Use of Bayesian Networks in Brazil High School Educational Database: Analysis of the Impact of Covid-19 on ENEM in Para between 2019 and 2022

Sandio Maciel Dos Santos*

Department of Science, Federal University of Para, Belem, Para, Brazil

Abstract

In Brazil, the National High School Examination (ENEM) serves as the primary assessment tool for students in their final year of high school. It is used for admission to higher education institutions, both public and private, as well as federally funded Brazilian government institutions. The COVID-19 pandemic in 2020 and 2022 caused significant disruptions to basic education, prompting reforms in the educational model, transitioning from in-person to remote (online) learning environments. This study aims to identify the main impacts of the COVID-19 pandemic, particularly in the state of Pará, located in the Amazon region of Brazil. To achieve this, we analyzed microdata from the ENEM editions between 2019 and 2022 to understand pre-pandemic trends and assess the impact of COVID-19 on student performance. The research findings reveal a correlation between family per capita income and participation in the ENEM among public school students, with a higher proportion of students from lower-income families not participating in the exam. Additionally, the absenteeism rate surpassed 100% compared to the previous year. Another noteworthy observation pertains to the education level of the responsible family member, indicating that higher levels of education correlate with better student performance.

Keywords: COVID-19 • Educational data mining • ENEM • Education

Introduction

In the attempt to manage confirmed cases of COVID-19, it became evident that there was no linear relationship between infections and social impact across various fundamental sectors of society, including health, education, and urban infrastructure [1]. This disparity in infection rates is closely linked to the levels of social inequality prevalent in affected populations, directly contributing to virus transmission [2-4]. Consequently, numerous efforts have been made to alleviate the repercussions of the COVID-19 pandemic, with education being one of the most profoundly affected sectors, leading to the adoption of remote learning initiatives [5-7].

Furthermore, studies such as Ariyo E, et al. [8] conducted a descriptive-analytical study utilizing binomial regression in Nigeria to elucidate the social dynamics between the surge in COVID-19 cases and remote learning performance among elementary education students. The study incorporated various social dimensions, including education level, gender, ethnicity, family composition, etc., across Nigeria's six most influential geopolitical zones. Findings from this research Ariyo E, et al. [8] revealed that family composition significantly influences the rise in COVID-19 infections, along with family income, where higher incomes correlate with better student performance.

Similarly, research such as Zhu W, et al. [9] employed electronic surveys with responses from parents of Chinese children, representing geographically diverse regions: Oriental (27.3%), Central (38.8%), and Western (33.9%)

China. The study collected social data such as age, relationship type, education level, family income, etc., to profile the students. Descriptive statistical analysis revealed that 51.3% of children experienced fatigue due to the extensive subjects and prolonged screen time, leading to discomfort and stress. Moreover, the majority (85.4%) reported that remote education failed to match the level of learning provided by face-to-face teaching.

In Brazil, various programs evaluate students' educational progress across both public and private networks, spanning from basic to higher education. The National Institute for Educational Studies and Research Anísio Teixeira (INEP) administers annual evaluation programs, serving as indicators of education quality throughout students' educational journey [10]. These evaluation programs include the Educational Census, ENADE, ENEM, etc., tailored to each educational cycle to gauge students' learning levels [11].

Given the socio-educational implications of the COVID-19 pandemic, it is evident that students completing their final years of high school during the pandemic experienced disparate performances, with a notable increase in ENEM dropouts, possibly linked to individual social inequalities [12].

Thus, this study proposes a data analysis using Data Science techniques to mine educational data from the microdata of ENEM from 2019 to 2022, examining pre-pandemic and pandemic-era COVID-19 periods, to ascertain the social dimensions exerting the greatest influence on student performance, particularly in the state of Pará.

The document is structured as follows: Section 2 outlines the materials and methods employed in the study; Section 3 details the results obtained using Bayesian networks. Finally, Section 4 provides a discussion on the findings.

Methodology

The methodology employed in this study involves the application of data science techniques, specifically Educational Data Mining [13,14] as the primary means of extracting knowledge from databases, to utilize the

***Address for Correspondence**: Sandio Maciel Dos Santos, Department of Science, Federal University of Para, Belem, Para, Brazil; E-mail: sandio.maciel@gmail.com*

acquired information for decision-making purposes. The analysis focuses on educational data about high school students and graduates to investigate the impacts of the COVID-19 pandemic.

For this study, databases from the ENEM exams of 2019 (pre-pandemic) and 2020-2022 (pandemic period) are selected. These years were chosen due to the significant increase in COVID-19 infections, along with the respective school census $K_{Q \leq 25\%}$es for the same periods, serving as microdata sources for the ENEM. This selection enables an examination of student performance amidst the challenges posed by the pandemic, particularly in the context of national exam resolutions aimed at determining the influence of school closures during periods of heightened epidemic risk [15-17].

The study categorizes performance using quartiles based on the minimum and maximum scores within each area of knowledge [18,19] as well as the number of dropouts per edition of the ξ exam. In this context, represents a vector $\overrightarrow{u}$ of numbers ordered in ascending order, Q indicates the analyzed percentile, and P ranges from 1 to 3, determining the percentiles of each quartile K (Equation 1).

$$K_Q = \frac{P(\vec{u}+1)}{4} \tag{1}$$

Unlike previous studies that have solely relied on average scores as a performance criterion [20-23] this study adopts a more comprehensive approach. It categorizes students' performance in the ENEM exam into four distinct groups, as illustrated in Table 1.

