_1234	.0142817	.0164575	0.87	0.386	-.0180766	.04664
perc_IRD_baixo_mediobaixo	-.011309	.006613	-1.71	0.088	-.0243113	.0016933
remuneracao_doc_2017	-.0000778	.0001884	-0.41	0.680	-.0004483	.0002927
dist_is_fun2	-.3136812	.028103	-11.16	0.000	-.3689367	-.2584257
percent_sem_comp_fun	-.0235046	.0340277	-0.69	0.490	-.0904091	.0434
percent_sem_internet_fun	.0177231	.0270738	0.65	0.513	-.0355089	.0709551
_cons	97.46204	1.64588	59.22	0.000	94.22594	100.6981

FONTE: Software STATA

NOTA: Elaboração do autor.

MODELO 8 - MODELO DE REGRESSÃO EXPLORATÓRIO - TAXA DE APROVAÇÃO - ENSINO MÉDIO

Source	SS	df	MS	Number of obs	=	395
Model	1843.68531	9	204.853924	F(9, 385)	=	12.79
Residual	6165.58648	385	16.0145103	Prob > F	=	0.0000
				R-squared	=	0.2302
				Adj R-squared	=	0.2122
Total	8009.2718	394	20.328101	Root MSE	=	4.0018

taxa_aprov_medio	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
doc_sem_sup_med_media	-.3182949	.7145659	-0.45	0.656	-1.723235	1.086645
esforco_doc_med_456	.0134148	.0204974	0.65	0.513	-.0268861	.0537156
qtd_acima_media_med	-2.175745	.416909	-5.22	0.000	-2.995449	-1.356042
percen_INSE_1234	-.0024803	.0195076	-0.13	0.899	-.0408351	.0358745
perc_IRD_baixo_mediobaixo	-.0117512	.0083306	-1.41	0.159	-.0281305	.004628
remuneracao_doc_2017	-.0000819	.000234	-0.35	0.727	-.000542	.0003782
dist_is_med	-.2662342	.0325295	-8.18	0.000	-.330192	-.2022765
percent_sem_comp_med	-.3057607	.2300565	-1.33	0.185	-.758085	.1465637
percent_sem_internet_med	.3521255	.2195866	1.60	0.110	-.0796135	.7838645
_cons	96.70601	2.431543	39.77	0.000	91.92524	101.4868

FONTE: Software STATA

NOTA: Elaboração do autor.

A seção seguinte apresentará uma estimativa para esses modelos desconsiderando as variáveis estatisticamente não-significantes.

5 MENSURANDO A INFLUÊNCIA DE INDICADORES EDUCACIONAIS: IDEB E A TAXA DE APROVAÇÃO

Esta seção tem como objetivo apresentar os resultados dos modelos finais considerando apenas as variáveis estatisticamente significantes para as variáveis dependentes IDEB e Taxa de Aprovação, bem como, interpretar os resultados estimados. Os testes diagnósticos que os validam podem ser encontrados no Apêndice B. Para esses modelos, serão utilizadas as mesmas variáveis dependentes da seção 4, após um processo de transformação logarítmica. O uso do logaritmo natural da variável dependente tem diversas utilidades em trabalhos de econometria aplicados. De acordo com Wooldridge (2019, p.206):

[...] quando $y > 0$, os modelos que usam $\log(y)$ como a variável dependente em geral satisfazem às hipóteses do MLC⁴ mais apropriadamente do que os modelos que usam o nível y . Variáveis estritamente positivas frequentemente possuem distribuições condicionais que são heterocedásticas ou concentradas; o uso do log pode aliviar, se não eliminar, ambos os problemas.

Além disso, modelos *log-lin*, como são conhecidos os modelos com variáveis dependentes logaritmizadas, mas com variáveis independentes em nível, têm a conveniência de terem seus parâmetros interpretados como percentuais. Portanto, os modelos apresentados nesta seção serão analisados segundo dois critérios (1) significância estatística e; (2) significância da magnitude. O intuito é mensurar, dentre as variáveis independentes, quais possuem maior importância relativa sobre o rendimento e o desempenho escolar.

Os modelos 9, 10 e 11, na seção 5.1, são os modelos que utilizam como variável dependente a nota municipal do IDEB, que inclui o rendimento dos alunos (taxa de aprovação) e o desempenho em termos das notas em português e matemática. Os modelos 12, 13 e 14, da seção 5.2, apresentam os resultados para a variável dependente Taxa de Aprovação, de modo a compreender as diferenças sobre esta variável sem a influência do desempenho, que será explorado na seção 6.

5.1 RESULTADOS DOS MODELOS FINAIS: IDEB

O modelo 9 apresenta os resultados do modelo para a variável dependente logaritmizada, IDEB (ensino fundamental anos iniciais), somente com as variáveis estatisticamente significantes. É importante notar que, em comparação com o modelo 3, uma vez que o modelo foi tornado mais parcimonioso, duas outras variáveis apresentaram

⁴ Modelo linear clássico.

significância estatística a 10%. De acordo com os resultados apresentados no modelo 9, escolas com um percentual de docentes sem ensino superior acima do quartil 75 possuem um desempenho 2% inferior no IDEB.

A segunda variável mais importante, em termos da magnitude do parâmetro, é a taxa de distorção idade-série. No entanto, em aspectos práticos, essa é a variável com maior importância no modelo: uma redução de 10% no percentual de alunos matriculados na escola com atraso escolar de 2 anos ou mais está associada a um aumento de 7,2% na média municipal do IDEB. As outras variáveis, ainda que significantes em termos estatísticos, são pouco significantes em termos de sua magnitude.

É importante salientar que o caráter agregado destas estimações inviabiliza uma interpretação mais precisa dos efeitos destas políticas no nível do aluno ou da escola. Mas estes resultados podem ser indicativos importantes na estratégia de melhora do desempenho e rendimento escolar, onde a formação dos professores e o desenvolvimento e manutenção dos alunos nas séries adequadas à sua idade, parecem ter relevância.

MODELO 9 - MODELO DE REGRESSÃO - IDEB - ENSINO FUNDAMENTAL ANOS INICIAIS

Source	SS	df	MS	Number of obs	=	397
Model	.539468817	5	.107893763	F(5, 391)	=	12.02
Residual	3.5103354	391	.00897784	Prob > F	=	0.0000
				R-squared	=	0.1332
				Adj R-squared	=	0.1221
Total	4.04980422	396	.010226778	Root MSE	=	.09475

ln_IDEB_fun1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
doc_sem_sup_fun1_media	-.0212396	.0109562	-1.94	0.053	-.04278 .0003009
perc_IRD_baixo_medioabaixo	-.0003388	.0001923	-1.76	0.079	-.0007169 .0000393
dist_is_fun1	-.0072529	.0010736	-6.76	0.000	-.0093636 -.0051421
percent_sem_comp_fun	-.0027591	.0010267	-2.69	0.008	-.0047776 -.0007406
percent_sem_internet_fun	.0014043	.000812	1.73	0.085	-.0001922 .0030008
_cons	1.928594	.0136289	141.51	0.000	1.901799 1.955389

FONTE: *Software* STATA

NOTA: Elaboração do autor.

O modelo 10 apresenta os resultados do modelo para a variável dependente logaritmizada, IDEB (ensino fundamental anos finais). Escolas com um percentual de docentes sem ensino superior acima do quartil 90 possuem um desempenho 4,7% inferior no IDEB. A exemplo do modelo anterior, além do percentual de docentes sem ensino superior, a taxa de distorção idade-série é a única variável do modelo com significância em termos de sua magnitude. Uma redução de 10% na distorção idade-série está associada a um aumento de 1,7% no desempenho médio municipal no IDEB.

MODELO 10 - MODELO DE REGRESSÃO - IDEB - ENSINO FUNDAMENTAL ANOS FINAIS

Source	SS	df	MS	Number of obs	=	389
Model	.16114144	4	.04028536	F(4, 384)	=	8.16
Residual	1.89503862	384	.004934996	Prob > F	=	0.0000
				R-squared	=	0.0784
				Adj R-squared	=	0.0688
Total	2.05618006	388	.005299433	Root MSE	=	.07025

ln_IDEB_fun2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
doc_sem_sup_fun2_media	-.047349	.0123603	-3.83	0.000	-.0716514 -.0230467
perc_IRD_baixo_mediobaixo	.0003191	.0001432	2.23	0.026	.0000376 .0006005
dist_is_fun2	-.0017682	.000608	-2.91	0.004	-.0029637 -.0005727
percent_sem_comp_fun	-.0009177	.0004346	-2.11	0.035	-.0017721 -.0000633
_cons	1.721282	.0137051	125.59	0.000	1.694336 1.748229

FONTE: *Software* STATA

NOTA: Elaboração do autor.

MODELO 11 - MODELO DE REGRESSÃO – IDEB – ENSINO MÉDIO

Source	SS	df	MS	Number of obs	=	366
Model	.311287727	2	.155643863	F(2, 363)	=	19.01
Residual	2.97251132	363	.008188736	Prob > F	=	0.0000
				R-squared	=	0.0948
				Adj R-squared	=	0.0898
Total	3.28379904	365	.00899671	Root MSE	=	.09049

ln_IDEB_med	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
qtd_acima_media_med	-.0146415	.0095344	-1.54	0.125	-.0333911 .004108
percen_INSE_1234	-.0026333	.0004566	-5.77	0.000	-.0035312 -.0017355
_cons	1.591968	.0202778	78.51	0.000	1.552091 1.631845

FONTE: *Software* STATA

NOTA: Elaboração do autor.

O modelo 11 apresenta os resultados do modelo para a variável dependente logaritmizada, IDEB (ensino médio). Apenas com duas variáveis independentes, como esperado após o modelo preliminar, uma delas, município com quantidade de alunos acima da média, resultou não significativa estatisticamente a 10%. Dado o tamanho da amostra e a significância de 12,5%, optou-se por mantê-la no modelo final, já que em termos de magnitude, é a variável mais importante do modelo.

Segundo o modelo 11, municípios com quantidade média de alunos em sala superior à média estadual estão associados a um IDEB 1% menor do que aqueles com quantidade abaixo da média. Essa evidência demonstra a importância do tamanho da sala no desempenho, mas não apresenta um tamanho ótimo em termos absolutos. Na literatura, evidências quanto ao tamanho da turma são ambíguas. Matavelli e Menezes-Filho (2020), utilizando uma regressão descontínua fuzzy em conjunto com variáveis instrumentais não encontraram evidências de que o tamanho da sala impacte o desempenho dos estudantes.

Camargo (2012), utilizando o mesmo método para escolas públicas do Rio Grande do Sul mostrou que apesar de não haver resultado estatisticamente significativo para o tamanho da turma, há uma relação positiva entre o nível socioeconômico e o desempenho dos alunos. Esse resultado corrobora a evidência do modelo 11: a cada 10% a mais de alunos nos níveis 1, 2, 3 e 4 do INSE, menor o desempenho no IDEB em 2,6%. Essa evidência ressalta a importância da condição socioeconômica no desempenho e rendimento do aluno no ensino médio. Evidência essa que não apareceu nos modelos para os níveis anteriores de ensino. Os modelos 12, 13 e 14 apresentam os resultados finais para a variável dependente logaritmizada, Taxa de Aprovação.

5.2 RESULTADOS DOS MODELOS FINAIS: TAXA DE APROVAÇÃO

Desconsiderando o fator desempenho no IDEB (que é a própria Taxa de Aprovação), a significância em termos da magnitude das variáveis percentual de professores sem ensino superior e distorção idade-série se tornam menores. Escolas com um percentual de docentes sem ensino superior acima do quartil 75 estão associadas a uma taxa de aprovação 0,4% inferior. Uma redução de 10% no percentual de alunos matriculados na escola com atraso escolar de 2 anos ou mais está associada a um aumento de 4,9% nas taxas de aprovação.

Essa evidência aponta a hipótese de que o desempenho em termos das notas em português e matemática podem ser mais sensíveis à formação do professor e à distorção idade-série. A seção 6 corrobora essa hipótese. É por isso que a variável dependente logaritmizada, Ideb, que contempla não só o rendimento do aluno mas também seu desempenho, apresenta resultados ligeiramente maiores. As variáveis relacionadas à regularidade docente e INSE, apesar de estatisticamente significantes, não possuem relevância em termos da sua magnitude. Ainda assim, é importante notar o sinal contraintuitivo da variável “Percentual de alunos nos níveis 1, 2, 3 e 4 do INSE”, positivo, indicando uma associação positiva entre nível socioeconômico baixo e Taxa de Aprovação. Esse resultado será mais explorado no modelo com a variável dependente “Desempenho em Português” e “Desempenho em Matemática” na seção 6.

MODELO 12 - MODELO DE REGRESSÃO – TAXA DE APROVAÇÃO – ENSINO FUNDAMENTAL ANOS INICIAIS

Source	SS	df	MS	Number of obs	=	399
Model	.198874233	4	.049718558	F(4, 394)	=	74.20
Residual	.264012883	394	.000670083	Prob > F	=	0.0000
				R-squared	=	0.4296
				Adj R-squared	=	0.4238
Total	.462887116	398	.001163033	Root MSE	=	.02589

ln_taxa_aprov_fun1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
doc_sem_sup_fun1_media	-.0046304	.0029862	-1.55	0.122	-.0105013 .0012404
percen_INSE_1234	.0002403	.0001227	1.96	0.051	-9.59e-07 .0004815
perc_IRD_baixo_mediobaixo	-.0001203	.0000526	-2.29	0.023	-.0002236 -.0000169
dist_is_fun1	-.0049056	.0002906	-16.88	0.000	-.005477 -.0043342
_cons	4.592025	.0063013	728.74	0.000	4.579637 4.604413

FONTE: *Software* STATA

NOTA: Elaboração do autor.

O modelo 13 apresenta os resultados para a variável dependente logaritmizada, Taxa de Aprovação (Ensino Fundamental Anos Finais). Municípios com quantidade média de alunos em sala superior à média estadual estão associados a taxas de aprovação 1,2% menores do que aqueles com quantidade abaixo da média. Uma redução de 10% no percentual de alunos matriculados na escola com atraso escolar de 2 anos ou mais está associada a um aumento de 3,3% nas taxas de aprovação. As demais variáveis independentes, ainda que apresentem significância estatística, não apresentam significância em termos de sua magnitude.

MODELO 13 - MODELO DE REGRESSÃO – TAXA DE APROVAÇÃO – ENSINO FUNDAMENTAL ANOS FINAIS

Source	SS	df	MS	Number of obs	=	397
Model	.183995196	4	.045998799	F(4, 392)	=	37.34
Residual	.482842938	392	.001231742	Prob > F	=	0.0000
				R-squared	=	0.2759
				Adj R-squared	=	0.2685
Total	.666838134	396	.001683935	Root MSE	=	.0351

ln_taxa_aprov_fun2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
esforco_doc_fun_456	.0004364	.0002258	1.93	0.054	-7.49e-06 .0008804
qtd_acima_media_fun2	-.0122794	.0066729	-1.84	0.066	-.0253986 .0008397
perc_IRD_baixo_mediobaixo	-.0001198	.0000707	-1.69	0.091	-.0002588 .0000192
dist_is_fun2	-.0033537	.0002999	-11.18	0.000	-.0039433 -.0027641
_cons	4.585681	.0137905	332.52	0.000	4.558568 4.612794

FONTE: *Software* STATA

NOTA: Elaboração do autor.

MODELO 14: MODELO DE REGRESSÃO – TAXA DE APROVAÇÃO – ENSINO MÉDIO

Source	SS	df	MS	Number of obs	=	399
				F(2, 396)	=	55.35
Model	.218653234	2	.109326617	Prob > F	=	0.0000
Residual	.782222217	396	.001975309	R-squared	=	0.2185
				Adj R-squared	=	0.2145
Total	1.00087545	398	.002514762	Root MSE	=	.04444

ln_taxa_aprov_med	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
gtd_acima_media_med	-.0260325	.0044681	-5.83	0.000	-.0348166 - .0172483
dist_is_med	-.0029437	.0003519	-8.37	0.000	-.0036354 - .0022519
_cons	4.575629	.0074959	610.42	0.000	4.560892 4.590366

FONTE: *Software* STATA

NOTA: Elaboração do autor.

Por fim, o modelo 14 apresenta os resultados para a variável dependente logaritmizada, Taxa de aprovação (Ensino Médio). Municípios com quantidade média de alunos em sala superior à média estadual estão associados a taxas de aprovação 2,6% menores do que aqueles com quantidade abaixo da média. Uma redução de 10% no percentual de alunos matriculados na escola com atraso escolar de 2 anos ou mais está associada a um aumento de 2,9% nas taxas de aprovação.

