FGV EBAPE

RAP
BRAZILIAN JOURNAL OF PUBLIC ADMINISTRATION

Impact of socioeconomic variables on Enem performance: a spatial and sociologic analysis

Rafael Oliveira Melo [1]
Anne Caroline de Freitas [2]
Eduardo de Rezende Francisco [3]
Marcelo Tadeu Motokane [4]

[1] Fundação Getulio Vargas / Educação Executiva, São Paulo / SP – Brazil
[2] Universidade de São Paulo / Programa de Pós-graduação Interunidades em Ensino de Ciências Instituto de Biociências, São Paulo/ SP - Brazil
[3] Fundação Getulio Vargas / Escola de Administração de Empresas de São Paulo, São Paulo / SP – Brazil
[4] Universidade de São Paulo / Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Departamento de Biologia, Ribeirão Preto / SP – Brazil

This study aims to identify variables with the highest impact on students' performance in the National Exam of High School (Enem), per municipality. The research adopted the educational sociology perspective and used spatial statistics to analyze the dynamics of territorial influence. We analyzed Enem 2018, measuring the degree of importance of each category of variables for the average and variance of students' grades in the exam – which is separated in an essay and an objective test – and used statistical modeling and geospatial analysis, such as linear and spatial regressions. The results indicate that, for the objective test, variables like percentage of students with scholarship, income, race, schooling, and education level of students' mothers are relevant to students' performance and dispersion of grades in each municipality. For the essay, variables were similar to the objective test but with less impact on the average and variance of the grades. This explicative factor increases when a spatial component is introduced in the model for the essay grades, indicating that there are other regional factors, besides socioeconomics, impacting the performance and dispersion per municipality. The results reinforce the sociological studies on education since the socioeconomic disparities reflected in the students' performance stand out, which are also pointed out by Bourdieu (1998, 2008, 2009) in his studies of the fundamentals of school productivity of inequalities. Studies with this integrated perspective are relevant to understand the influence of municipalities' location, evidencing specific gaps and, consequently, directing and influencing public actions to overcome inequalities.
Keywords: Enem; sociology; educational performance; statistical analysis; spatial regression.

Impacto das variáveis socioeconômicas no desempenho do Enem: uma análise espacial e sociológica

O objetivo deste trabalho é identificar variáveis com maior impacto no desempenho dos municípios no Exame Nacional do Ensino Médio (Enem), tanto para a prova objetiva quanto para a redação, com base em técnicas de estatística espacial, que permite analisar a dinâmica da influência territorial, e da perspectiva da sociologia da educação. Pretende-se mensurar o grau de importância de cada categoria de variáveis para a média e a variância das notas do Enem. Para isso, utilizam-se modelos estatísticos e geoespaciais, como regressão linear múltipla e regressão espacial, a partir do Enem de 2018. Para a prova objetiva, variáveis como o percentual de estudantes com bolsa, renda, raça, escolaridade e nível instrucional da mãe são fatores relevantes para o desempenho e a dispersão das notas dos estudantes de cada município. Para a redação, as variáveis são similares às da prova objetiva, mas com menor impacto na média e na dispersão das notas. Esse fator explicativo aumenta quando introduzimos um componente espacial no modelo para as notas de redação, indicando que outros fatores regionais, diferentes dos socioeconômicos, impactam o desempenho e a dispersão dos municípios. Os resultados reforçam os estudos da sociologia da educação, ao destacarem as disparidades socioeconômicas refletidas no desempenho estudantil, fundamentos da reprodução escolar das desigualdades mencionadas por Bourdieu (1998, 2008, 2009).

DOI: http://dx.doi.org/10.1590/0034-761220200843
ISSN: 1982-3134
Article received on Decemer 08, 2020 and accepted on August 30, 2021.
[Translated version] Note: All quotes in English translated by this article's translator.

Estudos com essa perspectiva integrada são relevantes para o entendimento da influência no nível da localidade dos municípios, podendo evidenciar lacunas específicas, direcionando e influenciando ações públicas visando à superação das desigualdades.
Palavras-chave: Enem; sociologia; desempenho educacional; análise estatística; regressão espacial.