Table 1 illustrates the discretization into three groups using the quartile method: $K_{Q \leq 25\%}$ for scores below 25%, $25\% < K_{Q<75\%}$ for scores between 26% and 74%, and $K_{Q>75\%}$ for scores above 75%. The variable ξ represents the number of dropouts. This demonstration is crucial for understanding the realistic impacts of COVID-19 on sociodemographic dimensions and their influence on student performance during educational interruptions.

The microdata from ENEM for 2019 and 2022 comprise 2.24, 1.88, and 1.40 gigabytes, respectively, each containing a set of 76 variables. In total, they represent more than 14 million instances, reflecting the number of exam participants nationwide. Among the 80 indicators analyzed, 22 were selected from the dataset as they showed a higher correlation with score performance. This aimed to streamline the construction of the representative Bayesian Network (BN) [24] for the problem at hand.

The variables representing scores in different areas of knowledge are grouped into four performance analysis groups, as depicted in Table 1. For the variable Monthly Family Income (Q006), which consists of income brackets (e.g., "from R$ 0.00 to R$ 998.00"), we opted to use the lowest salary and the number of people per household (Q005) to replace the original text and group them accordingly, as per the ENEM variables dictionary [10].

The data mining phase employs the Bayesian Network technique through the PGMPY library [25] due to its ease of configuration and usability, as well as its intuitive generation of probabilistic relationships and display of Conditional Probability Tables (CPTs) for each node. For visualization, the pyAgrum API is utilized [26]. Ultimately, statistical and probabilistic inferences drawn from the microdata from ENEM and the School Census aim to compare the sociodemographic effects of successive epidemic outbreaks, confirmed cases, and deaths on student performance. This includes identifying those most likely to be affected when a public health alert is declared.

Results

In this case study, we examine the repercussions of the COVID-19 pandemic on the academic performance of high school students, specifically focusing on those in the state of Pará, in the context of the National High School Examination (ENEM) in Brazil. The analysis utilizes microdata from the ENEM editions of 2019 and 2022, alongside microdata from the corresponding School Census in each respective year.

The representative microdata for the state of Pará encompass approximately 279.593, 330.883, and 185.978 participants for each analyzed

edition, with a negligible difference of 0.05% from the official numbers. As outlined in Table 2, the analysis comprises 12 pertinent parameters selected during the data processing phase. These variables shed light on the principal socio-behavioral changes among participants, influencing overall scores during the ENEM amid the COVID-19 pandemic in Brazil (Table 2).

As depicted in Table 2, participants whose family income does not exceed approximately 1 minimum wage exhibit the highest dropout rates from the ENEM during the years 2020 and 2021 compared to the previous year. This underscores the disproportionate impact of non-seasonal respiratory epidemic outbreaks, such as the COVID-19 pandemic, on individuals with lower income, given the prolonged closures of public institutions [27-29].

Furthermore, in a separate analysis involving participants with an overall score above 75%, those attending or previously enrolled in private schools during the pandemic demonstrated superior performance compared to their counterparts from public schools. Additionally, higher maternal occupation and education levels were associated with better student performance. These findings suggest a correlation between the availability of resources, such as computers and internet access, and student performance, as illustrated in Table 3.

The analysis of data from students scoring above 75% by administrative dependence reveals disparities in performance before the pandemic. Participants from public schools, especially those with mothers having elementary education and lower occupational status, exhibited lower performance compared to their counterparts from private institutions, whose mothers had higher levels of education and engaged in occupations requiring advanced education [30,31].

This initial analytical process aimed to elucidate the influence of social parameters on the performance of participants from Pará in the ENEM. To explore how the increase in cases of respiratory syndrome during the COVID-19 pandemic affected student performance, a Bayesian probabilistic analysis was conducted [32]. This involved employing techniques such as Hill-ClimbSearch, K2Score, and VariableElimination [33] with support from the pyAgrum library to visualize the Bayesian Networks (BNs) for the years 2019 and 2022, as depicted in Figure 1.

The BN derived from the 2019 data highlights key variables significantly impacting the performance of ENEM participants, including the father's education level, family income, computer dependency, and administrative dependence in the household, as depicted in Figure 1a. This organizational structure establishes a flow of probabilistic dependencies among the selected parameters. Applying the same methodology to structure BNs for the educational data from 2020 and 2022 (Figures 1b and 1c), directly influenced by the COVID-19 pandemic, reveals a notable shift: the presence of a computer in the household is no longer a primary variable of importance.

The selection of these parameters is determined by the BN itself, which identifies the most critical conditional dependencies for analysis. Consequently, when examining the inferences drawn from these variables, it becomes evident that higher levels of father's education correlate with better performance of participants (*) (Table 4) and [34]. On the other hand, the number of participants who abandoned the exam (ξ) increased by 19%, between the pre-post COVID-19 period, for parents with primary education and 8%, between the pre-COVID-19 period -post COVID-19, for those with

Table 1. Groups of students by overall average performance in ENEM in the state of Pará.