O quadro 2 apresenta um resumo dos resultados estimados nesta seção, separando em efeitos positivos e negativos e por variável dependente.

QUADRO 2 - RESUMO DOS RESULTADOS DO MODELO EXPLORATÓRIO

NÍVEIS DE ENSINO	IDEB - SINAIS POSITIVOS (+)	IDEB - SINAIS NEGATIVOS (-)	TAXA DE APROVAÇÃO – SINAIS POSITIVOS (+)	TAXA DE APROVAÇÃO – SINAIS NEGATIVOS (-)
Ensino fundamental - anos iniciais	Percentual de escolas sem internet*	Percentual de professores sem ensino superior acima do quartil 75 Taxa de distorção idade-série Percentual de escolas sem computadores Percentual de professores com Índice de Regularidade Docente Baixo e Médio Baixo ⁽¹⁾	Percentual de alunos nos níveis 1-4 do Inse	Percentual de professores sem ensino superior acima do quartil 75 ⁽²⁾ Percentual de professores com Índice de Regularidade Docente Baixo e Médio Baixo Taxa de distorção idade-série
Ensino fundamental - anos finais	Índice de Regularidade Docente Baixo	Percentual de professores sem ensino superior acima do quartil 90 Taxa de distorção idade-série Percentual de escolas sem computadores	Esforço Docente Níveis 4-6	Quantidade de alunos acima da média por sala Índice de Regularidade Docente Baixo Taxa de distorção idade-série
Ensino Médio		Quantidade de alunos acima da média por sala ⁽²⁾ Percentual de alunos nos níveis 1-4 do Inse		Quantidade de alunos acima da média por sala Taxa de distorção idade-série

FONTE: O autor (2021)

(1) Variáveis que não eram significantes no modelo exploratório mas que se tornaram no modelo final.

(2) Variáveis significantes a 12,5%.

Esses resultados estão em consonância com a literatura: Costa e Carnoy (2015), Gonçalves e Lourenço (2019) e Menezes-Filho *et al.* (2008) apresentam evidências de que a formação dos professores, bem como a manutenção dos alunos em séries adequadas estão relacionadas a maiores níveis de desempenho. Costa e Carnoy (2015) e Bruns, Costa e Cunha (2018) apresentam evidência da regularidade docente e da formação dos professores na taxa de aprovação.

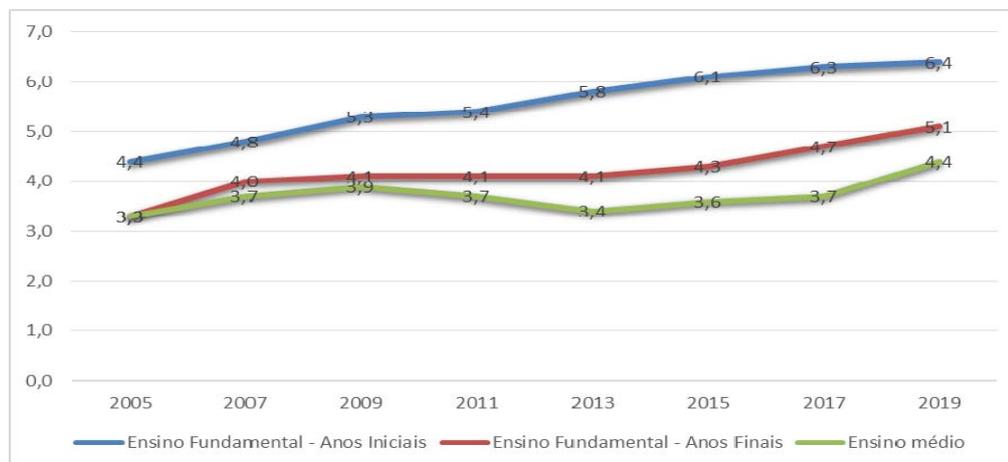
Por fim, três hipóteses importantes do modelo de regressão múltipla são a baixa correlação entre as variáveis independentes (a alta correlação indica multicolinearidade), a variância constante (o não cumprimento dessa hipótese é chamado heterocedasticidade) e a normalidade dos resíduos. Os testes corroborando essas hipóteses encontram-se no Anexo B.

6 MENSURANDO A INFLUÊNCIA DE INDICADORES EDUCACIONAIS: DESEMPENHO NO SAEB

A seção anterior apresentou os modelos para as variáveis dependentes IDEB e Taxa de Aprovação. Como explicado anteriormente, o IDEB é calculado a partir de dois componentes: a taxa de rendimento escolar (Taxa de Aprovação) e as médias de desempenho em Língua Portuguesa e Matemática do Saeb. Dada a baixa variação e o já elevado nível da Taxa de Aprovação (com médias estaduais de 95,29, 93,68 e 90,49 para os anos iniciais e finais do ensino fundamental e para o ensino médio, respectivamente), fica evidente o papel que o desempenho no Saeb tem na política pública como forma viável de influenciar na evolução do IDEB.

A figura 15 apresenta a série histórica bianual da média estadual do IDEB, de 2005 a 2019. A figura mostra tendências de crescimento distintas para cada nível de ensino. Para os anos iniciais do ensino fundamental, a taxa média de crescimento é de 5,5% entre 2005 e 2019, com forte desaceleração a partir de 2015. Para os anos finais do ensino fundamental, em que são registradas as maiores taxas de crescimento, a taxa média para o período ficou em 6,6%, com forte estagnação entre 2011 e 2013 e retomada a partir de 2015. Para o ensino médio, a taxa média de crescimento ficou em 4,6%, a menor do ensino básico, tendo inclusive queda registrada no IDEB entre 2011 e 2013 e forte retomada em 2019.

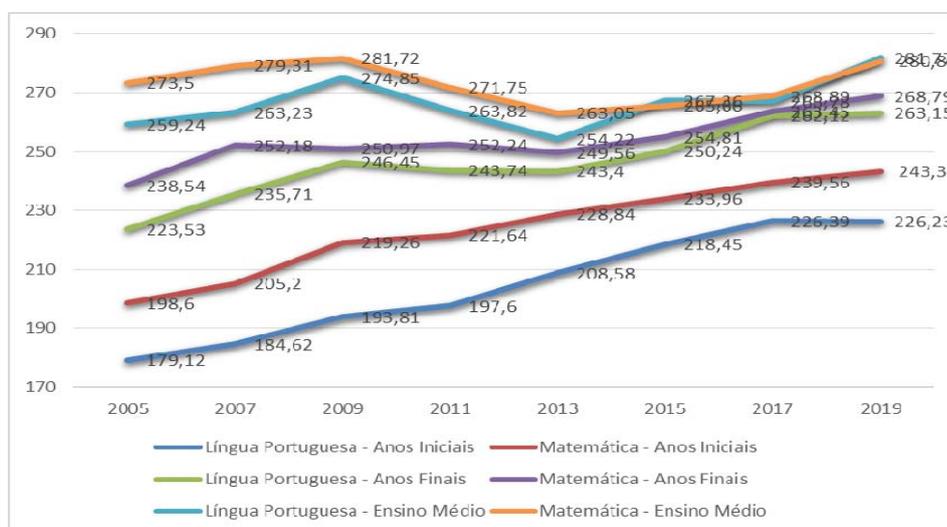
FIGURA 15 - SÉRIE HISTÓRICA DO IDEB ESTADUAL – ENSINO FUNDAMENTAL ANOS INICIAIS, FINAIS E ENSINO MÉDIO - 2005-2019



FONTE: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019)

A figura 16 apresenta a evolução histórica do desempenho em Língua Portuguesa e Matemática para os diferentes níveis de ensino da educação básica, entre 2005 e 2019. A análise das taxas de crescimento do período mostra que apenas a evolução do desempenho de matemática nos anos iniciais do ensino fundamental, não apresentou taxas negativas entre os anos analisados. No que concerne às taxas médias de crescimento, os anos iniciais do ensino fundamental apresentaram as maiores, registrando 3,4% para Língua Portuguesa e 3,0% para Matemática. Para os anos finais do ensino fundamental, as taxas médias ficaram em 2,4% e 1,7% para Língua Portuguesa e Matemática, respectivamente. Para o ensino médio, as taxas de crescimento médias ficaram em 1,3% e 0,4%, respectivamente, para Língua Portuguesa e Matemática.

FIGURA 16 - SÉRIE HISTÓRICA DO DESEMPENHO ESTADUAL NO SAEB EM LÍNGUA PORTUGUESA E MATEMÁTICA - ENSINO FUNDAMENTAL ANOS INICIAIS, FINAIS E ENSINO MÉDIO - 2005-2019

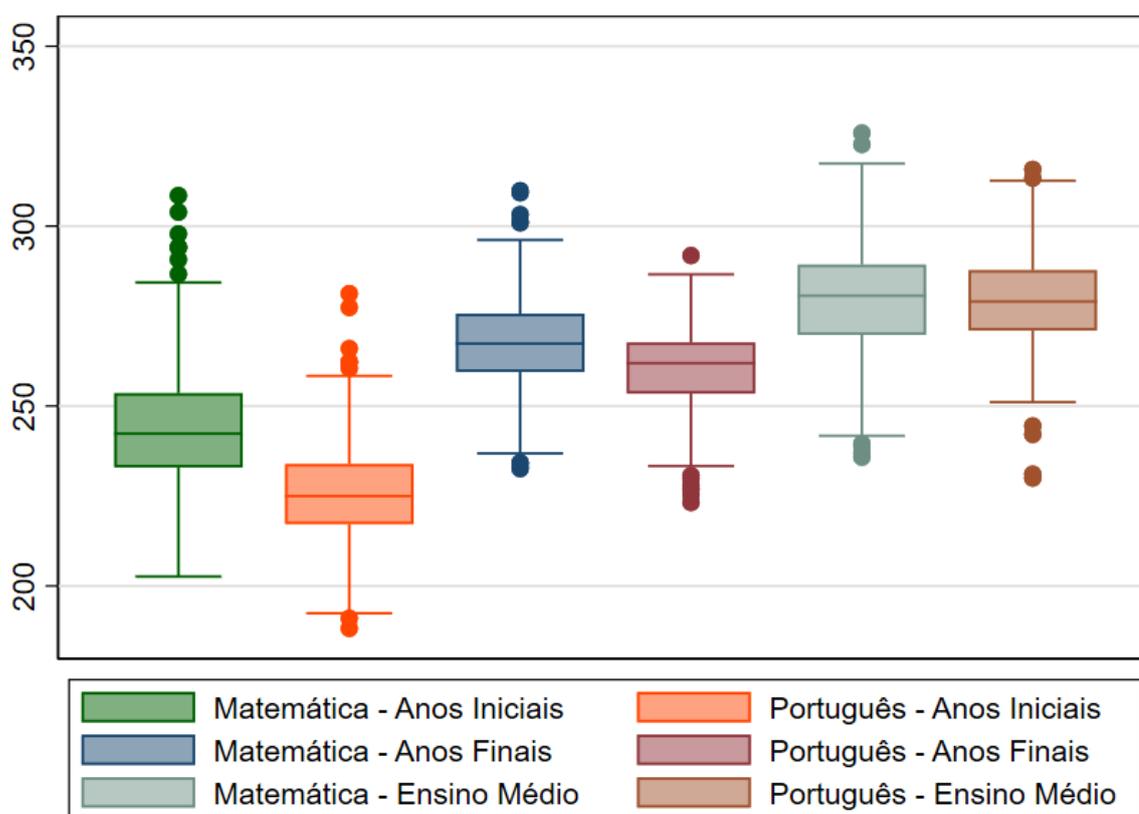


FONTE: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019)

Como já discutido, a evolução do desempenho no ensino médio demonstra ser esse o nível de ensino que requer maior atenção, dada a meta de aumentar o desempenho educacional, especialmente o desempenho em Matemática. É no ensino médio também que são registradas taxas de crescimento negativas mais significativas: -4% de 2009 para 2011, -3,6% de 2011 para 2013 e -0,2% de 2015 para 2017. Entretanto, de 2017 a 2019 é o desempenho no ensino médio que apresentou maior evolução, registrando 5,6% para Língua Portuguesa e 4,4% para Matemática, superior ao registrado no ensino fundamental, que melhorou em média 1% em cada biênio.

Portanto, compreender como o desempenho se relaciona às variáveis apresentadas na seção 3, bem como as diferenças em relação às evidências discutidas na seção 5, é fundamental para endossar critérios de atuação sobre a melhora do desempenho estadual. A figura 17 mostra que além das notas em matemática apresentarem maior variabilidade, exceto para matemática nos anos iniciais do ensino fundamental, a presença de *outliers* na cauda inferior indica municípios que estão muito abaixo da média, principalmente em português no ensino médio.

FIGURA 17 - GRÁFICO BOXPLOT DO DESEMPENHO ESTADUAL NO SAEB EM LÍNGUA PORTUGUESA E MATEMÁTICA – ENSINO FUNDAMENTAL ANOS INICIAIS, FINAIS E ENSINO MÉDIO EM 2019



FONTE: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019)

Assim, a subseção seguinte apresenta evidências empíricas de associação dos diversos indicadores do INEP com as medidas de desempenho em Português e Matemática, nos três níveis educacionais, anos iniciais do ensino fundamental, anos finais e ensino médio.

6.1 RESULTADOS DOS MODELOS FINAIS: DESEMPENHO EM PORTUGUÊS E MATEMÁTICA

Do mesmo modo da seção 5, os modelos apresentados nesta seção serão analisados segundo dois critérios (1) significância estatística e; (2) significância na magnitude. O intuito é mensurar, dentre as variáveis independentes, quais possuem maior importância relativa sobre o desempenho escolar em Português e Matemática. O modelo 15 apresenta os resultados do modelo de regressão ajustado para a variável dependente logaritimizada “Desempenho em Língua Portuguesa”.

MODELO 15 - MODELO DE REGRESSÃO - DESEMPENHO EM LÍNGUA PORTUGUESA - ENSINO FUNDAMENTAL - ANOS INICIAIS

Source	SS	df	MS	Number of obs	=	397
Model	.259687283	3	.086562428	F(3, 393)	=	26.92
Residual	1.26354276	393	.003215122	Prob > F	=	0.0000
Total	1.52323004	396	.003846541	R-squared	=	0.1705
				Adj R-squared	=	0.1642
				Root MSE	=	.0567

lnpt_ai	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
percen_INSE_1234	-.0022533	.00027	-8.35	0.000	-.0027841 - .0017225
perc_IRD_baixo_mediobaixo	-.0002788	.0001144	-2.44	0.015	-.0005036 - .0000539
percent_sem_comp_fun	-.0009677	.0003439	-2.81	0.005	-.0016438 - .0002916
_cons	5.528382	.0135605	407.68	0.000	5.501721 5.555042

FONTE: *Software* STATA

NOTA: Elaboração do autor.

A principal diferença entre os modelos com as variáveis dependentes IDEB e Taxa de Aprovação para os modelos com a variável dependente “Desempenho em Português” é a importância do nível socioeconômico. No modelo com a variável dependente “Taxa de Aprovação” a variável “Percentual de alunos nos níveis 1, 2, 3 e 4 do INSE” apareceu com um contraintuitivo sinal positivo. Dada a baixa variação das Taxas de Aprovação e a baixa relevância da variável em termos de magnitude, é interessante notar a inversão do sinal e aumento da magnitude quando da variável dependente desempenho.

De acordo com os resultados apresentados no modelo 15, corroborando a evidência apresentada por Camargo (2012), em termos de magnitude, o nível socioeconômico é a variável independente mais relevante no modelo. A cada 10% a mais de alunos nos níveis 1, 2, 3 e 4 do INSE, está associado um desempenho em Português 2,3% menor. Ainda que com

menor relevância em termos de magnitude, municípios com maior percentual de professores com Índice de Regularidade Docente média baixa e baixa e com maior percentual de escolas sem computadores estão associados a um menor desempenho.

O modelo 16, para “Desempenho em Matemática” apresenta evidências muito próximas do modelo para “Desempenho em Português” em termos do Índice de Regularidade Docente e acesso a computadores. A variável mais relevante do modelo é o nível socioeconômico, estimando que a cada 10% a mais de alunos nos níveis 1, 2, 3 e 4 do INSE, está associado um desempenho em Matemática 2,1% menor. A grande diferença é que para o desempenho em Matemática, a distorção idade-série parece ser mais importante. Sendo essa disciplina dependente de pré-requisitos para a evolução dos alunos, o atraso dos alunos parece prejudicar mais o desempenho do que em Português. De acordo com o modelo, uma redução de 10% na distorção idade-série está associada a um aumento de 1,2% no desempenho em Matemática.