Impacto de las variables socioeconómicas en el desempeño del Enem: un análisis espacial y sociológico

El objetivo de este trabajo es identificar variables con mayor impacto en el desempeño de los municipios en el Examen Nacional de Enseñanza Media (Enem), tanto para la prueba objetiva como para la prueba de redacción, a partir de técnicas de estadística espacial que permiten analizar la dinámica de la influencia territorial, y de la perspectiva de la sociología educacional. Pretendemos medir el grado de importancia de cada categoría de variables para la media y varianza de la puntuación del Enem. A tal fin, utilizamos modelos estadísticos y análisis geoespaciales, como regresión lineal y espacial, a partir del Enem 2018. Los resultados muestran que, para la prueba objetiva del Enem, variables como el porcentaje de alumnos becarios, renta, raza, escolaridad y nivel de instrucción de las madres son factores relevantes para el desempeño y dispersión de las notas de los estudiantes en cada municipio. Para el desempeño en redacción, las variables son similares a las de la prueba objetiva, no obstante, con menor impacto en la media y dispersión de la puntuación. Cuando se introduce un componente espacial en el modelo de puntuación de redacción, el factor explicativo aumenta, señalando que existen otros criterios regionales, distintos a los socioeconómicos, que impactan sobre el desempeño y dispersión de los municipios. Los resultados refuerzan los estudios de la sociología de la educación, ya que subrayan las disparidades socioeconómicas reflejadas en el desempeño de los estudiantes, que son fundamentos de la reproducción escolar de las desigualdades analizadas por Bourdieu (1998, 2008, 2009). Las investigaciones con esta perspectiva integrada son relevantes para la comprensión de la influencia en la regionalidad de los municipios, pudiendo revelar brechas específicas y, en consecuencia, dirigir e incidir en acciones públicas orientadas a la superación de desigualdades.
Palabras clave: Enem; sociología; desempeño educativo; análisis estadístico; regresión.

1. INTRODUCTION

According to the 1988 Brazilian Constitution, "education, which is the right of all and duty of the state and the family, shall be promoted and fostered with the cooperation of society, aiming at the full development of the person, their preparation for citizenship and their qualification for work" (Constituição da República Federativa do Brasil de 1988, our translation). One of the constitution's goals is to universalize education and eradicate illiteracy, providing a new panorama for democratic and citizen education in Brazil (Cury, 2014). This perspective shows that education is proposed as one of the main tools for overcoming socioeconomic paradigms, offering opportunities toward fairer citizenship.

According to Freire (1967), education is an instrument of liberation and social change, offering the opportunity to reflect on and problematize contradictions. However, the analysis of Brazilian education from a historical and social perspective shows unequal teaching conditions and access to schools (Libâneo, 2012).

Therefore, evaluation practices are crucial to map and monitor educational performance, considering these inequalities (Menezes & Pazello, 2005). Public policies on education require a set of objective and grounded variables to establish optimal criteria, leading to indicators to evaluate the students' performance in assessments applied on a large scale (Hartman, 1999).

One of the large-scale teaching assessments in Brazil is the National Secondary Education Examination (Enem), conducted annually by the National Institute of Educational Studies and

Research Anísio Teixeira (Inep). One objective of this exam is to analyze school performance at the end of basic education and collaborate with admission to higher education and is a substitute for entrance exams for some programs and higher education institutions (HEI). It is also used as a criterion in programs offering student loans (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [Inep], 2009).

Enem is regulated in the Brazilian education law (Lei nº 9.394, de 20 de dezembro de 1996). The law states that the Union has to assess students' performance to evaluate and improve the education quality and map priorities. Enem emerged in a remarkable historical and governmental context, in which large-scale assessments started to be implemented in Latin America, building a new configuration of capital accumulation and the state's role (Silva & Carvalho, 2010).