Groups	2019	2020	2021	2022
ξ	Inf %	Inf %	Inf %	Inf\%
KQ ≤ 25%	-inf - 443	-inf - 439	-inf – 443	-inf- 484
KQ<75%	444 - 546	440 - 544	444 – 572	484 – 602
KQ ≥ 75%	557 - inf+	545 - inf+	573 - inf+	603 - inf+

Note: inf%: absent participants; -inf: scores below 25%; inf+: scores above 75%

Table 2. Percentage of participants who opted for dropout in the ENEM.

Parameters	ENEM Edition							
	2019		2020		2021		2022	
ADM Dependence	Public	15.45	Public	41.49	Public	26.78	Public	33.52
Color and Race	Brown	69.16	Brown	66.84	Brown	64.25	Brown	64.71
Mother's Education Level	Elementary	39,85	Elementary	35.51	Elementary	30.60	Elementary	32.72
Father's Education Level	Elementary	45.44	Elementary	42.37	Elementary	38.53	Elementary	24.43
Mother's Occupation	Group 2	46.07	Group 2	46.98	Group 2	47.11	Group 2	46.07
Father's Occupation	Group 1	37.24	Group 1	32.31	Group 1	28.77	Group 1	31.21
Number of People	+4	45.10	+4	41.31	+4	41.07	+4	49.45
* Family Income	[0-1] Salary	63.73	[0-1] Salary	70.00	[0-1] Salary	62.41	[0-1] Salary	67.36
Number of Bathrooms	1	87.35	1	83.74	1	82.15	1	81.51
Number of Rooms	2	51.39	2	51.89	1	82.15	2	51.02
Computer	No	87.80	No	84.10	No	80.88	No	83.90
* Internet	No	64.79	Yes	52.46	Yes	69.42	Yes	70.31

Note: ADM: Administrative; Elementary: Elementary School; High: High School; and Group: Occupation details in the Table

Table 3. The percentage of participants who obtained a score above 75% on the ENEM.

Parameters	ENEM Edition							
	2019		2020		2021		2022	
ADM Dependence	Private	67.02	Private	48.55	Private	51.08	Private	43.05
Color and race	Brown	57.43	Brown	49.39	White	47.68	Brown	49.60
* Mother's Education Level	High	28.09	High	37.58	High	34.27	High	37.50
Father's Education Level	High	27.68	High	29.24	High	30.41	High	31.74
* Mother's Occupation	Group 2	30.99	Group 4	35.97	Group 4	33.76	Group 4	36.11
Father's Occupation	Group 4	23.14	Group 4	30.55	Group 4	29.89	Group 4	29.76
Number of People	4	30.99	4	33.06	4	32.98	4	35.51
Family Income	[0-1] Salary	32.23	[0-1] Salary	20.20	[0-1] Salary	15.20	[0-1] Salary	19.64
Number of Bathrooms	1	55.78	1	39.89	1	36.85	1	43.84
Number of Rooms	2	40.08	3	41.90	1	36.85	2	42.46
* Computer	No	53.30	Yes	62.00	Yes	67.00	Yes	57.00
* Internet	Yes	69.83	Yes	88.24	Yes	95.87	Yes	92.06

Notes: ADM: Administrative; Elementary: Elementary School; High: High School; and Group: Occupation details in the Table

higher education, and during the COVID-19 pandemic in 2020, dropout rates are approximately 31% for primary education and 10% for higher education, in 2020 and in 2021 they are approximated at 14% and 9% respectively (Table 4).

An important aspect to highlight is the conditional probability between administrative dependence and the availability of a computer in the household for educational purposes. Inferences reveal a notable correlation, particularly among public school students who have access to a computer, showing a significant relationship with their performance on ENEM scores. When assessing the scores of students classified in Group $K_{Q<75\%}$, a noteworthy disparity is observed between those with and without access to computers, indicating a significant increase in performance for the former. Specifically, among private school students, there is a 13% increase, as illustrated in Table 5.

Another crucial aspect to consider is the conditional probability between administrative dependence and the availability of a household computer for educational activities. Inferences indicate a significant correlation, particularly among public school students who have access to computers, showing notable improvements in ENEM scores compared to those without access. For students classified in Group $K_{Q<75\%}$, there is a noteworthy increase in scores among those with access to computers. Specifically, private school students exhibit a 20% increase, as detailed in Table 5.

Furthermore, an analysis of participants' declared family income reveals a strong relationship between higher family income (C6) (*) and student scores, as illustrated in Table 6. Consistent with this inference, an examination of pre-established family income ranges demonstrates a decline in performance among students reporting incomes of up to 1 salary (C1). But those achieving scores of $K_{Q≥75\%}$ notably, there was a reduction of approximately 6.5% among the number of participants in this income range.

In a more detailed analysis, the impact of performance is assessed concerning the administrative dependence and family income of the participants. It becomes evident that the proportion of students in public schools decreased to Group $K_{Q≥75\%}$ when linked to incomes of up to 1 minimum wage. Conversely, the number of dropouts increased by 30%. Figure 2 provides a visual representation of the performance of participants based on family income (Figure 2).