MODELO 16 -MODELO DE REGRESSÃO - DESEMPENHO EM MATEMÁTICA - ENSINO FUNDAMENTAL - ANOS INICIAIS

Source	SS	df	MS	Number of obs	=	397
Model	.264384916	4	.066096229	F(4, 392)	=	16.32
Residual	1.5880272	392	.00405109	Prob > F	=	0.0000
				R-squared	=	0.1427
				Adj R-squared	=	0.1340
Total	1.85241211	396	.004677808	Root MSE	=	.06365

	lnmt_ai	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	percen_INSE_1234	-.0021236	.0003103	-6.84	0.000	-.0027337 - .0015135
	perc_IRD_baixo_mediobaixo	-.0002996	.0001285	-2.33	0.020	-.0005522 - .0000047
	dist_is_fun1	-.0012519	.0007362	-1.70	0.090	-.0026994 .0001955
	percent_sem_comp_fun	-.0009231	.0003873	-2.38	0.018	-.0016844 - .0001617
	_cons	5.608509	.0155726	360.15	0.000	5.577893 5.639125

FONTE: *Software STATA*

NOTA: Elaboração do autor.

Com relação ao ensino fundamental anos finais, os modelos 17 e 18 apresentam os resultados para as variáveis dependentes “Desempenho em Português” e “Desempenho em Matemática”. Assim como nos anos iniciais, a evidência da relação desempenho e nível socioeconômico é aparente, relação esta que não apareceu nos modelos para as variáveis dependentes “Ideb” e “Taxa de Aprovação”. Destaca-se também a relevância da variável “Percentual de docentes sem ensino superior acima do quartil 90” tanto para “Desempenho em Português” quanto para “Desempenho em Matemática”, que apareceu no modelo “Ideb” mas não no modelo “Taxa de Aprovação”.

MODELO 17 - MODELO DE REGRESSÃO – DESEMPENHO EM LÍNGUA PORTUGUESA – ENSINO FUNDAMENTAL - ANOS FINAIS

Source	SS	df	MS	Number of obs	=	389
Model	.141838072	5	.028367614	F(5, 383)	=	17.49
Residual	.621366424	383	.001622367	Prob > F	=	0.0000
				R-squared	=	0.1858
				Adj R-squared	=	0.1752
Total	.763204496	388	.001967022	Root MSE	=	.04028

lnpt_af	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
doc_sem_sup_fun2_media	-.0236133	.0071846	-3.29	0.001	-.0377395	-.009487
percen_INSE_1234	-.0013646	.0002006	-6.80	0.000	-.001759	-.0009703
perc_IRD_baixo_mediobaixo	.0001462	.0000825	1.77	0.077	-.0000159	.0003084
dist_is_fun2	-.0007063	.0003523	-2.00	0.046	-.0013991	-.0000136
percent_sem_comp_fun	-.0005562	.0002494	-2.23	0.026	-.0010464	-.0000659
_cons	5.63055	.0111529	504.85	0.000	5.608621	5.652478

FONTE: *Software* STATA

NOTA: Elaboração do autor.

MODELO 18 - MODELO DE REGRESSÃO – DESEMPENHO EM MATEMÁTICA – ENSINO FUNDAMENTAL – ANOS FINAIS

Source	SS	df	MS	Number of obs	=	389
Model	.137859903	4	.034464976	F(4, 384)	=	17.12
Residual	.773268565	384	.00201372	Prob > F	=	0.0000
				R-squared	=	0.1513
				Adj R-squared	=	0.1425
Total	.911128468	388	.002348269	Root MSE	=	.04487

lnmt_af	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
doc_sem_sup_fun2_media	-.0181958	.0080031	-2.27	0.024	-.0339312	-.0024603
percen_INSE_1234	-.0014824	.0002225	-6.66	0.000	-.0019198	-.001045
dist_is_fun2	-.0007842	.0003925	-2.00	0.046	-.0015559	-.0000125
percent_sem_comp_fun	-.0004657	.0002778	-1.68	0.094	-.0010119	.0000805
_cons	5.669347	.0113526	499.39	0.000	5.647026	5.691668

FONTE: *Software* STATA

NOTA: Elaboração do autor.

Escolas com um percentual de docentes sem ensino superior acima do quartil 90 possuem um desempenho 2,4% inferior em Português e 1,8% em Matemática. A segunda variável mais importante em termos de magnitude é o nível socioeconômico, estimando que cada 10% a mais de alunos nos níveis 1, 2, 3 e 4 do INSE, está associado a um desempenho 1,4% menor em Português e 1,5% menor em Matemática. Para a variável “Taxa de distorção idade-série”, uma redução de 10% na distorção idade-série está associada a um aumento de 0,7% no desempenho médio em Português e 0,8% no desempenho em matemática. Nota-se que, diferentemente dos modelos para os anos iniciais, nos anos finais a distorção idade-série é significativa tanto no desempenho em

Matemática como no de Português. No entanto, tanto a distorção idade-série quanto o percentual de escolas sem computadores, apesar da associação negativa estatisticamente significativa, têm uma magnitude muito pequena.

Os modelos 19 e 20 apresentam os resultados para as variáveis dependentes “Desempenho em Português” e “Desempenho em Matemática” para o ensino médio. De acordo com os modelos, municípios com quantidade média de alunos em sala superior à média estadual estão associados a um desempenho em torno de 1% maior do que aqueles com quantidade abaixo da média. Para os modelos “Ideb” e “Taxa de Aprovação” o sinal desta variável resultou negativo, indicando que turmas maiores, apesar de estarem associadas a menores taxas de aprovação, estão associadas a um desempenho maior em Português e Matemática.

MODELO 19 - MODELO DE REGRESSÃO - DESEMPENHO EM LÍNGUA PORTUGUESA - ENSINO MÉDIO

Source	SS	df	MS	Number of obs	=	366
Model	.172634393	4	.043158598	F(4, 361)	=	25.43
Residual	.612646043	361	.00169708	Prob > F	=	0.0000
				R-squared	=	0.2198
				Adj R-squared	=	0.2112
Total	.785280435	365	.002151453	Root MSE	=	.0412

lnpt_em	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
esforco_doc_med_456	.0004717	.0002208	2.14	0.033	.0000375	.0009059
qtd_acima_media_med	.0077272	.0046046	1.68	0.094	-.001328	.0167824
percen_INSE_1234	-.001624	.000218	-7.45	0.000	-.0020527	-.0011953
dist_is_med	-.0016182	.0003473	-4.66	0.000	-.0023011	-.0009352
_cons	5.688966	.0221486	256.85	0.000	5.645409	5.732522

FONTE: *Software* STATA

NOTA: Elaboração do autor.

MODELO 20 - MODELO DE REGRESSÃO - DESEMPENHO EM MATEMÁTICA - ENSINO MÉDIO

Source	SS	df	MS	Number of obs	=	366
Model	.298350258	5	.059670052	F(5, 360)	=	25.95
Residual	.827815872	360	.002299489	Prob > F	=	0.0000
				R-squared	=	0.2649
				Adj R-squared	=	0.2547
Total	1.12616613	365	.003085387	Root MSE	=	.04795

lnmt_em	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
esforco_doc_med_456	.0006386	.0002598	2.46	0.014	.0001277	.0011496
qtd_acima_media_med	.0093058	.0053621	1.74	0.084	-.0012393	.0198508
percen_INSE_1234	-.0018949	.0002547	-7.44	0.000	-.0023958	-.001394
perc_IRD_baixo_medioabaixo	.0002165	.0001037	2.09	0.038	.0000126	.0004204
dist_is_med	-.0024039	.0004081	-5.89	0.000	-.0032065	-.0016013
_cons	5.692174	.0275932	206.29	0.000	5.637909	5.746438

FONTE: *Software* STATA

NOTA: Elaboração do autor.

Como discutido na seção 5.1, a evidência do tamanho da sala no desempenho (MATAVELLI; MENEZES-FILHO, 2020; CAMARGO, 2012) é inconclusiva, o que indica que talvez o efeito positivo sobre o desempenho seja anulado pelo efeito negativo sobre a taxa de aprovação.

Quanto à distorção idade-série, o modelo estima que uma redução de 10% na distorção idade-série está associada a um aumento de 1,6% no desempenho médio em Português e 2,4% no desempenho em matemática. No que diz respeito ao nível socioeconômico, a cada 10% a mais de alunos nos níveis 1, 2, 3 e 4 do INSE, o desempenho é 1,6% menor em Português e 1,9% menor em Matemática. Essa evidência ressalta a importância da condição socioeconômica no desempenho do aluno, evidência essa que não apareceu nos modelos “Ideb” e “Taxa de Aprovação” para os níveis do fundamental, mas apareceu para todos os níveis de ensino nos modelos “Desempenho em Português” e “Desempenho em Matemática”.

Por fim, municípios em que o nível de esforço (sobrecarga) dos professores é alto e o Índice de Regularidade Docente é baixo, estão associados a maiores desempenhos em Português e Matemática no ensino médio. Apesar do resultado contraintuitivo, para estas variáveis não há relevância em termos de magnitude.

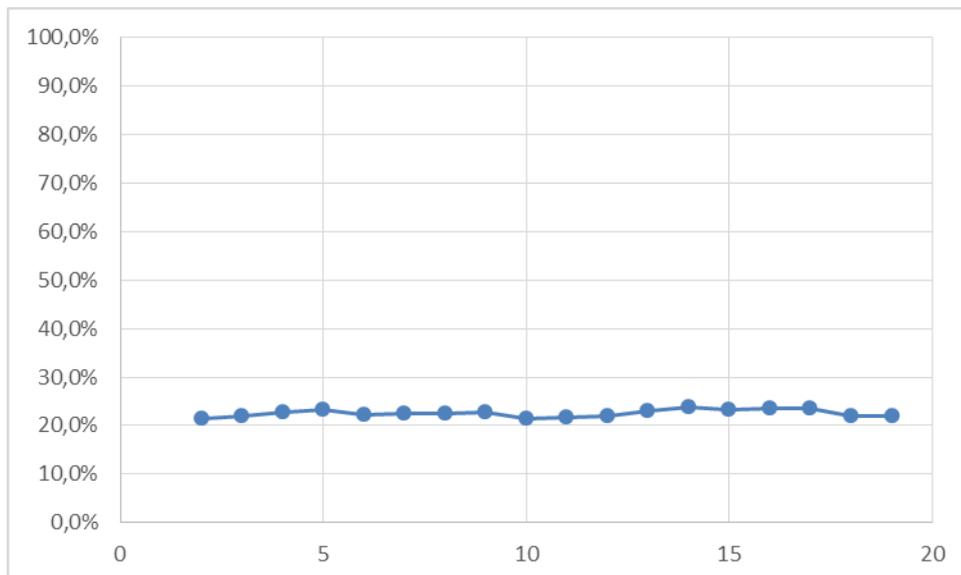
7 POLÍTICAS EDUCACIONAIS, ANOS MÉDIOS DE ESCOLARIDADE E PRODUTIVIDADE TOTAL DOS FATORES

Essa seção tem como objetivo apresentar um panorama descritivo e exploratório da evolução temporal do investimento agregado em educação no Paraná de 2002 a 2019, bem como das variáveis taxa de aprovação, reprovação, abandono e distorção idade-série, de 2006 a 2019, complementando a evolução temporal das variáveis de desempenho apresentadas na seção 6. Discute-se também, brevemente, a questão da qualidade educacional e algumas evidências da literatura brasileira e internacional a esse respeito. Por fim, será apresentada uma estimativa para a produtividade total dos fatores paranaense através da metodologia dos resíduos de Solow apresentada por Jones (2015) e Barbosa, Petterino e Ferreira (2015) e sua correlação com os anos médios de escolaridade.

7.1 EVOLUÇÃO TEMPORAL DAS VARIÁVEIS EDUCACIONAIS PARANAENSES

Os dados do percentual de despesas paranaenses de ações voltadas à educação e ações voltadas para a ciência e tecnologia, em relação ao total, são oriundos do Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro (SICONFI) da Secretaria do Tesouro Nacional (STN). As figuras 18 e 19 apresentam a evolução temporal deste percentual entre 2002 a 2019. No que concerne às despesas de ações voltadas à educação, é possível notar uma estabilidade nos investimentos, com uma média de 22,5% das despesas totais.

FIGURA 18 - PERCENTUAL DE DESPESAS PARANAENSES DE AÇÕES VOLTADAS À EDUCAÇÃO EM RELAÇÃO AO TOTAL DE DESPESAS ESTADUAIS

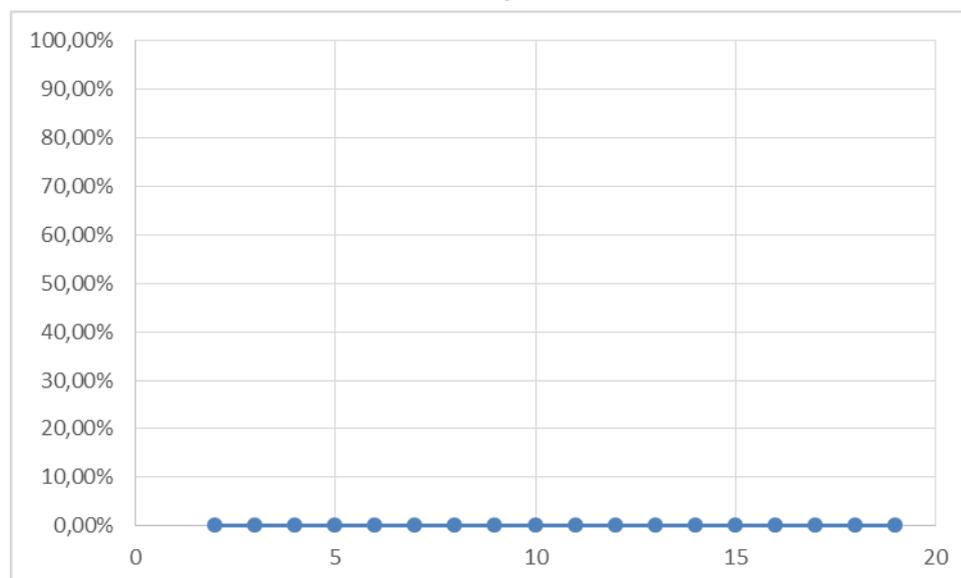


FONTE: Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro (SICONFI) da Secretaria do Tesouro Nacional (STN)

NOTA: Elaboração do autor.

Sobre o percentual de despesas de ações voltadas à ciência e tecnologia, em relação ao total, a figura 19 mostra que esses investimentos não tiveram grande relevância no orçamento do estado, ficando com média de 0,03% no período, chegando a um máximo de 0,07% em 2011 e um mínimo de 0,004% em 2005.

FIGURA 19 - PERCENTUAL DE DESPESAS PARANAENSES DE AÇÕES VOLTADAS PARA A CIÊNCIA E TECNOLOGIA EM RELAÇÃO AO TOTAL DE DESPESAS ESTADUAIS



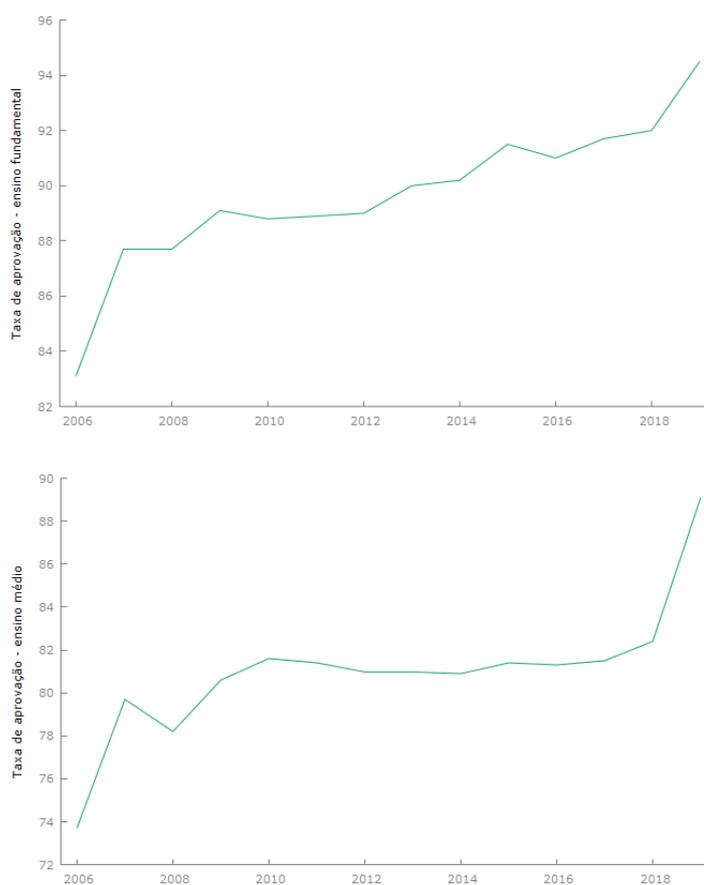
FONTE: Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro (SICONFI) da Secretaria do Tesouro Nacional (STN)

NOTA: Elaboração do autor.