Large-scale assessments began in the seventh century in China (Travitzki, 2013). Currently, several countries adopt this practice, such as SAT and ACT (USA), A-levels (UK), Baccalauréate (France), PAU (Spain), Abitur (Germany), Gaokao (China), PSU (Chile), HSC (Australia), and USE (Russia). Generally speaking, one of the most relevant international exams is the Programme for International Student Assessment (PISA) of the Organisation for Economic Co-operation and Development (OECD). This exam considers students' performance in reading, mathematics, and science. In 2018, Brazil presented the following results: i) Reading – OECD 487, Brazil 413; Brazilian range in the ranking: 55th and 59th; ii) Mathematics – OECD 489, Brazil 384; Brazilian range in the ranking: 69th and 72nd; iii) Science – OECD 489, Brazil 404; Brazilian range in the ranking: 64th and 67th (Inep, 2018c).

The Brazilian Enem assesses specific and more general skills and is equivalent to international exams such as the Chinese Gaokao and the US SAT (Travitzki, 2013). The exam has 180 objective questions divided into areas of knowledge, including an essay, in which students must present a dissertation-argumentative text based on a proposed subject (Inep, 2009). Enem is a national exam and an essential indicator in the Brazilian education system.

However, the factors influencing students' performance and access to education in Brazil have to be carefully analyzed, considering the persistent inequalities "among students, schools, classes of a given school, and the regions where the schools are located" (Torres, Bichir, Gomes & Carpim, 2006, our translation). These different factors affecting students are also reflected in the large-scale standardized assessments (Gremaud, Felício & Biondi, 2007; Lobo, Cassuce & Cirino, 2016).

Against this backdrop, this article aims to identify variables with a more significant impact on the students' performance in Enem per municipality, observing both the exam's objective questions and the essay. The study measures the relevance of each category of variables and the average and variance of the grades. Statistical models and geospatial analysis were used to examine data referring to Enem 2018 (microdata and school census) throughout Brazil.

This study contributes by pointing out possible measures to improve Brazilian education based on understanding variables significantly impacting the students' grades in the Enem, leading to tailor-made policies and guidelines for municipalities.

2. THEORETICAL FRAMEWORK

Student performance can be influenced by social circumstances beyond the school's organizational issues. In this sense, the sociology of education helps to understand the effect of socioeconomic and cultural contexts on education. This study examines the variables affecting students' performance based on Pierre Bourdieu's (1930-2002) contributions.

Bourdieu (1998) points out that the school, as an institution, believes in its neutrality supported by a set of absolute and global rules that supposedly guarantee the same chances for everyone. Another conception observed in this institution reinforces that assessment is an objective, fair, and democratic tool, which works for everyone. According to this traditional perspective, school failure is the inability of certain individuals to succeed in school life, leading to the myth of gift and competence (Muzzeti, 1999). For Bourdieu, the chances of educational success are not the same, and some students have more favorable social conditions to meet educational demands. Therefore, educational inequalities are rooted in the social environment, and the idea of performance based on individual's gift must be overcome.

When recognizing that social derivation plays a preponderant role in education (Bourdieu & Passeron, 2014), Bourdieu's sociological contributions help analyze the students' paths. They help realize that the dominant hegemonic classes tend to have greater educational success, differently from the popular classes (Lima, Ostermann & Rezende, 2013).

For Bourdieu, cultural capital – which is correlated to the origin of different groups – is the determining factor shaping a person's educational path:

> […] the notion of cultural capital was imposed, firstly, as an indispensable hypothesis to address the inequality of school performance of children from different social classes. It connects school success, i.e., the specific benefits that children from different classes and fractions can obtain in the school market with the distribution of cultural capital between classes and class fractions (Bourdieu, 1998, p. 73, our translation).

Bourdieu's (2008) ideas highlight that the educational system is configured based on an "aristocratic culture," in which the language in educational practices favors the ruling class, reproducing and preserving social order values. Therefore, the school contributes to the continuity of a specific culture, privileging talents and merits.