In Figure 2, a notable increase in participants in the $K_{Q<75\%}$ Group between 2020 and 2021 from private schools is observed. This migration may be attributed to the challenges posed by remote learning during the peaks of confirmed COVID-19 cases in Brazil. In contrast, students from public schools account for the majority of dropouts in the national exam [31].

A more specific analysis of educational data from the state of Pará, focusing on the relationship between its six mesoregions and the school
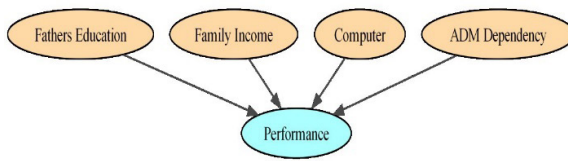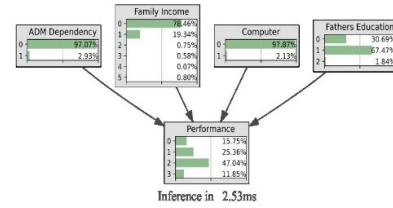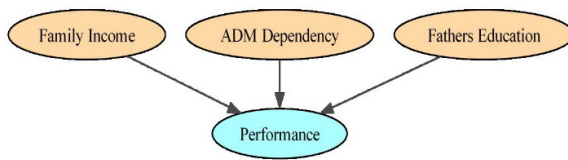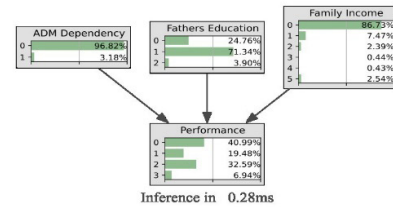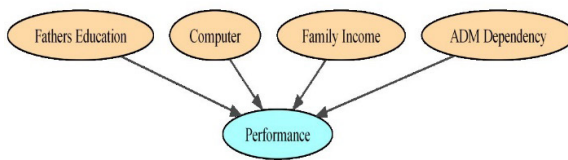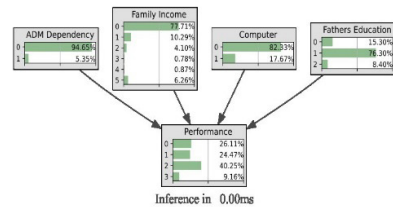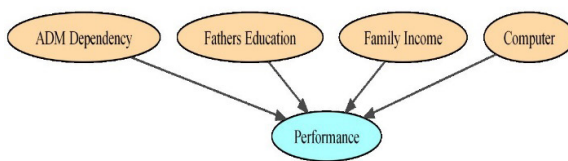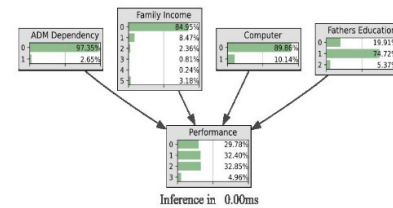
Figure 1. Bayesian network with ENEM data between 2019 and 2022, which correspond to the pre-in-post COVID-19 pandemic. Notes: Blue indicates the desired class variable, while orange represents parameters that are related to the class variable, as per the relationship test conducted by the hill-climb search algorithm and K2Score.

Table 4. Relationship between the father's level of education and the student's performance.

ENEM Edition		Father's Level of Education		
	Group	Elementary School	High School	University Education
2019	ξ	* 20.68	13.58	* 10.88
	$K_{Q \leq 25\%}$	30.55	23.05	19.25
	$K_{Q<75\%}$	41.95	49.55	47.39
	$K_{Q \geq 75\%}$	6.81	13.82	22.48 *
2020	ξ	* 51.36	38.01	29.60 *
	$K_{Q \leq 25\%}$	21.83	19.01	13.24
	$K_{Q<75\%}$	24.05	35.35	36.36
	$K_{Q \geq 75\%}$	2.76	7.64	20.80 *
2021	ξ	* 34.73	25.30	* 17.75
	$K_{Q \leq 25\%}$	29.34	24.08	19.20
	$K_{Q<75\%}$	32.11	41.82	40.86
	$K_{Q \geq 75\%}$	3.82	8.80	20.19 *
2022	ξ	* 39.76	27.90	* 18.92
	$K_{Q \leq 25\%}$	35.51	32.06	25.72
	$K_{Q<75\%}$	23.20	34.92	39.88
	$K_{Q \geq 75\%}$	1.54	5.12	15.98 *

Table 5. Relationship between administrative dependence and families that have at least one computer at home in the 2019 ENEM.

Group	Administrative Dependence		
	Public	Private	Computer
ξ	16.09	10.79	None
$K_{Q \leq 25\%}$	25.92	17.26	None
$K_{Q < 75\%}$	47.30	48.25	None
$K_{Q \geq 75\%}$	* 10.68	23.70	None
Ξ	7.21	3.51	At least one
$K_{Q \leq 25\%}$	11.34	5.29	At least one
$K_{Q < 75\%}$	34.57	24.25	At least one
$K_{Q \geq 75\%}$	* 46.87	66.95	At least one

Table 6. Relationship between family income and student performance.