A figura 20 apresenta a evolução temporal para a variável Taxa de Aprovação para o ensino fundamental (anos iniciais e finais em um mesmo gráfico) e ensino médio. Mais uma vez é possível observar três tendências distintas. A primeira tendência, de 2006 a 2007, apresentou uma taxa de crescimento anual de 6%, seguida de um período de estagnação de 2008 a 2017 em que a taxa de crescimento média foi zero. De 2018 a 2019 a taxa de crescimento ficou em 3%.

Para o ensino médio, verifica-se também a ocorrência de três tendências distintas. No entanto, de 2006 a 2007 e de 2018 a 2019, a taxa média de crescimento para a taxa de aprovação no ensino médio foi de 8%, acima do crescimento registrado para o ensino fundamental. É importante perceber que as taxas de aprovação para o ensino fundamental já se encontravam, em 2019, em quase 95%, enquanto para o ensino médio, encontram-se ainda abaixo dos 90%.

FIGURA 20 - EVOLUÇÃO DA TAXA DE APROVAÇÃO PARA O ENSINO FUNDAMENTAL E ENSINO MÉDIO



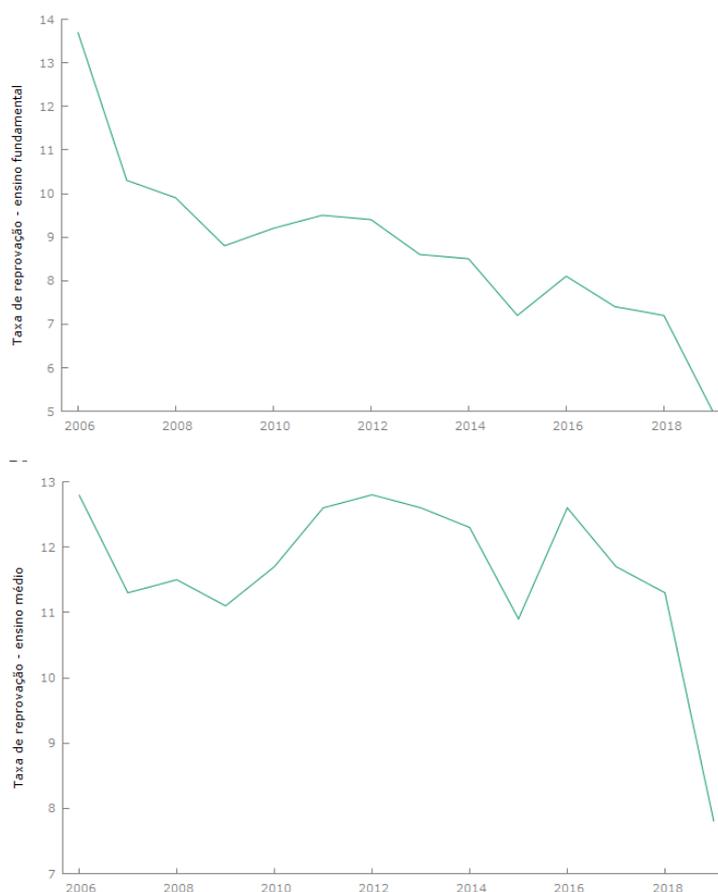
FONTE: Ministério da Educação / Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - MEC/INEP

NOTA: Elaboração do autor.

A figura 21 apresenta a evolução temporal para a variável Taxa de Reprovação para o ensino fundamental (anos iniciais e finais em um mesmo gráfico) e ensino médio. Para o ensino fundamental, quatro tendências podem ser observadas. A primeira, de 2006 a 2009, é possível observar uma tendência de queda em que a taxa de crescimento média no período foi de -13%. De 2009 a 2012, observou-se uma piora, com aumentos da Taxa de Reprovação a uma taxa média de 3,9%. Entre 2012 e 2016, a taxa média de crescimento ficou em -2,7%, só piorada devido ao aumento da taxa de 2015 a 2016. O último período analisado, de 2017 a 2019, apresentou taxa de crescimento médio de -13,8%.

Para o ensino médio também é possível notar as três tendências observadas. No entanto, a queda observada no primeiro período é menor: taxa média de crescimento de -4%. A piora registrada entre 2009 e 2012 para o ensino médio também mostrou-se mais acentuada que no ensino fundamental, com uma taxa média de 4,9%. De 2012 a 2016, os movimentos de melhora na taxa de reprovação foram completamente absorvidos pelos movimentos de piora, resultando em uma taxa média de crescimento de 0,1%. No último período analisado, de 2017 a 2019, assim como no ensino fundamental, a taxa média de crescimento foi de -13,8%.

FIGURA 21 - EVOLUÇÃO DA TAXA DE REPROVAÇÃO PARA O ENSINO FUNDAMENTAL E ENSINO MÉDIO

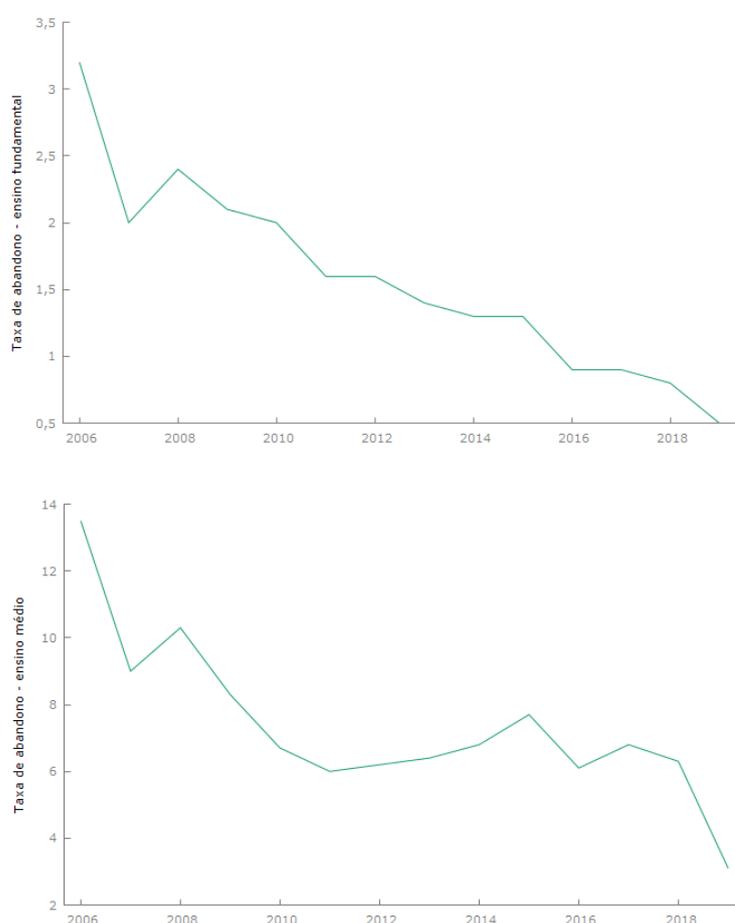


FONTE: Ministério da Educação / Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - MEC/INEP

NOTA: Elaboração do autor.

A figura 22 apresenta a evolução temporal para a variável Taxa de Abandono para o ensino fundamental (anos iniciais e finais em um mesmo gráfico) e ensino médio. Para o ensino fundamental, três tendências são observadas. A primeira, de 2005 para 2006, registrou uma taxa de crescimento de -37,5%, seguida de uma piora, com taxa de crescimento de 20%, de 2006 para 2007. Ainda que com períodos de estagnação (2011 para 2012, 2014 para 2015 e 2016 para 2017), a taxa média de crescimento no período de 2009 a 2019 foi de -12,4%, chegando a uma taxa de abandono de 0,5% em 2019, no ensino médio.

FIGURA 22 - EVOLUÇÃO DA TAXA DE ABANDONO PARA O ENSINO FUNDAMENTAL E ENSINO MÉDIO



FONTE: Ministério da Educação / Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - MEC/INEP

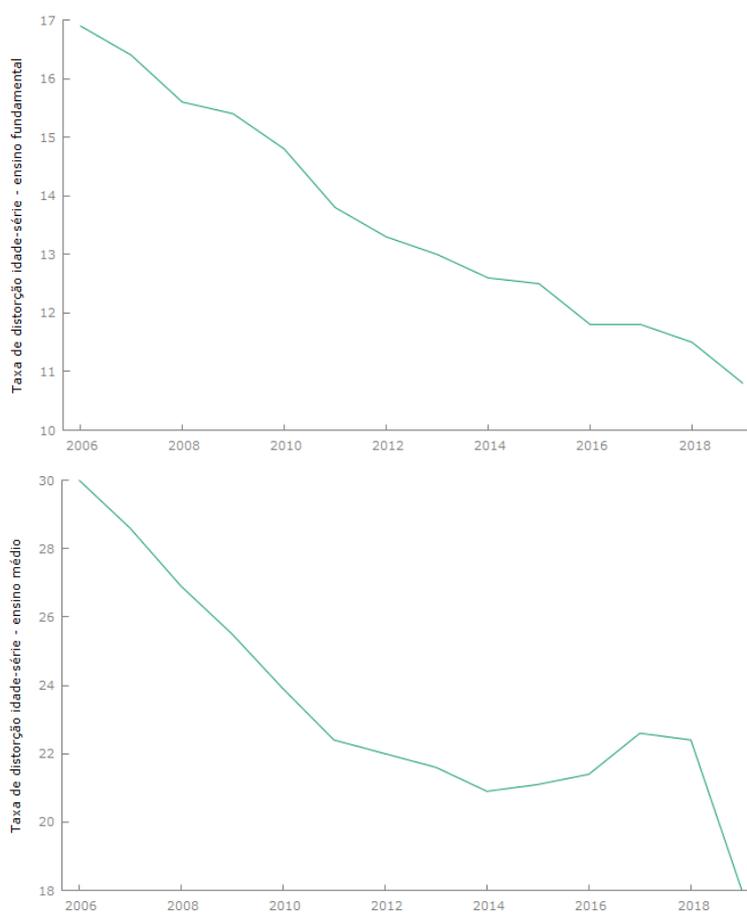
NOTA: Elaboração do autor.

Para o ensino médio, como já discutido anteriormente, o abandono é um problema maior, já que neste período o custo de oportunidade da educação começa a se tornar maior. Na figura 24 é possível notar três tendências distintas. De 2006 a 2011, ainda que com uma leve piora de 2007 para 2008, a tendência era de queda, registrando uma taxa

média de crescimento de -13,6%. No entanto, de 2012 a 2017, exceto pela redução registrada entre 2015 e 2016, houve aumento da taxa de abandono em todos os anos. Por fim, 2018 e 2019 registraram melhoras notáveis na taxa de abandono, com reduções de -7,4% de 2017 para 2018 e de -50,8% de 2018 para 2019.

Por fim, a figura 23 apresenta a evolução temporal para a variável Taxa de Distorção Idade-Série para o ensino fundamental (anos iniciais e finais em um mesmo gráfico) e ensino médio. A distorção idade-série mostra a porcentagem dos alunos matriculados que têm idade pelo menos 2 anos maior do que a idade esperada para aquela série. Para o ensino fundamental é possível notar uma queda vertiginosa em praticamente todo o período (apenas uma estagnação entre 2016 e 2017), com taxa média de crescimento de -3,4%. Ainda assim, em 2019, a porcentagem de alunos matriculados em séries não adequadas era de 10,8%.

FIGURA 23 - EVOLUÇÃO DA TAXA DE DISTORÇÃO IDADE-SÉRIE PARA O ENSINO FUNDAMENTAL E ENSINO MÉDIO



FONTE: Ministério da Educação, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - MEC/INEP

NOTA: Elaboração do autor.

No entanto, o problema da distorção idade-série se mostra mais relevante no ensino médio. De 2006 a 2014 a taxa média de crescimento registrada foi de -4,4%. Mas entre 2015 e 2017 a taxa média de crescimento foi de 2,7%. Apenas de 2017 para 2018 e de 2018 para 2019 as taxas voltaram a decrescer, com taxas de crescimento de -0,9% e -19,6%, respectivamente. Ainda que a redução no último período tenha sido grande, em 2019 a distorção idade-série era um problema para 18% dos alunos matriculados.

É patente que o desempenho em termos das taxas de aprovação, reprovação, abandono e distorção idade-série apresenta grande defasagem, principalmente, no ensino médio, assim como apresentado na seção em relação ao desempenho em Português e Matemática. Ainda que os modelos das seções anteriores evidenciem associações relevantes entre mudanças na escola e aumentos no desempenho, é importante salientar que o debate da qualidade educacional, em especial no Brasil, é bastante complexo.

7.2 A QUESTÃO DA QUALIDADE EDUCACIONAL

Quando discute-se qualidade educacional é importante ter em mente que o conceito geral de qualidade utilizado pelas políticas públicas brasileiras é bastante bem definido, sendo a maioria dessas políticas relacionadas à expansão dos insumos escolares, ainda que a evidência de que insumos e melhora da educação seja um tópico discutível. A heterogeneidade das escolas brasileiras e, especialmente para este estudo, paranaenses, pressupõe a existência de impactos distintos para cada nível de qualidade. Os histogramas apresentados na seção 3 são uma evidência dessa heterogeneidade.

Quando se foca, por exemplo, em anos médios de estudo, o debate se centra nos níveis médios educacionais (quantidade), com pouco destaque para o papel da qualidade educacional. Trabalhos como Hanushek e Kimko (2000) mostram que a inclusão de variáveis de qualidade em seus modelos de crescimento representa um grande aumento no poder de explicação dos modelos. No entanto, como apontado por Menezes-Filho (2009), utilizando uma amostra de municípios brasileiros, só existe relação entre gastos educacionais e desempenho nos municípios com maiores níveis de qualidade de ensino.

Santos (2015), discutindo questões acerca das diferenças entre os achados da literatura brasileira e a literatura internacional para ensino infantil, argumenta que essas diferenças refletem a heterogeneidade da qualidade das escolas frequentadas por pobres e ricos e que os impactos da educação podem ser distintos a depender da qualidade dessas escolas. O autor mostra que, diferentemente da literatura internacional, são os menos vulneráveis que mais se beneficiam das políticas públicas educacionais brasileiras.

De acordo com Paes de Barros *et al.* (2001), examinando um universo de indivíduos entre 11 e 25 anos em áreas urbanas das regiões Nordeste e Sudeste, investigaram quatro tipos de determinantes do desempenho educacional: i) a disponibilidade e qualidade dos serviços educacionais; ii) a atratividade do mercado de trabalho local (custo de oportunidade do tempo); iii) a disponibilidade de recursos familiares (financeiros e não-financeiros) e; iv) o volume de recursos da comunidade em que o indivíduo vive.

Os autores encontraram que, dentre os quatro determinantes, características familiares como a escolaridade dos pais (em particular, da mãe) são o determinante mais importante na explicação do desempenho educacional, inclusive superior à importância da renda *per capita*. Essa evidência corrobora diversos trabalhos que identificam a mobilidade intergeracional como importante preditor da desigualdade educacional (BARROS, 2001; MAHLMEISTER *et al.*, 2014) e corrobora os resultados deste trabalho que relacionam melhores níveis socioeconômicos a maiores níveis de desempenho em Português e Matemática.

Portanto, ainda que as evidências aqui encontradas apresentem magnitudes pequenas, a investigação dos efeitos heterogêneos sobre os diferentes níveis de desempenho e de qualidade devem ser levadas em consideração. A correlação entre educação e crescimento econômico é bem estabelecida na literatura (TOPEL, 1999, KRUEGER; LINDALH, 2001; LANGE; TOPEL, 2006). Mas a relação entre o mecanismo do aumento da produtividade e a evolução do capital humano é bastante ambígua, como discutido na próxima seção.