This vision extends to assessment tools, especially those used on a large scale. According to Lima (2015), when applying a single assessment tool to students from different socioeconomic realities, the tendency is for the result to reflect social characteristics, as the assessment incorporates the dominant culture. The socioeconomic strata more distant from the prevailing hegemonic culture will have unfavorable performance in these assessments, serving as a basis for a conservative school and building a regulatory tool for neoliberalism (Blini & Backes, 2019).

According to Blini and Backes (2019), this perspective based on large-scale assessments blames students for low performance, diminishing the state's responsibility of offering quality education.

For Perrenoud, these assessments adopt market-based justifications such as the demand for comparative data. This approach neutralizes the impacts of location:

> [...] system evaluations, which allow comparing data, do not have such restrictions and can "take the goals of education seriously." This means not only building another picture of the inequalities based on tests standardization but also less favorably estimating the system effectiveness (Perrenoud, 2003, p. 11, our translation).

Thus, such assessments external to the school reality are restricted to easily measurable data (content and technical expertise), giving less importance to other elements, such as "reasoning, imagination, autonomy, solidarity, citizenship, body balance, or an ear for music" (Perrenoud, 2003, p. 12, our translation). It is faster and cheaper to stick to objective and written tests restricted to cognitive disciplinary issues and technical skills.

Bourdieu emphasizes that the role of school assessment goes beyond the mere measurement of learning. It inculcates cultural and moral values in students (C. M. M. Nogueira & M. A. Nogueira, 2002). Bourdieu's theory has several contributions, focusing on a macro-social perspective and highlighting the non-neutrality of school institutions in reproducing such social stains. However, the literature points out limitations in his arguments, among them the effects of school on such inequalities – i.e., Bourdieu neglects micro-sociological contributions of the institution. Indeed, there are divergences in how each school/teacher intervenes in this process, as argued by authors examining the school effect (Soares, 2004).

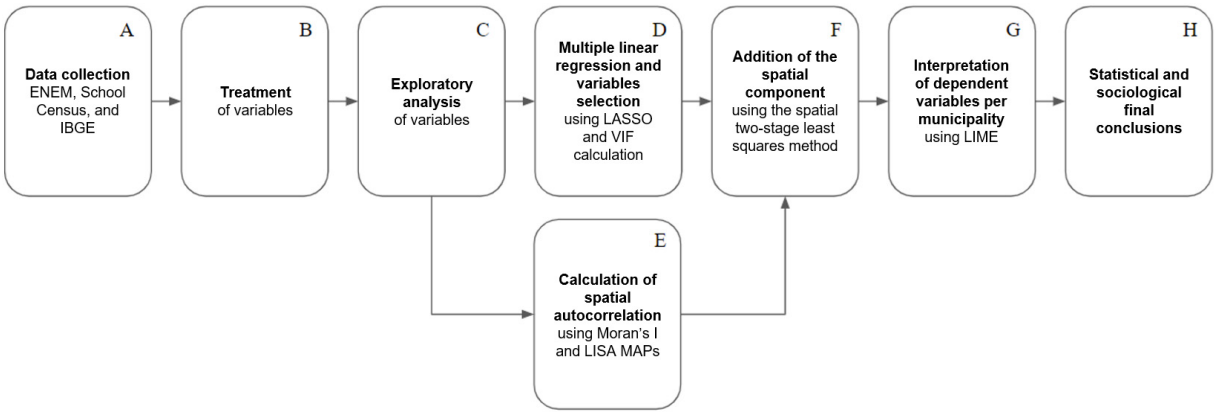
However, longitudinal studies are needed to observe this issue in depth (Alves & Soares, 2007). Despite the data pointing out the school effect, the macro-social factors are evident, especially in a country where ethnic, economic, and social inequalities play an imperative role in the educational system (Louzano, 2013). Furthermore, Arias (2009) highlights that, despite the contributions of school effect studies, 65% to 95% of the variation in school grades are explained by contextual factors – compelling variables that affect school performance and are external to the institution.

Given such a sociological panorama, it is crucial to identify and understand the social concepts that permeate society and influence the school context, observing the role of institutions in reinforcing or transposing this panorama (Silva, 2017).