ENEM Edition	Family Income							
	2019		2020		2021		2022	
Group	C1	C6	C1	C6	C1	C6	C1	C6
ξ	16.22	6.39	42.32	22.53	27.84	12.85	30.96	15.79
$K_{Q \leq 25\%}$	22.51	10.41	20.35	8.07	26.49	9.50	33.96	14.71
$K_{Q < 75\%}$	47.01	41.06	31.94	41.59	39.59	39.85	31.61	39.59
$K_{Q \geq 75\%}$	(+) 10.25	41.14 (*)	(+) 5.39	37.81 (*)	(+) 6.08	37.80 (*)	(+) 3.47	29.90 (*)

Note: C1: Income from zero up to 1 minimum wage; C6: more than 3 minimum wages
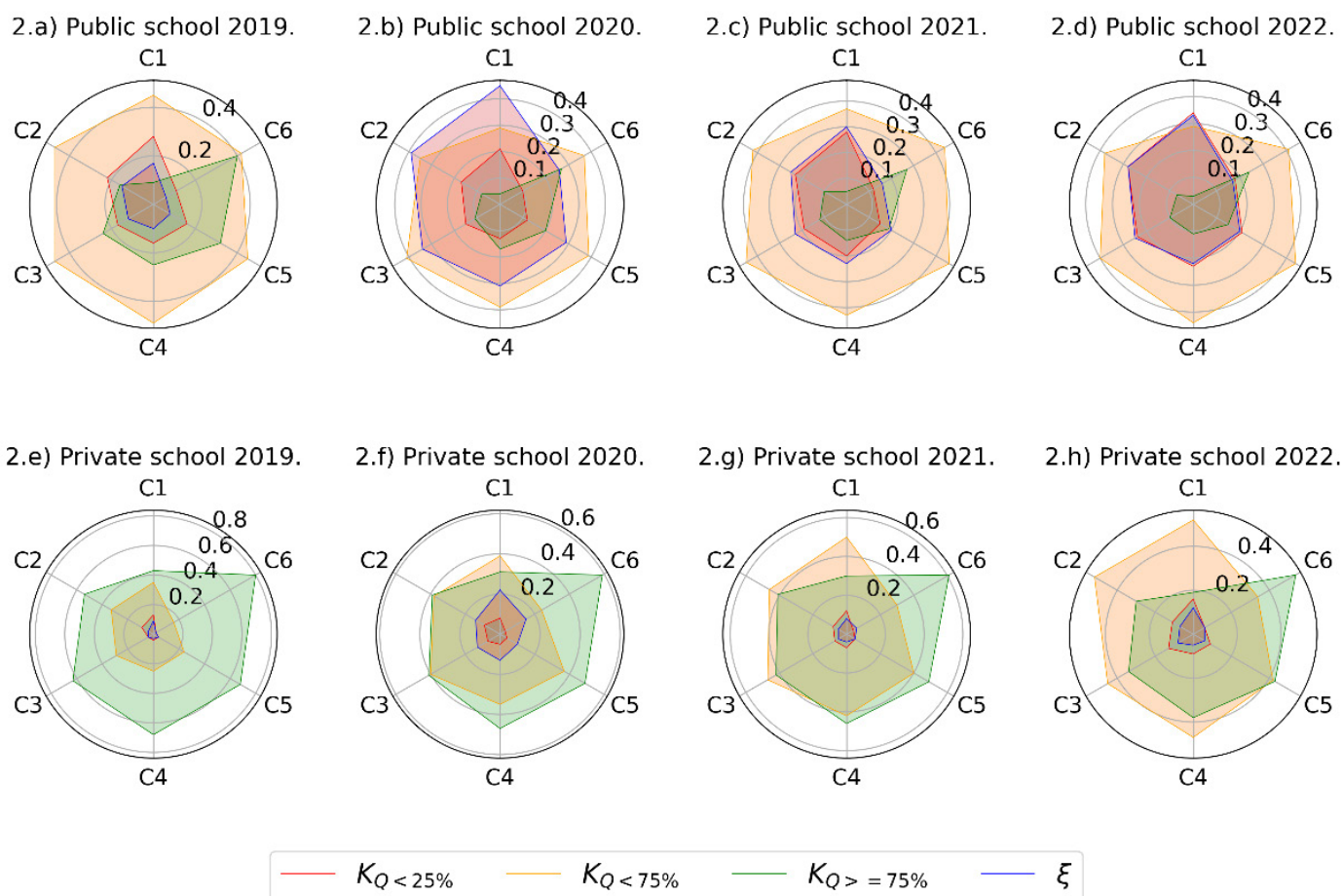

Figure 2. Performance radar of students through family income and administrative dependence. Note: C1: Up to 1 minimum wage; C2: 1.5 minimum wages; C3: 2 minimum wages; C4: 2.5 minimum wages; C5: 3 minimum wages; and C6: More than 3 minimum wages. The colors represent performance percentages: Blue: Dropouts; Orange: Score between [0-25]; Green: Score between [26-74]; Red: Score between [75-100].