7.3 CORRELAÇÃO ENTRE A PRODUTIVIDADE TOTAL DOS FATORES (PTF) E ANOS MÉDIOS DE ESTUDOS PARA O PARANÁ

Esta seção tem por objetivo estimar a produtividade total dos fatores para o Paraná, de 2001 a 2019, através de um modelo de crescimento com capital humano. Seguindo Jones (2015), utilizou-se uma função de produção agregada como apresentada por Romer (1990), em que o estoque de capital físico, capital humano e a tecnologia (aqui identificada como a produtividade total dos fatores, PTF) se combinam para gerar o produto. Considerando uma transformação logarítmica no intuito de linearizar a estimação, tem-se:

$$\ln y_t = \ln A_t + \alpha \ln k_t + \beta E_t \quad (1)$$

em que y representa o produto por trabalhador, k representa o capital físico por trabalhador e E o capital humano. Seguindo Barbosa, Petterino e Ferreira (2015), tem-se que a produtividade total dos fatores é:

$$a_t = \ln y_t - \alpha \ln k_t - \beta E_t \quad (2)$$

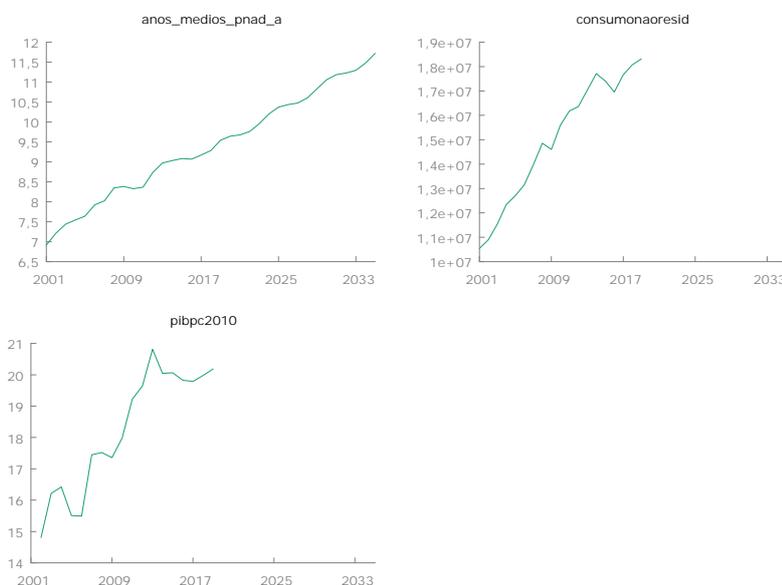
onde $a_t = \log(A_t)$. Da equação (2), a PTF pode ser estimada através dos resíduos da regressão (1).

Para identificar o produto por trabalhador, foi utilizada a variável Produto Interno Bruto per capita, oriunda do Instituto Brasileiro de Geografia e Estatística (IBGE) a preços de 2010. Para identificar o capital físico por trabalhador, foi utilizada, a exemplo de Barbosa, Petterino e Ferreira (2015), o consumo não residencial de energia elétrica. Essa variável foi construída a partir da soma das variáveis “Energia Elétrica no Setor Secundário (Indústria) - Consumo (Mwh)”, “Energia Elétrica no Setor Comercial - Consumo (Mwh)”, “Energia Elétrica no Consumo Livre - Consumo (Mwh)”, “Energia Elétrica em Outras Classes - Consumo (Mwh)” e “Energia

Elétrica Rural – Consumidores”. Para o componente capital humano foi utilizada a série ajustada dos anos de escolaridade da Pesquisa Nacional por Amostra de Domicílios (PNAD) estimada pelo Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES)⁵. Os dados remanescentes foram retirados da Base de Dados do Estado (BDEweb).

A figura 24 apresenta os gráficos das três variáveis, produto *per capita*, consumo não residencial de energia elétrica e anos de escolaridade, para o Paraná, de 2001 a 2019. É possível observar que, sendo os dados não estacionários, estes deverão passar por pelo menos uma diferenciação. Após passarem por uma diferenciação as séries se tornaram estacionárias (hipótese nula de raiz unitária pode ser rejeitada para as três séries), de modo que previsões foram feitas considerando as séries como integradas de primeira ordem.

FIGURA 24 - SÉRIES ORIGINAIS DO PRODUTO PER CAPITA, CONSUMO NÃO RESIDENCIAL DE ENERGIA ELÉTRICA ANOS DE ESCOLARIDADE, DE 2001 A 2019



FONTE: Pesquisa Nacional por Amostra de Domicílios (PNAD)

NOTA: Estimada pelo Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES).

A exemplo da previsão dos anos médios de escolaridade, de modo a ampliar a série, um modelo ARIMA foi utilizado para prever as variáveis PIB per capita paranaense e Consumo de Energia Não Residencial para os anos de 2020 a 2035. O modelo 21 apresenta os resultados de um modelo ARIMA (1, 1, 1) para a série do PIB per capita paranaense. A figura 25 apresenta as previsões para esta variável. Para 2035, a previsão para o PIB per capita paranaense anual é de R\$ 26.150,00, a preços de 2010.

⁵ Mais detalhes no Apêndice A.

MODELO 21 - ARIMA (1, 1, 1) PARA PIB PER CAPITA PARANAENSE

Modelo 1 - ARIMA, usando as observações 2003-2019 (T = 17)

Variável dependente: (1-L) pibpc2010

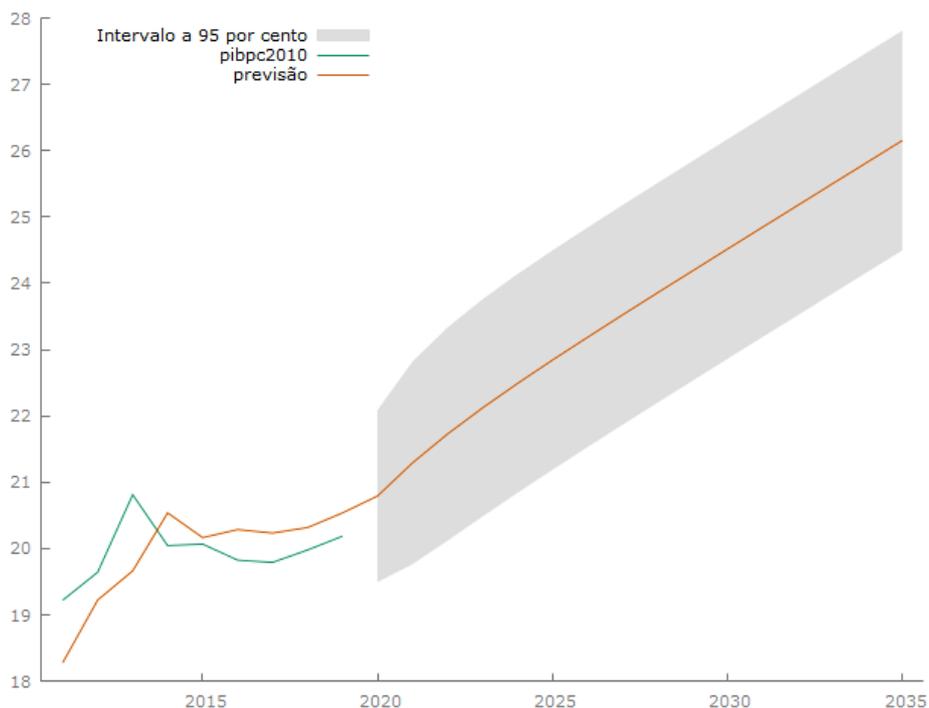
Erros padrão baseados na Hessiana

	Coefficiente	Erro Padrão	z	p-valor	
const	0,326841	0,0622199	5,253	<0,0001	***
phi_1	0,622639	0,230517	2,701	0,0069	***
theta_1	-1,00000	0,173081	-5,778	<0,0001	***
Média var. dependente	0,316580	D.P. var. dependente		0,759781	
Média de inovações	0,063886	D.P. das inovações		0,661193	
R-quadrado	0,868482	R-quadrado ajustado		0,859715	
Log da verossimilhança	-17,88895	Critério de Akaike		43,77790	
Critério de Schwarz	47,11075	Critério Hannan-Quinn		44,10919	
	<i>Real</i>	<i>Imaginária</i>	<i>Módulo</i>	<i>Frequência</i>	
AR					
Raiz 1	1,6061	0,0000	1,6061	0,0000	
MA					
Raiz 1	1,0000	0,0000	1,0000	0,0000	

FONTE: *Software* GRETL

NOTA: Elaboração do autor.

FIGURA 2 - PREVISÃO PARA PIB PER CAPITA PARANAENSE DE 2020 A 2035



Para intervalos de confiança de 95%, $z(0,025) = 1,96$

	pibpc2010	previsão	erro padrão	intervalo a 95%
2020		20,791437	0,661193	19,495522 - 22,087352
2021		21,291696	0,778885	19,765111 - 22,818282
2022		21,726515	0,819980	20,119384 - 23,333646
2023		22,120587	0,835368	20,483296 - 23,757878
2024		22,489288	0,841258	20,840453 - 24,138124
2025		22,842194	0,843530	21,188905 - 24,495483
2026		23,185263	0,844410	21,530251 - 24,840276
2027		23,522209	0,844750	21,866529 - 25,177889
2028		23,855342	0,844882	22,199403 - 25,511281
2029		24,186101	0,844933	22,530061 - 25,842140
2030		24,515381	0,844953	22,859303 - 26,171459
2031		24,843741	0,844961	23,187648 - 26,499834
2032		25,171528	0,844964	23,515429 - 26,827627
2033		25,498958	0,844965	23,842857 - 27,155059
2034		25,826166	0,844966	24,170064 - 27,482268
2035		26,153235	0,844966	24,497133 - 27,809338

FONTE: Software GRETL

O modelo 22 apresenta os resultados de um modelo ARIMA (2, 1, 2) para a série do Consumo de Energia Elétrica Não Residencial. A figura 26 apresenta as previsões para esta variável. O Consumo de Energia Elétrica Não Residencial é previsto atingir em 2035, 25.709.327 megawatts-hora (mwh).

MODELO 22 - ARIMA (2, 1, 2) PARA CONSUMO DE ENERGIA ELÉTRICA NÃO RESIDENCIAL

Modelo 9: ARIMA, usando as observações 2002-2019 (T = 18)

Variável dependente: (1-L) consumonaoresid

Erros padrão baseados na Hessiana

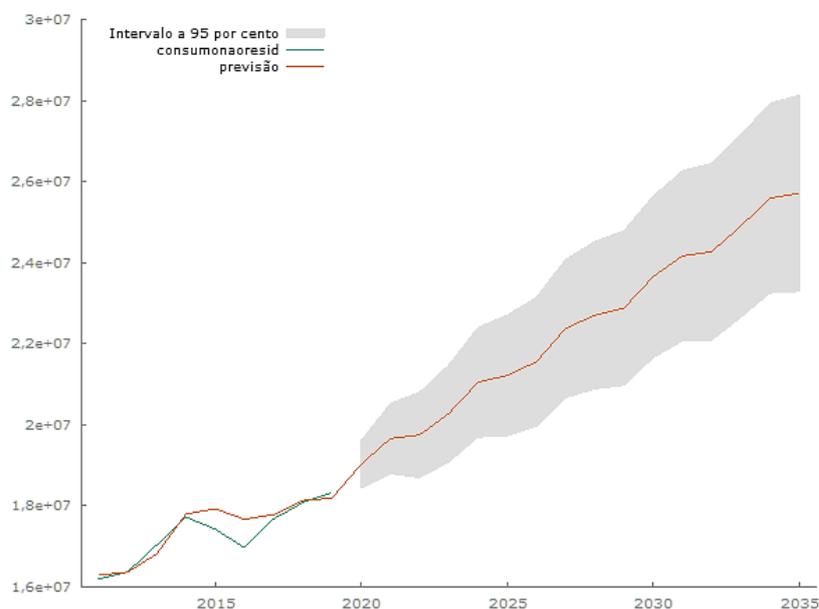
	Coefficiente	Erro Padrão	z	p-valor	
const	455991	72980,9	6,248	<0,0001	***
phi_1	-0,676643	0,182720	-3,703	0,0002	***
phi_2	-0,994980	0,0309524	-32,15	<0,0001	***
theta_1	0,764399	0,442227	1,729	0,0839	*
theta_2	0,999990	0,322627	3,100	0,0019	***
Média var. dependente	431702,0	D.P. var. dependente		417810,0	
Média de inovações	-12020,31	D.P. das inovações		297716,3	
R-quadrado	0,985695	R-quadrado ajustado		0,982629	
Log da verossimilhança	-254,5669	Critério de Akaike		521,1338	
Critério de Schwarz	526,4760	Critério Hannan-Quinn		521,8704	

	<i>Real</i>	<i>Imaginária</i>	<i>Módulo</i>	<i>Frequência</i>
AR				
Raiz 1	-0,3400	-0,9431	1,0025	-0,3051
Raiz 2	-0,3400	0,9431	1,0025	0,3051
MA				
Raiz 1	-0,3822	-0,9241	1,0000	-0,3124
Raiz 2	-0,3822	0,9241	1,0000	0,3124

FONTE: *Software* GRETL

NOTA: Elaboração do autor.

FIGURA 26 - PREVISÃO PARA CONSUMO DE ENERGIA ELÉTRICA NÃO RESIDENCIAL
PARA 2020 A 2035



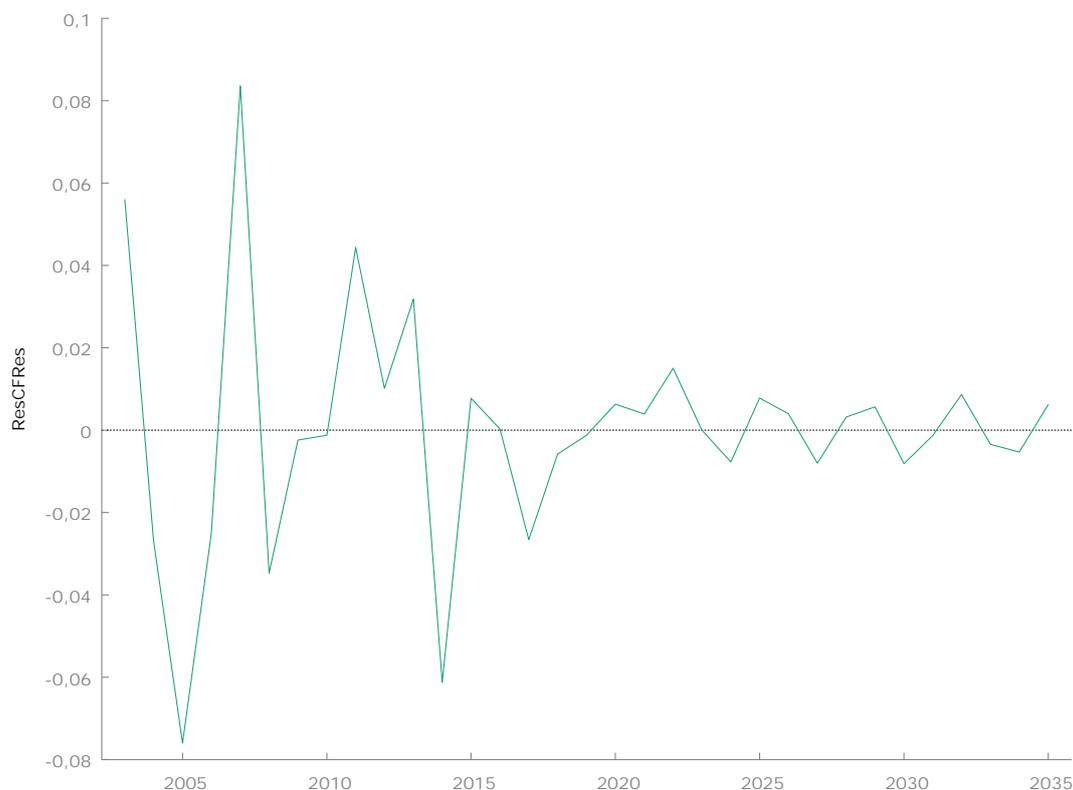
consumonaoresid	previsão	erro padrão	intervalo a 95%
2020	19019078,01	297716,308	18435564,77 - 19602591,25
2021	19651484,63	439896,705	18789302,93 - 20513666,32
2022	19740456,48	536806,746	18688334,59 - 20792578,37
2023	20269258,96	611378,864	19070978,40 - 21467539,51
2024	21041159,88	689550,275	19689666,17 - 22392653,58
2025	21210947,47	758542,037	19724232,40 - 22697662,54
2026	21546272,57	812744,310	19953322,99 - 23139222,14
2027	22368678,50	870230,528	20663058,01 - 24074298,99
2028	22696798,32	927951,664	20878046,48 - 24515550,16
2029	22874737,62	973874,431	20965978,81 - 24783496,43
2030	23646100,13	1019768,742	21647390,13 - 25644810,14
2031	24165353,79	1070088,313	22068019,24 - 26262688,34
2032	24264750,96	1112053,727	22085165,71 - 26444336,22
2033	24899084,13	1150667,041	22643818,17 - 27154350,09
2034	25589205,51	1194845,941	23247350,50 - 27931060,52
2035	25709327,72	1234679,030	23289401,29 - 28129254,15

FONTE: *Software* GRETL

NOTA: Elaboração do autor.