3. METHODOLOGY

We used data from Enem, the Brazilian School Census, and the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística [IBGE], 2018, 2019; Inep, 2018). The variables were treated for exploratory analysis, submitted to linear regression and variable selection. The exploratory analysis allowed calculating the spatial autocorrelation and produced the linear regression model with the spatial component. Then, an analysis per municipality was conducted, using a model for interpretation of the mapped dependent variables. Finally, we discussed the data based on a statistical and sociological point of view. Figure 1 shows the research steps, and sections 3.1 to 3.5 detail the data analysis and discussion.

FIGURE 1 METHODOLOGICAL STEPS (A-H METHODOLOGICAL PATH AND ANALYTICAL TOOLS)



Source: Elaborated by the authors.

3.1. Data collection and sample

Data from three sources were used: microdata from Enem 2018 (Inep, 2018b), microdata from the 2018 School Census (Inep, 2019a), and statistics from IBGE (2018).

Enem's microdata provided the grades of each area of knowledge and essay-writing per student. This source also offered data for socioeconomic variables, obtained through a questionnaire the students fill out when registering to do the exam (Inep, 2018b).

The School Census microdata provided information about infrastructure and subjects offered by schools, such as availability of rooms, laboratories, canteens, electronic equipment, 3rd-year classes with lessons for each school subject – for example, Portuguese, mathematics, Spanish, and English – and the number of administrative staff per school (Inep, 2018a).

The IBGE data were used to understand the country's socioeconomic factors (IBGE, 2018). IBGE offered data for the variables Gini Index, Gross Domestic Product per capita (GDP per capita), and percentage of population below the poverty line.

3.2. Variables selection and treatment

Enem data were collected per student, whereas the School Census and IBGE statistics were collected per municipality. The first part of the data treatment was understanding each source's data according to variables.

The next step consisted of processing and transforming the data through feature engineering, creating dummies and new variables derived from the original data sources (Section 3.1). The final database presents information on 79 variables collected at the municipality level. Information was found for 5,548 of the 5,570 municipalities in Brazil (IBGE, 2019), which represents 99.6% of the total.

Dependent variables are based on the average and variance for the grades students obtained in Enem, per municipality: a) average grade of the objective test; b) variance of the objective test;

c) average grade of the essay; d) variance of the essay. The independent variables used to predict the average grades and variances of Enem were classified into five categories: economic, race, mother's education, scholarship, and school infrastructure. A color was attributed to each category of variables to facilitate identifying them in the charts and maps (Box 1).

BOX 1 DESCRIPTION OF THE CATEGORIES OF DEPENDENT AND INDEPENDENT VARIABLES AND THEIR COLORS IN THE CHARTS AND MAPS

Categories of variable	Color used in charts and maps	Example
Average and variance for the grades students obtained in Enem, per municipality	Brown	Average grade of the objective test
Economic variables	Red	Percentage of students from high-income households
Racial variables	Purple	Percentage of black students
Mother's education	Blue	Percentage of students whose mothers completed elementary school
Scholarship variables	Orange	Percentage of students that received a scholarship during high school
Variables of school infrastructure	Green	Percentage of schools with a canteen

Source: Elaborated by the authors.

The average represents the students' performance in Enem per municipality, while the variance measures the dispersion of the average, representing the amplitude of distance between the students' grades in relation to the municipality average. Dependent variables are presented in brown.

3.3. Exploratory Analysis

We carried out an exploratory analysis of the dependent and independent variables to understand their behavior and raise hypotheses to be proven through statistical techniques. The method used for this exploratory analysis was the development of chloroplectic maps at the municipal level, with the Jenks natural breaks optimization (Jenks, 1967), with six ranges (0 to 5). The ranges were differentiated in the charts by different tones of the colors used for each category of variables (Box 1).

In addition, to explore municipality location, we analyzed the municipality of Itapipoca, a city in the North of the state of Ceará. The study of Itapipoca allows characterizing in which range the

municipality is positioned when observing the variables presented above. Afterward, the local interpretation of the regression model was made, using this municipality as an example.