3.a) Edition 2019 **3.b) Edition 2020** **3.c) Edition 2021** **3.d) Edition 2022**

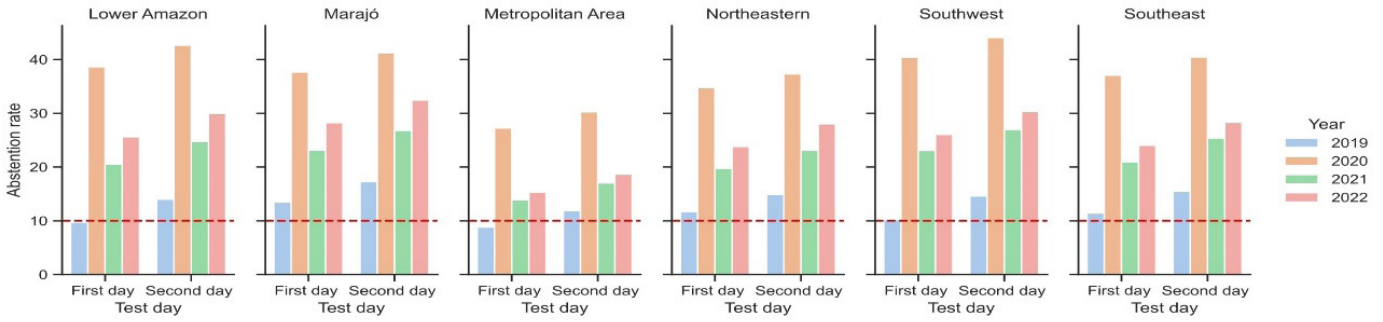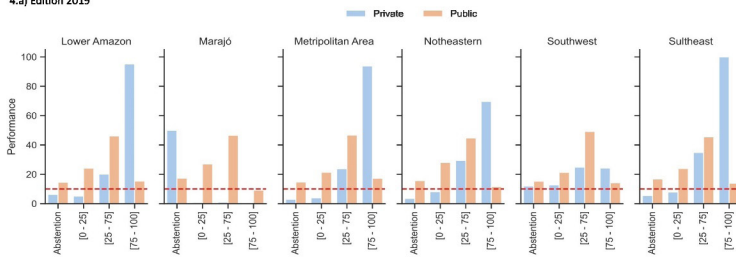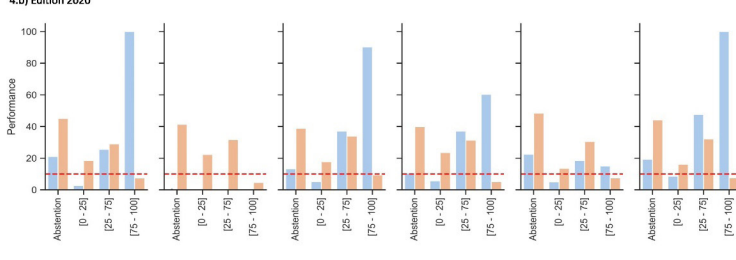**3.e) Abstention by mesoregion of Pará**

Figure 3. Abstention of students from the state of Pará by mesoregion. Note: MA: Belém Metropolitan Area; Rate: abstention rate.

4.a) Edition 2019
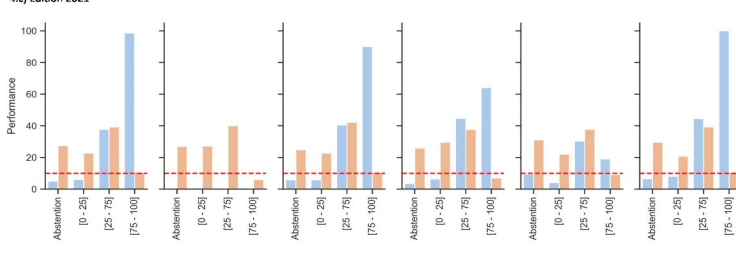
4.b) Edition 2020

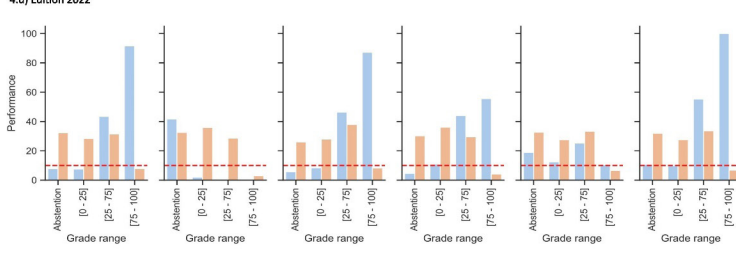**4.c) Edition 2021**

4.d) Edition 2022

Figure 4. Comparison of the performance of Pará's students by mesoregion in ENEM between 2019 and 2022.

census, sheds light on whether the impact of the COVID-19 pandemic had uniform effects on abstention rates and overall participant performance. Figure 3 displays the mesoregions of Pará for clearer reader comprehension, alongside the percentages of abstention among ENEM participants in the 2019 and 2022 editions (Figure 3).

Among the mesoregions of Pará depicted in Figure 4, it is noteworthy that only the Belém Metropolitan Region and the Northeastern Pará region experienced a significant rebound in reducing abstention rates in the ENEM during the COVID-19 pandemic between 2020 and 2021. Conversely, the other regions maintained persistently high dropout rates from the ENEM during the same period, with percentages exceeding 20% (Figure 4).