Utilizando as previsões apresentadas na seção 2 para os anos médios de escolaridade, estimando as equações (1) e (2), a figura 27 apresenta a estimativa para a Produtividade Total dos Fatores como os resíduos da regressão (1). Considerando a estacionariedade da variável “ResCFRes”, um modelo exploratório ARMAX foi estimado a fim de compreender de maneira exploratória relação entre a variável “anos médios de estudo” e “produtividade total dos fatores”. Foi utilizado como critério da escola dos parâmetros o menor valor para o Critério de Akaike, além da significância estatística da maior parte das variáveis do modelo. É importante salientar que dada a limitada série de dados, esses resultados devem ser tratados apenas a título de exercício exploratório.

FIGURA 27 - ESTIMATIVA DA PRODUTIVIDADE TOTAL DOS FATORES BASEADA NAS PREVISÕES



FONTE: *Software* GRETL

NOTA: Elaboração do autor.

O modelo 22 apresenta uma relação negativa, porém não significativa, entre os anos médios de estudo e a produtividade total dos fatores. Outros modelos, com critérios de Akaike maiores, apresentaram sinal positivamente mas igualmente não significantes. Ainda que aparentemente contraintuitivo, esse resultado corrobora trabalhos como Gomes *et al.* (2003) e Bonelli e Bacha (2012), que encontraram resultados nulos ou negativos na associação entre capital humano e PTF. Os autores não negam o possível efeito do capital humano sobre a produtividade, mas argumentam contra a métrica utilizada para capital humano, que mensura apenas a quantidade educacional.

MODELO 2 - ARMAX (2, 0, 1) PARA PRODUTIVIDADE TOTAL DOS FATORES

Variável dependente: ResCFRes

ERROS PADRÃO BASEADOS NA HESSIANA

	<i>Coefficiente</i>	<i>Erro Padrão</i>	<i>z</i>	<i>p-valor</i>	
const	0,00158764	0,00481352	0,3298	0,7415	
phi_1	0,471115	0,171399	2,749	0,0060	***
phi_2	-0,300215	0,182492	-1,645	0,1000	*
theta_1	-1,00000	0,0857958	-11,66	<0,0001	***
d_l_anos_medios_pnad_a	-0,110756	0,345735	-0,3204	0,7487	
Média var. dependente	2,50e-18		D.P. var. dependente	0,028967	
Média de inovações	-0,000317		D.P. das inovações	0,022292	
R-quadrado	0,400084		R-quadrado ajustado	0,338024	
Log da verossimilhança	76,95598		Critério de Akaike	-141,9120	
Critério de Schwarz	-132,9329		Critério Hannan-Quinn	-138,8908	
	<i>Real</i>	<i>Imaginária</i>	<i>Módulo</i>	<i>Frequência</i>	
AR					
Raiz 1	0,7846	-1,6478	1,8251	-0,1793	
Raiz 2	0,7846	1,6478	1,8251	0,1793	
MA					
Raiz 1	1,0000	0,0000	1,0000	0,0000	

FONTE: *Software* GRETL

NOTA: Elaboração do autor.

Barbosa Filho, Pessoa e Veloso (2010), investigando a evolução da produtividade total dos fatores (PTF) para a economia brasileira de 1992 a 2007, construíram uma medida específica de capital humano, levando em conta os diversos níveis de escolaridade e experiência. Os autores encontraram que o capital humano manteve-se estacionário durante o período analisado. Uma possível interpretação é que a expansão educacional no Brasil ocorreu em um período em que a demanda por trabalho qualificado não acompanhou a oferta, devido ao baixo crescimento da PTF.

Para os autores, a baixa contribuição do capital humano para o crescimento econômico do Brasil mesmo diante da elevação da escolaridade média, corrobora o argumento de que esses aumentos na quantidade educacional devam vir acompanhados de política de aumento da qualidade educacional. Esses resultados reforçam o discutido na seção 7.2. A heterogeneidade das escolas paranaenses em termos da qualidade sugere a necessidade de investigar esses impactos heterogêneos, ainda que a relação entre os investimentos em educação e o crescimento seja reconhecidamente relevante.

8 CONSIDERAÇÕES FINAIS

Esse estudo pretendeu estabelecer evidências a respeito dos determinantes do aumento da escolaridade, bem como da qualidade educacional no estado do Paraná, apresentando mensurando relações entre indicadores educacionais disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), dados do *Subnational Human Development Index* (SHDI) e da Pesquisa Nacional por Amostra de Domicílios (PNAD). O principal objetivo foi mensurar quanto a política pública pode interceder na qualidade da educação básica paranaense (ensino fundamental e médio). Taxas de aprendizado abaixo do adequado, evasão escolar e o desempenho aquém das metas estabelecidas, são problemas importantes que carecem de evidências a respeito de como a política pública deve agir.

Este trabalho apresentou previsões para a variáveis anos médios de estudo, considerando pessoas de 25 anos ou mais de idade, para o Paraná, de 2020 a 2035. As previsões foram feitas a partir de modelos de série-temporal ARIMA (p, d, q) considerando duas séries históricas: a primeira, da *Subnational Human Development Index* (SHDI), que segue o padrão utilizado nos relatórios internacionais do Índice de Desenvolvimento Humano (IDH); e a segunda com base em dados da Pesquisa Nacional por Amostra de Domicílios (PNAD).

De acordo com a previsão para os dados do SHDI, seguindo a tendência atual, o Paraná sairá de 8,13 anos médios de escolaridade em 2019 para 10,20 em 2035. Para a previsão utilizando os dados da PNAD, o Paraná sairia de 9,54 anos médios de escolaridade em 2019 para 11,72. O objetivo das políticas públicas educacionais deve ser, portanto, extrapolar essa tendência atual.

A evidência da ação dos professores, da gestão escolar por parte dos diretores (LEITHWOOD *et al.*, 2004; BIONDI; FELÍCIO, 2007; ROBINSON; LLOYD; ROWE, 2008; COSTA; CARNOY, 2015; GONÇALVES; LOURENÇO, 2019) e da relevância do aspecto socioeconômico dos alunos na qualidade educacional corrobora os achados deste relatório. A importância da formação dos professores e da regularidade docente em todos os níveis educacionais da educação básica e a evidência de que a distorção idade-série está relacionada a níveis piores de desempenho corrobora a evidência da literatura em defesa do programa de progressão continuada (MENEZES-FILHO *et al.*, 2008).

Os modelos estimados neste trabalho apresentaram evidência empírica da importância do nível socioeconômico no desempenho em Português e Matemática dos alunos em todos os níveis da educação básica. O mesmo é verdadeiro para a distorção idade-série, ainda que para Português nos anos iniciais do ensino fundamental esse coeficiente não seja estatisticamente significativo.

Dada a evolução histórica dos indicadores IDEB, Taxa de Aprovação e Desempenho em Português e Matemática, fica clara a viabilidade de se acessar uma melhora no IDEB através do desempenho, ainda aquém das metas estabelecidas. O acesso a computadores e a regularidade docente são ainda entraves importantes na redução das defasagens escolares tanto em termos de aprovação quanto em termos de desempenho, assim como a evidência de formação dos professores.

A partir dos resultados dos modelos exploratórios levando em consideração as variáveis dependentes “Ideb” e “Taxa de Aprovação”, a situação socioeconômica medida pelo Indicador de Nível Socioeconômico (INSE) do Inep só parece apresentar relevância no ensino médio. Esse resultado corrobora a evidência do aumento do custo de oportunidade de estudar dos alunos nesta etapa do ciclo educacional, uma vez que nesta idade os alunos enfrentam um *trade-off* entre estudar e trabalhar. No entanto, para os modelos com variável dependente “Desempenho em Português” e “Desempenho em Matemática”, o nível socioeconômico resultou relevante em termos de significância e magnitude em todos os níveis da educação básica.

Outra característica marcante para o ensino médio, tanto sobre o desempenho no IDEB como nas Taxas de Aprovação, é a quantidade de alunos em sala de aula que, quanto maiores estão associadas a menores valores daquelas variáveis dependentes. Mas para os modelos sobre desempenho em Português e Matemática, essa evidência não só não se verifica, como muda de sinal. Portanto, acompanhando os achados da literatura Matavelli e Menezes-Filho (2020) e Camargo (2012) a evidência do tamanho da sala no desempenho é inconclusiva.

A falta de dados observacionais em sala de aula dificulta a compreensão dos mecanismos pelos quais a efetividade dos professores em sala de aula pode ser melhorada. Portanto, pesquisas futuras devem levar em conta o papel da coleta de dados primários, bem como a utilização de pesquisas experimentais ou quase-experimentais que possam trazer luz às relações de causalidade existentes na sala de aula. A evidência neste relatório de não relação entre as medidas de remuneração dos professores e o desempenho e rendimento escolares, corrobora a evidência da literatura de que características observáveis como idade, educação, qualificações e regime de trabalho não explicam as diferenças de efetividade dos professores (ARAÚJO *et al.*, 2016, KANE; STAIGER, 2008).

Por fim, a última seção apresentou um panorama descritivo e exploratório da evolução temporal do percentual dos investimentos agregados em educação, ciência e tecnologia, sobre o total de despesas estaduais e das variáveis Taxa de Abandono, Reprovação, Aprovação e distorção idade-série para o Paraná, complementando a evolução temporal do desempenho em Português e Matemática. Fica patente a grande

defasagem, principalmente, no ensino médio, em relação às taxas de abandono e de distorção idade-série.

A estimação de correlação negativa e não significativa, ainda que exploratória, entre a medida de produtividade total dos fatores e os anos médios de estudo para o Paraná, corrobora o debate da importância da qualidade educacional e da relevância de se levar em conta a heterogeneidade das escolas paranaenses. Discutiu-se também a necessidade de medidas de educação que levem em conta não só a quantidade educacional, mas também a qualidade desta. Esses resultados, consonantes com a literatura, sugerem que estudos futuros se debrucem sobre essas questões, principalmente no que se refere aos impactos heterogêneos dos investimentos em educação.

REFERÊNCIAS

- ARAÚJO, M. C.; CARNEIRO, P.; CRUZ-AGUAYO, Y.; SCHADY, N. **Teacher quality** and learning outcomes in kindergarten. **The Quarterly Journal of Economics**, v.131, n.3, p.1415-1453, 2016.
- BARBOSA FILHO, F. H.; PESSÔA, S. A.; VELOSO, F. Evolução da produtividade total dos fatores na economia brasileira com ênfase no capital humano - 1992-2007. **Rev. Bras. Econ.**, v.64, n.2, 2010.
- BARBOSA, M. P.; PETTERINI, F. C.; FERREIRA, R. T. Composição do capital humano, crescimento econômico e produtividade total dos fatores nos municípios brasileiros. *In: XX ENCONTRO REGIONAL DE ECONOMIA*, 2015. Fortaleza. **Anais [...]**. Fortaleza, 2015.
- BARROS, R. P. de *et al.* Determinantes do desempenho educacional no Brasil. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v.31, n.1, p.1-42, 2001.
- BIONDI, R. L.; FELÍCIO, F. Atributos escolares e o desempenho dos estudantes: uma análise em painel dos dados do SAEB. Brasília: INEP, 2007. (Texto para discussão).
- BONELLI, R.; BACHA, E. Crescimento brasileiro revisitado. *In: VELOSO, F.; FERREIRA, P. C.; GIAMBIAGI, F.; PESSÔA, S. Desenvolvimento econômico – uma perspectiva brasileira*. Rio de Janeiro: Campus, 2012.
- BRASIL. Instituto Nacional de Estudos e Pesquisas Anísio Teixeira. **Saeb 2019**: indicador de nível socioeconômico do Saeb 2019. Brasília: INEP, 2021. Nota técnica.
- BRUNS, B.; COSTA, L.; CUNHA, N. Through the looking glass: can classroom observation and coaching improve teacher performance in Brazil? **Economics of Education Review**, 2018.
- CAMARGO, J. **O efeito do tamanho da turma sobre o desempenho escolar**: uma avaliação do impacto da "enturmação" no ensino fundamental do Rio Grande do Sul. 2012. Dissertação (Mestrado em Economia) – Programa de Pós-Graduação em Economia, Faculdade de Ciências Econômicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.
- COSTA, L. O.; CARNOY, M. The effectiveness of an early-grade literacy intervention on the cognitive achievement of brazilian students. **Educational Evaluation and Policy Analysis Mont**, v.20, n.10, p.1-24, 2015.
- GOMES, V.; PESSÔA, S. A.; VELOSO, F. Evolução da produtividade total dos fatores na economia brasileira: uma análise comparativa. **Pesquisa e Planejamento Econômico**, v.33, n.3, p.389-434, 2003.
- GONÇALVES, J.; LOURENÇO, F. C. P. Avaliação de impacto do programa ensino médio inovador sobre a frequência escolar dos alunos em Santa Catarina. **Revista CAMINE: Caminhos da Educação**, Franca, v.11, n.2, 2019.

GUJARATI, D. N. **Econometria básica**. Tradução: Denise Durante, Mônica Rosemberg, Maria Lúcia G. L. Rosa. 5.ed. Porto Alegre: AMGH, 2011.

HANUSHEK, E. A.; KIMKO, D. Schooling, labor force quality, and the growth of nations. **American Economic Review**, v.90, n.5, p.1184-1208, 2000.

JONES, C. **Introdução à teoria do crescimento econômico**. Tradução: Cristiana de Assis Serra. 3.ed. Rio de Janeiro: Elsevier, 2015.

KANE, T. J.; STAIGER, D. O. Estimating teacher impacts on student achievement: an experimental evaluation. Technical Report. **National Bureau of Economic Research**, 2008.

KRUEGER, A. B.; LINDAHL, M. Education for growth: why and for whom? **Journal of Economic Literature**, v.39, n.4, p.1101-1136, 2001.

LANGE, F.; TOPEL, R. The social value of education and human capital. *In*: HANUSHEK, E.; WELCH, F. (ed.). **Handbook of the Economics of Education**.. North-Holland, 2006.

LEITHWOOD, K. *et al.* **How leadership influences student learning**. Minneapolis; Toronto: Center for Applied Research and Educational Improvement; Ontario Institute for Studies in Education, 2004.

MAHLMEISTER *et al.* **Revisando a mobilidade intergeracional de educação no Brasil**. São Paulo: Insper, 2014.

MATAVELLI, I. R.; MENEZES-FILHO, N. A. Efeitos de tamanho da sala no desempenho dos alunos: evidências usando regressões descontínuas no Brasil. **Revista Brasileira de Economia**, v.74, n.3, 2020.

MENEZES-FILHO, N.; AMARAL, L. F. L. E. A relação entre gastos educacionais e desempenho escolar. **Insper Working Paper**, 2009.

MENEZES-FILHO, N.; VASCONCELLOS, L.; WERLANG, S. R. C.; BIONDI, R. L. Avaliando o Impacto da progressão continuada nas taxas de rendimento e desempenho escolar do Brasil. **Revista Brasileira de Economia**, FGV, 2008.

PAES DE BARROS, R.; MENDONÇA, R. SANTOS, D. D., QUINTAES, G. Determinantes do desempenho educacional no Brasil. **Pesq. Plan. Econ.**, Rio de Janeiro, v.31, n.1, p.1-42, abr. 2001.

ROBINSON, V. M. J.; LLOYD, C. A.; ROWE, K. J. The impact of leadership on student outcomes: an analysis of the differential effects of leadership types. **Educational Administration Quarterly**, Columbus, v.44, n.5, p.635-74, 2008.