3.4. Statistical Techniques

This study used five statistical techniques: multiple linear regression, least absolute shrinkage and selection operator (Lasso), variance inflation factor (VIF), spatial two-stage least squares (STLS), and local interpretable model-agnostic explanations (LIME).

The multiple linear regression model was applied to the average and the variance of the objective test and essay grades per municipality: average grade of the objective test, average grade of the essay, variance grade of the objective test, and variance grade of the essay.

The Lasso method was used to conduct the multiple linear regression model (Tibshirani, 2017). The method helped select the independent variables that influence the average variables and the variance of the essay and objective tests. The K-FOLD method was used (with k=5) to avoid overfitting (Stone, 1974).

The VIF calculation (Fox & Monette, 1992) was performed to assess and eliminate multicollinearity, and the calculation for the variables selected by Lasso, eliminating those that presented a VIF greater than 10. Box 3 shows the adjusted R^2 of the final regression models of all dependent variables, while Graph 1 (items A to C) shows the importance of each independent variable.

Moran's I (1948) was used for a descriptive analysis of the dependent variables' spatial component (Silva, Rebouças, Abreu & Ribeiro, 2018). The Queen contiguity matrix of order 1 was selected as the neighborhood matrix. We used the statistical technique Lisa maps (local indicator of spatial association), which allowed us to observe the formation of spatial clusters where there is a relationship between the variables of neighboring municipalities (Silva, Rebouças, Abreu & Ribeiro, 2018).

The selected variables were subjected to the spatial regression model, using the STLS method (Anselin, 2012), with the same neighborhood matrix used to calculate Moran's I.

The Lime method (Ribeiro, Singh & Guestrin, 2016) was used to understand the variables that positively and negatively impact the average and variance of each of the municipalities. As an example, we observed the municipality of Itapipoca in the state of Ceará.

3.5. Tools adopted

Data treatment was conducted using the R 4.0.2 software (R Development Core Team, 2020). These tools contributed to carrying out cross-sections between the three data sources – Enem, School Census, and IBGE, as mentioned in section 3.2).

Exploratory analysis and linear, spatial, and local explanation regression models were conducted with the open-source software Python (Python Software Foundation, 2020). In addition to the Scikit Learn package, for linear regression, Python's pysal, geopandas, and esda were used for the spatial statistics analysis, and matplotlib to build the charts.

4. RESULTS

4.1. Exploratory Analysis

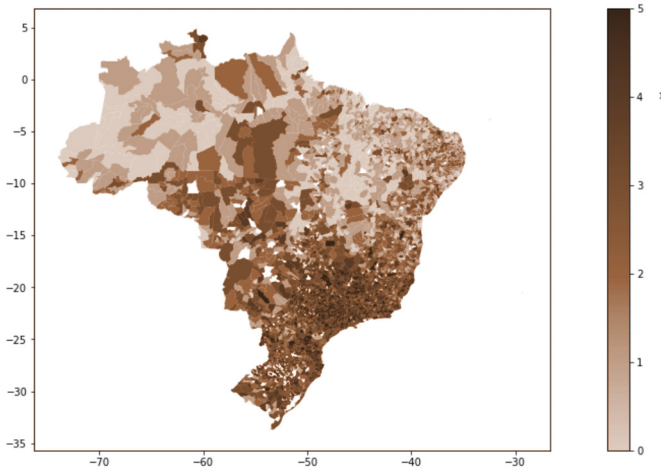
Figures 2 and 3 show the chloropletic maps created using the Jenks natural breaks optimization method (Jenks, 1967) (as mentioned in section 3.3). The figures show histograms of the number of municipalities for each of the ranges defined by the algorithm.

Figure 2 represents two of the four dependent variables analyzed in this study: average grade of the objective test and average grade of the essay. Figure 3 shows three of the 79 independent variables used in this study: percentage of students from high-income households (representing an economic variable); percentage of students whose mothers work in highly specialized jobs (representing a variable related to mothers' education); and percentage of schools with a canteen (representing a variable related to school infrastructure).