The Marajó region emerged as one of the most severely affected areas following the onset of the COVID-19 pandemic. Notably, between 2020 and 2022, students from public schools experienced a substantial decline in performance, with fewer than 10% achieving scores above 75% in the assessments. Moreover, it is crucial to highlight the surge in absent students during the ENEM exams from private schools. This trend could be linked to travel restrictions resulting from lockdowns and the closure of educational institutions on the island, as illustrated in Figure 4.

Discussion

The case study presented in this article addresses the systematic relationship between waves of COVID-19 infection and the Brazilian educational system, with a specific focus on the state of Pará. It analyzes the main impacts of interruptions in both public and private high school institutions. The analysis is based on the performance of students in the last year of high school who take the ENEM exam. The analysis of ENEM educational data revealed that students with lower per capita income (up to 1 minimum wage) faced greater difficulties in acquiring the curriculum content. This trend is evidenced by Table 6, which shows that the lower the income, the higher the percentage of dropout students. Additionally, it is important to consider other correlated factors, such as population density per square meter, where a higher number of people per area may increase the chances of virus transmission. Another relevant aspect is the relationship between the administrative dependence of schools and the level of education of parents, which also influences students' performance on the ENEM. It was noticed that students from public schools showed a higher dropout rate when their parents had lower levels of education than high school [35,36].

Conclusion

It is noteworthy that private educational institutions had a smaller impact on the teaching-learning process of their students compared to public institutions. Furthermore, despite small positive or negative fluctuations, the differences in the percentage of students who obtained scores above 75% are very small. A relevant data point is related to the performance of private institutions in Pará on the ENEM. In 2019, these institutions occupied the top positions in the ENEM ranking and maintained that position in 2020 and 2022, even with the transition to remote learning. This highlights the ability of private institutions to maintain high performance standards for their students, even in the face of challenges such as distance learning. Thus, variables such as population density and the educational level of parents also significantly influenced student performance. While private educational institutions showed less disruption in the teaching-learning process compared to public ones, consistently maintaining high performance standards in the ENEM, even during periods of remote learning.

Supplementary Information

None declared.

Conflict of Interest

The authors declare that they have no competing interests.

References

1. Green, Heidi, Ritin Fernandez and Catherine MacPhail. "The social determinants of health and health outcomes among adults during the COVID-19 pandemic: A systematic review." *Public Health Nurs* 38 (2021): 942-952.

2. Paremoer, Lauren, Sulakshana Nandi, Hani Serag and Fran Baum. "Covid-19 pandemic and the social determinants of health." *BMJ* 372 (2021):129.

3. Park, Ashley and Omer A. Awan. "COVID-19 and virtual medical student education." *Acad Radiol* 30 (2023): 773-775.

4. Julio, Silva and Marcelo Ribeiro-Alves. "Social inequalities and the pandemic of COVID-19: The case of Rio de Janeiro." *J Epidemiol Community Health* 75 (2021):975–979.

5. Alqahtani, Ammar Y., and Albraa A. Rajkhan. "E-learning critical success factors during the covid-19 pandemic: A comprehensive analysis of e-learning managerial perspectives." *Educ Sci* 10 (2020):216.

6. Godoy, Lívea Dornela, Raquel Falcoski, Roberta Monteiro Incrocci and Fabiana Maris Versuti, et al. "The psychological impact of the COVID-19 pandemic in remote learning in higher education." *Educ Sci* 11 (2021): 473.

7. Cássia, do Carmo Andrade Lisbôa and Maria Cristina do Amaral Moreira "A avaliação da aprendizagem durante o ensino remoto imposto pela pandemia da Covid-19." *Revista Contexto & Educacao* 39 (2024): e14253-e14253.

8. Ariyo, Esther, Micheal Amurtiya, Olaleye Yemisi Lydia and Ariyo Oludare, et al. "Socio-demographic determinants of children home learning experiences during COVID 19 school closure." *Int J Educ Res Open* 3 (2022): 100111.

9. Zhu, Wenting, Qianqian Liu and Xiumin Hong. "Implementation and challenges of online education during the COVID-19 outbreak: A national survey of children and parents in China." *Early Child Res Q* 61 (2022): 209-219.

10. Brasil. "National institute of educational studies and research anísio teixeira — national institute of educational studies and research anísio teixeira - INEP." (2024).

11. Neto, Nelson Weber, Raimundo C. Soares, Luciano Reis Coutinho and Ariel Soares Teles. "A Pandemia da COVID-19 impactou o ENEM? Uma Analise Comparativa de Dados dos Anos de 2019 e 2020.*" Revista Novas Tecnol Educ* 20 (2022): 223-232.

12. Dayene, Mendes Silva Gonçalves and Luiz Ismael Pereira. "Abandono escolar no ensino médio: uma análise comparativa antes e durante a pandemia em minas gerais." *Políticas Educacionais* 18 (2024):1-22.

13. Andrea, Filatro. "Data Science na Educacao: Presencial, a Distancia e Corporativa: Presencial, a Distância e Corporativa."*Saraiva Educacao* (2020).

14. Mouromtsev, Dmitry and Mathieu d'Aquin. "Open data for education: Linked, shared, and reusable data for teaching and learning." 1ˢᵗ ed. 2016 edition ed. Switzerland, *Springer* (2016).