SANTOS, D. D. **Impactos do ensino infantil sobre o aprendizado**: benefícios positivos, mas desiguais. 2015. Tese (Livre-Docência) – Departamento de Economia, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, 2015.

SOUZA, A. M. A relevância dos indicadores educacionais para educação básica: informação e decisões. **Revista Meta: Avaliação**, Rio de Janeiro, v.2, n.5, p.153-179, maio 2010.

TOPEL, R. Labor markets and economic growth. *In*: ASHENFELTER, O.; CARD, D. (ed.). **Handbook of Labor Economics**, North-Holland, p.2943-2984, 1999.

WOOLDRIDGE, J. M. **Introdução à econometria**: uma abordagem moderna. Tradução: Priscilla Rodrigues da Silva Lopes e Livia Marina Koeppi. São Paulo: Cengage Learning, 2019.

APÊNDICE A - AJUSTE DAS SÉRIES DOS INDICADORES ANOS DE ESTUDO E TAXAS DE CONCLUSÃO DO ENSINO FUNDAMENTAL E MÉDIO

Para realizar a projeção do indicador Anos de estudo era necessário contar com uma série histórica que contivesse um número de observações capaz de garantir maior robustez aos resultados alcançados. No Brasil, a principal fonte para o seu cálculo é a Pesquisa Nacional por Amostra de Domicílios (PNAD), realizada pelo IBGE, a qual passou por um processo de reformulação que resultou na implementação, desde 2012, de uma nova versão – a PNAD Contínua.

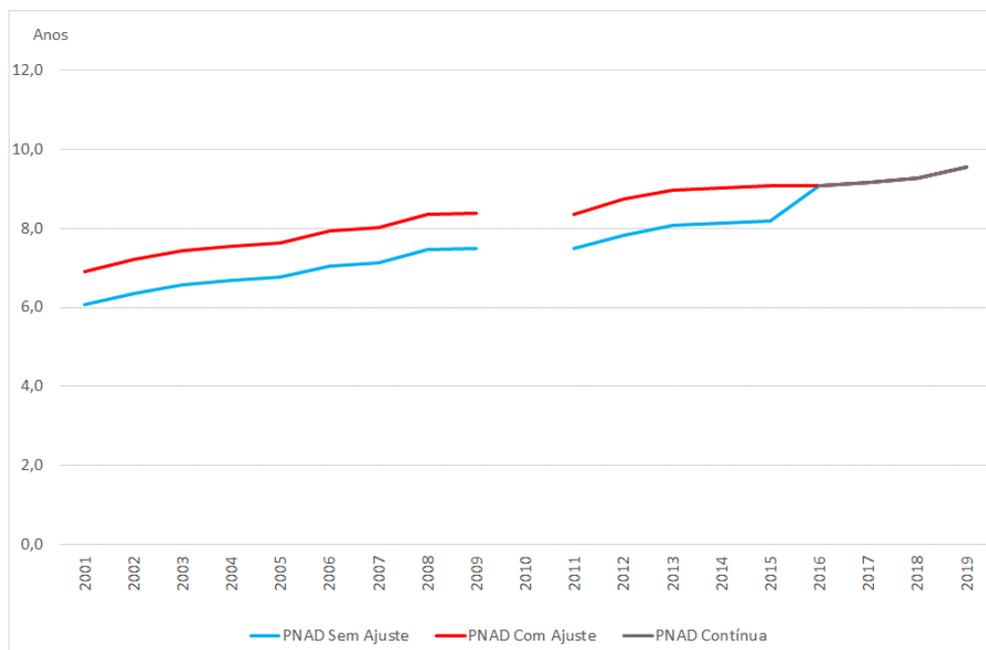
A PNAD Contínua combina a produção de indicadores mensais, trimestrais e anuais, sendo que o tema Educação é abordado em suplemento anual, com base em dados coletados durante o segundo trimestre do ano, sendo que a série histórica disponível remonta ao ano de 2016 e seguintes. No caso específico do indicador Anos de estudo ele foi padronizado para corresponder à nova seriação do Ensino Fundamental, o qual, em conformidade com o estabelecido pela Lei n.º 11.274, de 2006, passou a contar com 9 anos para a conclusão desta etapa da educação básica no país.

Segundo o IBGE, a nova padronização deste indicador consistiu em estabelecer correspondências entre a antiga e a atual seriação do ensino fundamental, com rebatimento sobre as demais etapas de ensino – médio e superior, sendo sempre considerado como referência a última série concluída com aprovação. Para o ajuste que será apresentado a seguir, cabe destacar que a nova padronização realizada pelo IBGE estabelece que as pessoas que concluíram o 1.º ano do ensino fundamental na modalidade 8 anos seriam consideradas como tendo 2 anos de estudo; e as que concluíram o 1º ano do ensino médio contabilizariam 10 anos de estudo⁶.

Para retroagir a série histórica foi necessário recorrer aos dados da versão anterior da PNAD, que foi realizada até o ano de 2015. Nela, o padrão de contagem dos anos de estudo tinha por base a antiga modalidade do Ensino Fundamental, de 8 anos. Assim, ao se combinar os dados desta pesquisa, entre 2001 e 2015, com os da PNAD Contínua, para o período 2016/2019, percebe-se uma quebra de tendência associada aos padrões diferentes de contabilização das séries do ensino fundamental (gráfico A1).

⁶ IBGE, Pesquisa Nacional por Amostra de Domicílios Contínua Notas técnicas - v. 1.7, p. 29.

GRÁFICO A1 - ANOS MÉDIOS DE ESTUDO DA POPULAÇÃO DE 25 ANOS E MAIS DE IDADE - PARANÁ - 2001-2019



FONTE: IBGE-PNAD (2001-2015), PNAD Contínua (2016-2019)

NOTA: Sem dado para 2010 pois a PNAD antiga não era realizada em anos censitários.

QUADRO A1 - COMPATIBILIZAÇÃO DO INDICADOR ANOS DE ESTUDO NA PNAD ANTIGA

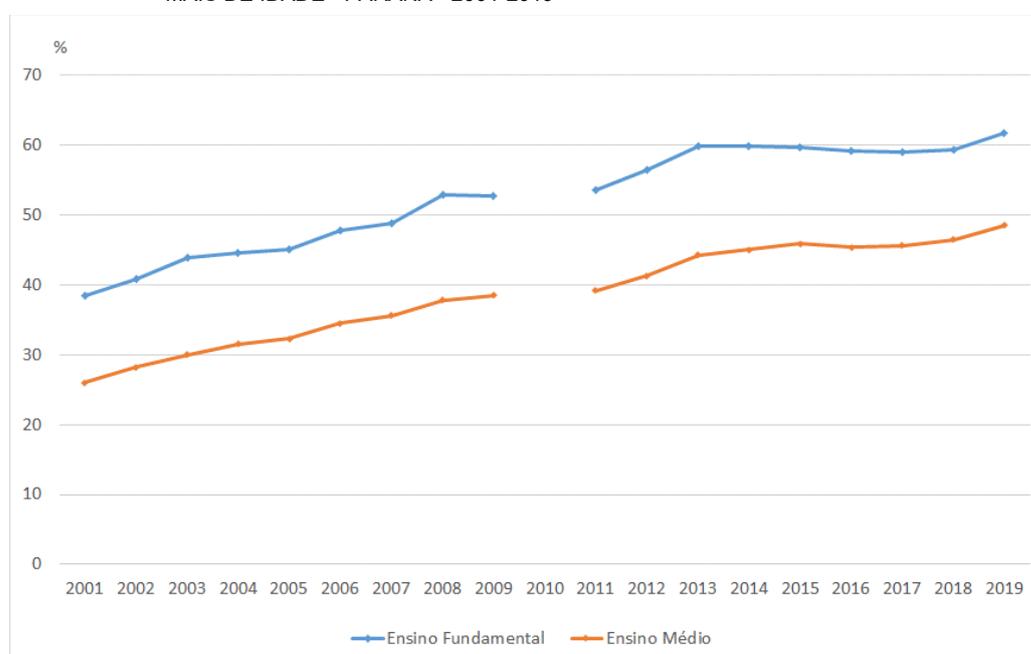
PNAD CONTÍNUA		PNAD ANTIGA		
	Anos		Anos	Ajuste
Sem instrução e menos de 1 ano	0	Sem instrução e menos de 1 ano	0	0
1 ano	1			
2 anos	2	1 ano	1	2
3 anos	3	2 anos	2	3
4 anos	4	3 anos	3	4
5 anos	5	4 anos	4	5
6 anos	6	5 anos	5	6
7 anos	7	6 anos	6	7
8 anos	8	7 anos	7	8
9 anos	9	8 anos	8	9
10 anos	10	9 anos	9	10
11 anos	11	10 anos	10	11
12 anos	12	11 anos	11	12
13 anos	13	12 anos	12	13
14 anos	14	13 anos	13	14
15 anos	15	14 anos	14	15
16 anos ou mais	16	15 anos ou mais	15	16

O ajuste realizado consistiu em atribuir para a seriação da modalidade 8 anos a mesma pontuação que o IBGE considerou nas correspondências que estabeleceu a partir dos dados da PNAD Contínua. Assim, quem tinha 1 ano de estudo na PNAD Antiga passou a contabilizar dois anos; quem tinha 9 anos (que correspondia ao primeiro ano do ensino médio) passou a contar 10 anos de estudo (quadro A1). Como resultado deste ajuste

verifica-se que as variações na série deixam de configurar uma quebra de tendência, relacionando-se à mudança de longo prazo esperada para a área educacional.

As PNADs serviram também para calcular duas taxas referentes à conclusão, pela população adulta, das etapas da educação básica – fundamental e médio. A forma ideal seria calcular essas taxas a partir da variável derivada Nível de instrução que o IBGE disponibiliza nas bases de dados da PNAD Contínua, mas que não está disponível nas bases anteriores. A solução foi utilizar a variável Anos de estudo como proxy do nível de instrução mais elevado alcançado pelas pessoas, sendo que a taxa de concluintes do Ensino fundamental considera todos aqueles que tinham 9 anos e mais de estudo e a taxa de concluintes do Ensino Médio aqueles com 12 anos e mais de estudo (gráfico A2).

GRÁFICO A2 - TAXA DE CONCLUINTES DO ENSINO FUNDAMENTAL E MÉDIO NA POPULAÇÃO DE 25 ANOS E MAIS DE IDADE - PARANÁ - 2001-2019



FONTE: IBGE - PNAD (2001-/2015) e PNAD Contínua (2016-2019)

NOTA: Sem dado para 2010 pois a PNAD antiga não era realizada em anos censitários.

Como a PNAD Contínua disponibiliza em suas bases a variável Nível de Instrução foi possível testar a aderência entre as taxas apresentadas no gráfico acima, para o período 2016/2019, construídas a partir da proxy Anos de estudo, com as que se obteriam calculando as taxas a partir do nível de instrução mais elevado. Para o ensino fundamental os resultados são iguais; já para a taxa de concluintes do ensino médio há pequenas diferenças, nunca ultrapassando a ordem de 0,5 pontos percentuais – sempre menor quando se utiliza o Nível de instrução como base de cálculo. Este fato pode estar relacionado ao fato de que uma parcela dos estudantes do ensino médio frequenta curso de

4 anos de duração e, portanto, concluiria a educação básica somente após 13 anos de estudo, aspecto melhor apreendido pela variável Nível de instrução.

A tabela A1 apresenta as séries históricas deste conjunto de indicadores, todas referentes à população de 25 anos e mais de idade. Além dos indicadores construídos com dados da PNAD, apresenta também o indicador Anos de estudo divulgado pelo Global Data Lab⁷, que realiza padronização para o cálculo do Índice de Desenvolvimento Humano (HDI) para subunidades nacionais, tornando o indicador comparável internacionalmente.

TABELA A1 - ANOS MÉDIOS DE ESTUDO E TAXA DE CONCLUINTES DO ENSINO FUNDAMENTAL E MÉDIO NA POPULAÇÃO DE 25 ANOS E MAIS DE IDADE - PARANÁ - 2001-2019

ANO	PNAD		HDI ⁽¹⁾	TAXA DE CONCLUINTES (%)	
	Sem Ajuste	Com Ajuste		Ensino Fundamental	Ensino Médio
2001	6,07	6,92	5,88	38,4	26,0
2002	6,34	7,22	6,09	40,8	28,2
2003	6,56	7,44	6,29	43,9	30,0
2004	6,67	7,54	6,29	44,5	31,5
2005	6,76	7,64	6,40	45,0	32,3
2006	7,03	7,93	6,50	47,7	34,5
2007	7,13	8,02	6,60	48,8	35,6
2008	7,45	8,35	6,81	52,8	37,8
2009	7,49	8,38	6,91	52,7	38,5
2010	-	-	7,01	-	-
2011	7,48	8,37	7,22	53,6	39,2
2012	7,82	8,73	7,42	56,4	41,3
2013	8,07	8,97	7,52	59,9	44,3
2014	8,12	9,03	7,52	59,8	45,0
2015	8,18	9,08	7,72	59,7	45,9
2016	9,07	9,07	7,83	59,1	45,4
2017	9,17	9,17	7,93	59,0	45,6
2018	9,28	9,28	7,93	59,4	46,5
2019	9,54	9,54	8,13	61,7	48,5

FONTE: IBGE: PNAD (2001-2015), PNAD Contínua (2016-2019)

NOTAS: Os ajustes decorrentes da padronização dos anos de estudo referem-se ao período 2001/2015.

Sem dado para 2010 pois a PNAD antiga não era realizada em anos censitários.

(1) Global Data Lab.

⁷ Ver <https://globaldatalab.org/shdi/>.

APÊNDICE B - TESTES DE MULTICOLINEARIDADE, HETEROCEDASTICIDADE E NORMALIDADE PARA OS MODELOS IDEB E TAXA DE APROVAÇÃO

Três hipóteses importantes do modelo de regressão múltipla são a baixa correlação entre as variáveis independentes, a variância constante e a normalidade dos resíduos. A figura B1 apresenta os resultados dos testes Fator de Inflação da Variância (VIF) a fim de testar a multicolinearidade nos modelos estimados, para os seis modelos.

FIGURA B1 - RESULTADO DOS TESTES FATOR DE INFLAÇÃO DA VARIÂNCIA (VIF)

Variable	VIF	1/VIF	Variable	VIF	1/VIF
percen~p_fun	3.19	0.313241	perc_IRD_b~o	1.02	0.977190
percen~t_fun	3.19	0.313942	doc_~1_media	1.02	0.977274
doc_~1_media	1.03	0.974409	dist_is_fun1	1.01	0.989581
perc_IRD_b~o	1.02	0.977978	percen_~1234	1.00	0.997219
dist_is_fun1	1.02	0.984539			
Mean VIF	1.89		Mean VIF	1.01	
Variable	VIF	1/VIF	Variable	VIF	1/VIF
dist_is_fun2	1.02	0.982638	esforc~n_456	1.06	0.943611
percen~p_fun	1.02	0.982870	qtd_acima_~2	1.03	0.967912
doc_~2_media	1.01	0.988784	dist_is_fun2	1.03	0.972290
perc_IRD_b~o	1.00	0.998754	perc_IRD_b~o	1.01	0.994242
Mean VIF	1.01		Mean VIF	1.03	
Variable	VIF	1/VIF	Variable	VIF	1/VIF
percen_~1234	1.01	0.988056	dist_is_med	1.00	0.995834
qtd_acima_~d	1.01	0.988056	qtd_acima_~d	1.00	0.995834
Mean VIF	1.01		Mean VIF	1.00	

Modelos Ideb

Modelos Taxa de aprovação

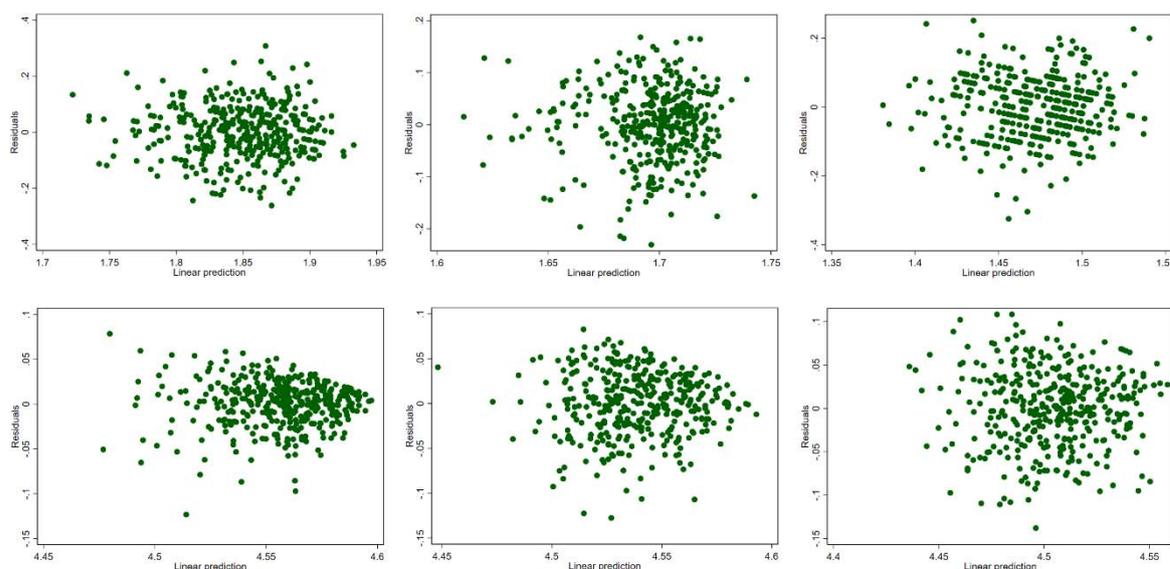
FONTE: *Software* Estatístico STATA 15.1

NOTA: Elaboração do autor.