15. Laertes, Pereira Junior, Simone Nasser Matos and Helyane Bronoski Borges. "Socio-economic status and COVID-19–related cases and fatalities." *Pub Health* 189 (2020):129-134.

16. Karakose, Turgut. "The impact of the COVID-19 epidemic on higher education: Opportunities and implications for policy and practice." *Educ Process Int J* 10 (2021):7-12.

17. Reimers, Fernando M. "Learning from a pandemic. The impact of COVID-19 on

education around the world." Primary and secondary education during Covid-19: Disruptions to educational opportunity during a pandemic (2022): 1-37.

18. Bendikson, Linda and John Hattie and Viviane Robinson. "Identifying the comparative academic performance of secondary schools." *J Educ Admin* 49 (2011):433–449.

19. Waheed, Hajra, Saeed-Ul Hassan, Naif Radi Aljohani and Julie Hardman, et al. "Predicting academic performance of students from VLE big data using deep learning models."*Comput Hum Behav* 104 (2020):106189.

1. Boneti, Lindomar Wessler and Geisa Melo Oliveira. "Enem: Análise do desempenho escolar nas edições de 2009 a 2013." *Revista Espaço Pedagógico* 24 (2017).

20. Guilherme, Ferrari Bravin, Luciana Lee and Silvia das Dores Rissino. "Mineraçao de dados educacionais na base de dados do enem 2015." *Brazilian J Prod Eng* 5 (2019):186–201.

21. Rafael, Vinicios do Carmo, Wesllei Felipe Heckler and Juliano Varella de Carvalho. "Uma Análise do Desempenho dos Estudantes do Rio Grande do Sul no ENEM 2019." *Revista Novas Tecnol Educ* 18 (2020): 378-387.

22. Fernando, Lang da Silveira, Marcia Cristina Bernardes Barbosa, Roberto da Silva. "Exame Nacional do Ensino Médio (ENEM): Uma análise crítica." *Revista Brasileira de Ensino de Física* 37 (2015):1101.

23. Murphy, Kevin Patrick. "Dynamic Bayesian networks: Representation, inference and learning." University of California, Berkeley (2002).

24. Ankan, Ankur and Abinash Panda. "Pgmpy: Probabilistic graphical models using python." *SciPy* (2015):6-11.

25. Ducamp, Gaspard, Christophe Gonzales and Pierre-Henri Wuillemin. "aGrUM/ pyAgrum: A toolbox to build models and algorithms for probabilistic graphical models in python." *Inter Conf Probabilist Graph Models. PMLR* (2020).

26. Janderson Ferreira Dutra, Joao Batista Firmino Júnior and Damires Yluska de Souza Fernandes. "Fatores que podem interferir no desempenho de estudantes no ENEM: uma revisão sistemática da literatura." *Revista Brasileira de Informatica Educacao* 31 (2023): 323-351.

27. Cesar, Augusto Andrade Ferreira, Tarcísio da Costa Lobato and Brena do Nascimento Carvalho. "Enem no Norte do Brasil: Uma análise comparativa dos determinantes do desempenho entre o ano de 2019 e o ano pandêmico de 2020." (2022).

28. Torres, Ronaldo, Mateus Machado de Pereira, Reisoli Bender Filho and Fernanda Cigainski Lisbinski. "Determinantes do desempenho dos participantes da prova do enem: Evidências para o rio grande do sul." *Desenvolvimento em Questao* 18 (2020): 352–368.

29. Luanda, Carol de Sousa Fernandes, Francisco Bezerra Mendes, Jodean Alves da Silva and Rodrigo Fonseca Da Silva. "Análise do desempenho em matemática e suas tecnologias dos participantes do ENEM 2021 em Barra do Corda, Maranhão: Uma comparação entre alunos de escolas públicas e privadas por meio de regressão logística." *Contribuciones Las Ciencias Sociales* 16 (2023): 33822-33835.

30. Diego, Navarro, Matheus Ianello, Felipe Muneratto and Graciella Watanabe. "Impactos do Conhecimento das Ciências Naturais para o Desempenho no ENEM: Considerações sobre a Desigualdade Científico-tecnológica para a Justica Social." *Revista Brasileira de Pesquisa em Educacao em Ciencias* (2021): e26002-30.

31. Mark, Chavira and Adnan Darwiche. "Compiling Bayesian networks using variable elimination." *IJCAI* 2443 (2007).

32. Koller, Daphne and Nir Friedman. "Probabilistic graphical models: Principles and techniques." *MIT Press* (2009).

33. Suhonen, Tuomo and Hannu Karhunen. "The intergenerational effects of parental higher education: Evidence from changes in university accessibility." *J Pub Econ* 176 (2019): 195-217.

34. Leandro, Morais Coelho de Oliveira, Luciane Zanin and Flavia Florio. "Professores do ensino fundamental público: Literacia em saúde e fatores associados." *Revista Contexto & Educacao* 39 (2024):e13673–e13673.

35. Laertes, Pereira Junior, Simone Nasser Matos and Helyane Bronoski Borges. "Análise dos Perfis de Alunos do Ensino Superior sobre a Realização de Aulas na Modalidade a Distância Durante Pandemia da Covid-19 Usando Algoritmos de Aprendizagem de Máquina." *Revista Novas Tecnologias na Educação* 18 (2020):336–345.