Esse teste avalia o quanto a variância de um coeficiente aumenta se os seus preditores estiverem correlacionados. Se não houver multicolinearidade, os valores do teste serão iguais 1. Tratando-se de dados econômicos, o problema da multicolinearidade é uma questão de grau. Ainda que alguma multicolinearidade nos modelos seja aceita, o ideal é que o teste VIF esteja abaixo de 5. Valores VIF acima de 10 indicam alta multicolinearidade. Para os modelos em questão, apenas o primeiro modelo apresentou um VIF acima de 3, como esperado devido à reconstrução das variáveis contínuas em discretas.

A figura B2 apresenta evidência de não violação do pressuposto de homoscedasticidade, ou seja, os modelos parecem não apresentar o problema da heterocedasticidade, pois os resíduos não apresentam tendência quando plotados contra os valores previstos.

FIGURA B2 - TESTE GRÁFICO PARA HETEROCEDASTICIDADE



FONTE: *Software* Estatístico STATA 15.1

NOTA: Elaboração do autor.

FIGURA B3 - TESTE DE WHITE PARA HETEROCEDASTICIDADE

```
White's special test statistic : 1.497446 Chi-sq( 2) P-value = .473
White's special test statistic : 2.833111 Chi-sq( 2) P-value = .2425
White's special test statistic : 2.000617 Chi-sq( 2) P-value = .3678
White's special test statistic : 35.98901 Chi-sq( 2) P-value = 1.5e-08
White's special test statistic : 16.10103 Chi-sq( 2) P-value = 3.2e-04
White's special test statistic : 11.96984 Chi-sq( 2) P-value = .0025
```

FONTE: *Software* Estatístico STATA 15.1

NOTA: Elaboração do autor.

A figura B3 apresenta os resultados a partir do teste de White, sob a hipótese nula de homoscedasticidade. Apenas os modelos com variável dependente IDEB não rejeitam a hipótese nula a 10%. Ainda assim, dada a reduzida dimensão da base de dados, não há evidência de que há um problema de heterocedasticidade ocasionando viés sobre os resultados.

A figura B4 apresenta os resultados do teste de normalidade de Shapiro-Wilk. O teste retorna uma estatística W para cada um dos seis modelos estimados, com uma

significância associada. Valores p acima de 0,05 indicam normalidade da distribuição. Apenas os modelos 9 e 11, para a variável dependente IDEB indicam normalidade.

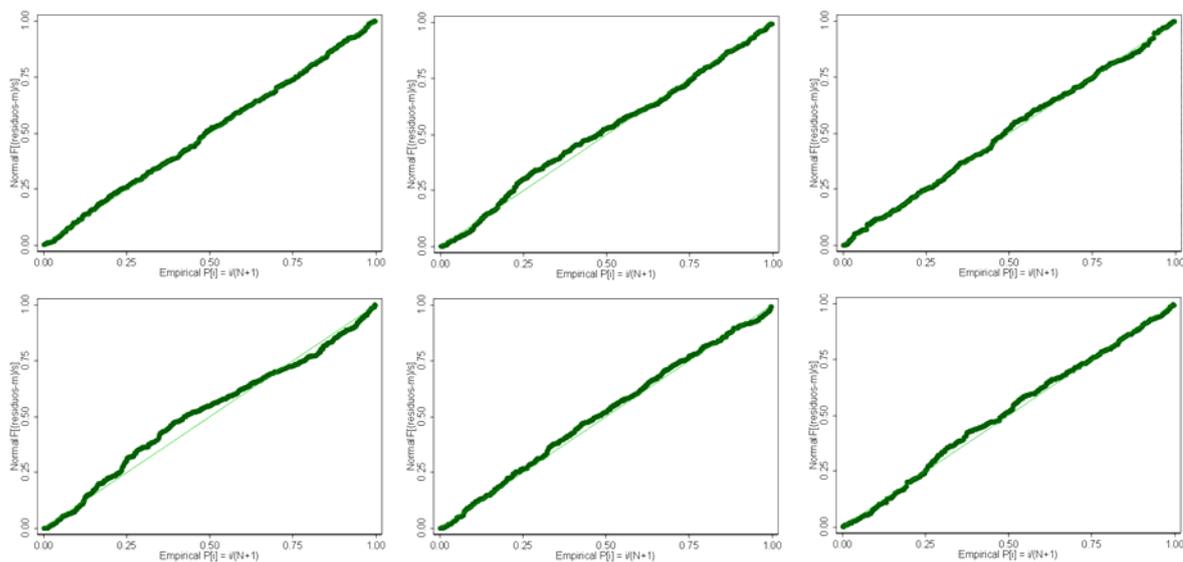
FIGURA B4 - RESULTADO DOS TESTES DE NORMALIDADE SHAPIRO-WILK

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	397	0.99709	0.795	-0.546	0.70755
Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	389	0.98814	3.183	2.752	0.00296
Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	366	0.99359	1.629	1.156	0.12388
Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	399	0.96676	9.129	5.261	0.00000
Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	397	0.98259	4.761	3.712	0.00010
Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	399	0.99287	1.958	1.599	0.05492

FONTE: *Software* Estatístico STATA 15.1

NOTA: Elaboração do autor.

FIGURA B5 - TESTE GRÁFICO DE NORMALIDADE DOS RESÍDUOS



FONTE: *Software* Estatístico STATA 15.1

NOTA: Elaboração do autor.

Dada a dimensão reduzida da base de dados, a figura B5 apresenta um teste gráfico de normalidade dos resíduos. O ajuste razoável do modelo à linha diagonal evidencia que os resíduos não estão longe de serem normalmente distribuídos.

APÊNDICE C - TESTES DE MULTICOLINEARIDADE, HETEROCEDASTICIDADE E NORMALIDADE PARA OS MODELOS DE DESEMPENHO

A figura C1 apresenta os resultados dos testes Fator de Inflação da Variância (VIF) a fim de testar a multicolinearidade nos modelos estimados, para os seis modelos.

FIGURA C1 - RESULTADO DOS TESTES FATOR DE INFLAÇÃO DA VARIÂNCIA (VIF)

Variable	VIF	1/VIF	Variable	VIF	1/VIF
percen_~1234	1.01	0.989992	percen_~1234	1.06	0.944010
perc_IRD_b~o	1.01	0.990145	dist_is_fun1	1.06	0.944744
percen~p_fun	1.00	0.999817	perc_IRD_b~o	1.01	0.988585
Mean VIF	1.01		percen~p_fun	1.01	0.993429
			Mean VIF	1.03	
Variable	VIF	1/VIF	Variable	VIF	1/VIF
percen_~1234	1.06	0.940633	percen_~1234	1.05	0.949250
doc_~2_media	1.04	0.962087	dist_is_fun2	1.04	0.962290
dist_is_fun2	1.04	0.962186	doc_~2_media	1.04	0.962392
percen~p_fun	1.02	0.981341	percen~p_fun	1.02	0.981403
perc_IRD_b~o	1.01	0.989688	Mean VIF	1.04	
Mean VIF	1.03				
Variable	VIF	1/VIF	Variable	VIF	1/VIF
qtd_acima_~d	1.14	0.877955	qtd_acima_~d	1.14	0.877221
percen_~1234	1.11	0.898209	percen_~1234	1.12	0.891447
esforc~d_456	1.03	0.973842	perc_IRD_b~o	1.06	0.946547
dist_is_med	1.01	0.986274	esforc~d_456	1.05	0.952908
Mean VIF	1.07		dist_is_med	1.03	0.967754
			Mean VIF	1.08	

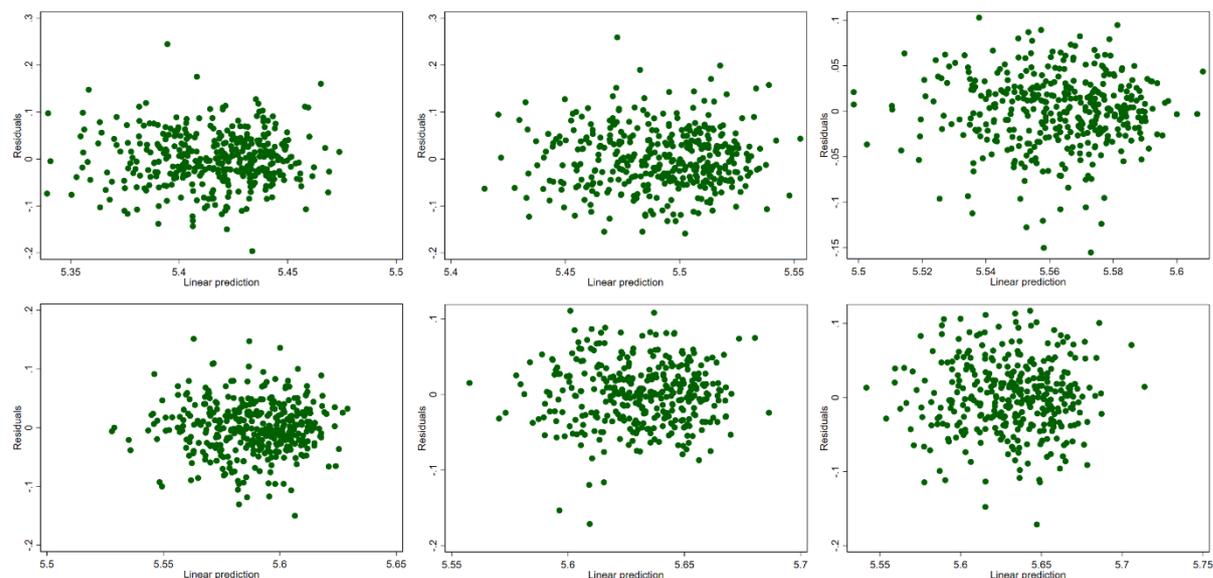
FONTE: *Software* Estatístico STATA 15.1

NOTA: Elaboração do autor.

Esse teste avalia o quanto a variância de um coeficiente aumenta se os seus preditores estiverem correlacionados. Se não houver multicolinearidade, os valores do teste serão iguais 1. O ideal é que o teste VIF esteja abaixo de 5. Para todos os modelos, os valores VIF foram em torno de 1.

A figura C2 apresenta evidência de não violação do pressuposto de homoscedasticidade, ou seja, os modelos parecem não apresentar o problema da heterocedasticidade, pois os resíduos não apresentam tendência quando plotados contra os valores previstos.

FIGURA C2 - TESTE GRÁFICO PARA HETEROCEDASTICIDADE



FONTE: *Software* Estatístico STATA 15.1

NOTA: Elaboração do autor.

FIGURA C3 - TESTE DE WHITE PARA HETEROCEDASTICIDADE

```

White's special test statistic : 3.89438 Chi-sq( 2) P-value = .1427
White's special test statistic : 1.273825 Chi-sq( 2) P-value = .5289
White's special test statistic : 7.145411 Chi-sq( 2) P-value = .0281
White's special test statistic : 3.825757 Chi-sq( 2) P-value = .1477
White's special test statistic : 5.107332 Chi-sq( 2) P-value = .0778
White's special test statistic : .7111925 Chi-sq( 2) P-value = .7008

```

FONTE: *Software* Estatístico STATA 15.1

NOTA: Elaboração do autor.

A figura C3 apresenta os resultados a partir do teste de White, sob a hipótese nula de homoscedasticidade. Apenas os modelos com variável dependente Português para os anos finais do ensino fundamental e ensino médio rejeitam a hipótese nula a 10%. Ainda assim, dada a reduzida dimensão da base de dados, não há evidência de que há um problema de heterocedasticidade ocasionando viés sobre os resultados.

A figura C4 apresenta os resultados do teste de normalidade de Shapiro-Wilk. O teste retorna uma estatística W para cada um dos seis modelos estimados, com uma significância associada. Valores p acima de 0,05 indicam normalidade da distribuição. Apenas os modelos “Desempenho em Português” para os anos iniciais do ensino fundamental e “Desempenho em Matemática” para os anos finais do ensino fundamental e Ensino Médio indicam normalidade.

FIGURA C4 - RESULTADO DOS TESTES DE NORMALIDADE SHAPIRO-WILK

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	397	0.99298	1.920	1.551	0.06042

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	397	0.99056	2.582	2.257	0.01202

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	389	0.97861	5.742	4.154	0.00002

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	389	0.99269	1.964	1.604	0.05439

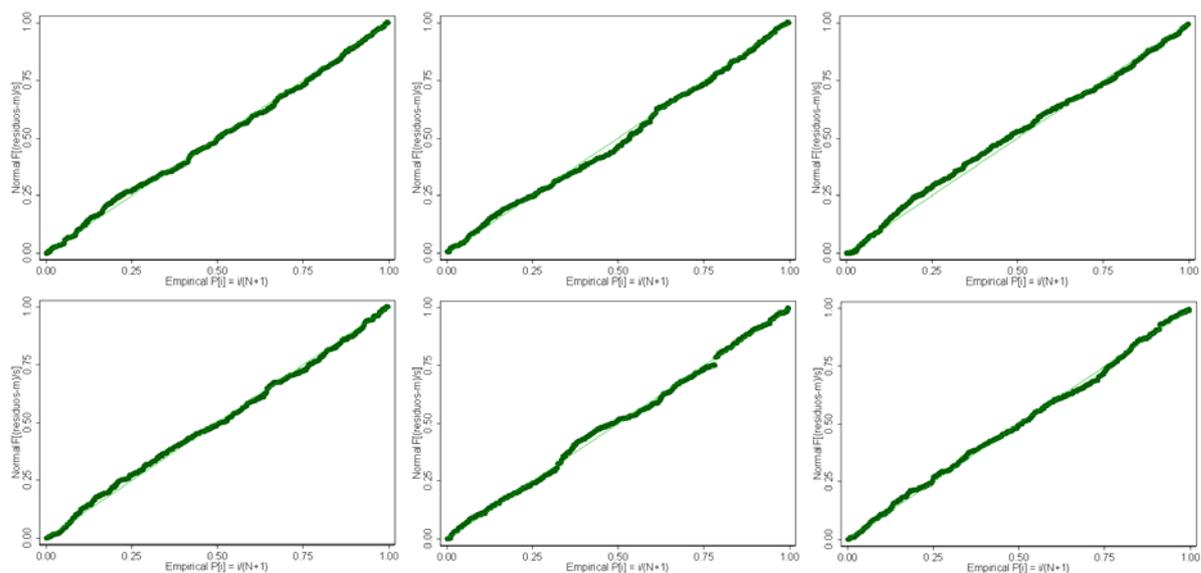
Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	366	0.98846	2.935	2.552	0.00536

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
residuos	366	0.99415	1.489	0.943	0.17292

FONTE: *Software* Estatístico STATA 15.1

NOTA: Elaboração do autor.

FIGURA C5. TESTE GRÁFICO DE NORMALIDADE DOS RESÍDUOS

FONTE: *Software* Estatístico STATA 15.1

NOTA: Elaboração do autor.

Dada a dimensão reduzida da base de dados, a figura C5 apresenta um teste gráfico de normalidade dos resíduos. O ajuste razoável do modelo à linha diagonal evidencia que os resíduos não estão longe de serem normalmente distribuídos.