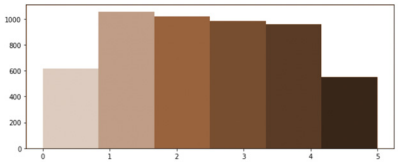
For dependent variables, there is a large regional difference in the average grade for the objective test (Figure 2 A.I and A.II). The average grade for the essay (Figure B.I and B.II) shows the same regions, but the concentration of municipalities in range 4 and 5 in the South and Southeast regions is not as large as the average for the objective test.

FIGURE 2 MAPS (LEFT) AND HISTOGRAMS WITH FREQUENCY DISTRIBUTIONS (RIGHT) OF DEPENDENT VARIABLES (RANGE 0, LESS INTENSITY, TO RANGE 5, MORE INTENSITY)

A.I) Average of grades obtained in Enem's objective test



A.II) Histogram of A.I

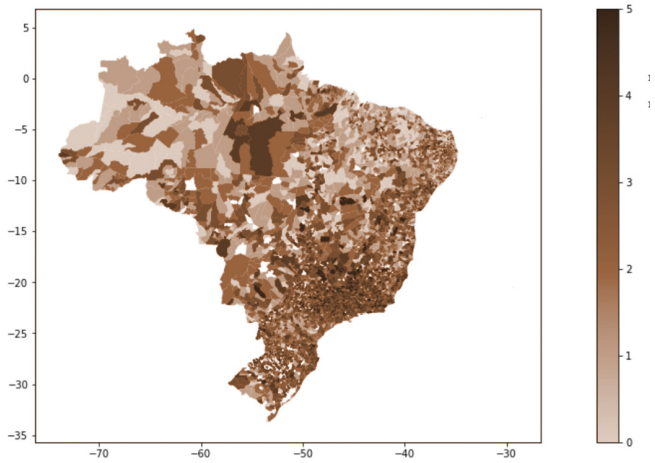


Average grade of the objective test:

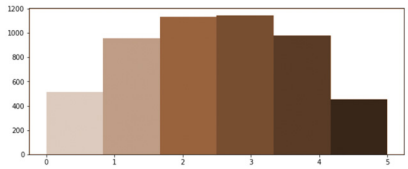
Range 0: 439.2 - 477.5	Range 3: 507.5 - 522.7
Range 1: 477.5 - 492.9	Range 4: 522.8 - 541.6
Range 2: 492.9 - 507.4	Range 5: 541.7 - 586.2

Continue

B.I) Average of grades obtained in Enem's essay



B.II) Histogram of B.I



Average grade of the essay:

Range 0: 208.0 - 399.2	Range 3: 474.1 - 509.1
Range 1: 399.4 - 440.8	Range 4: 509.1 - 553.3
Range 2: 440.8 - 474.1	Range 5: 553.6 - 694.3

Note: Maps elaborated based on the Jenks natural breaks classification method (Jenks, 1967).
Source: Elaborated by the authors.

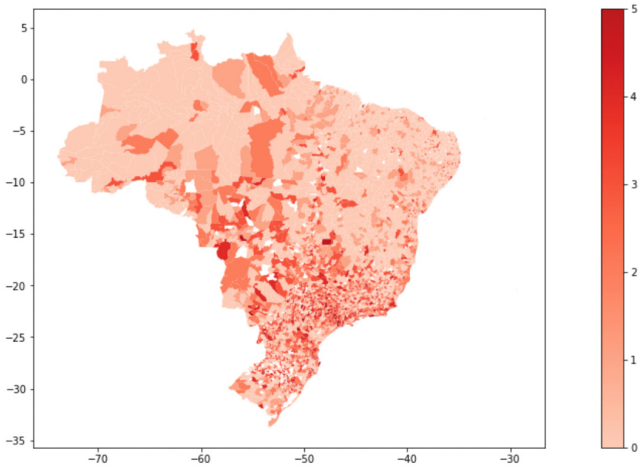
For the independent variables, Figure 3 shows that the percentage of high income families is higher in municipalities in the South, Southeast, and southern part of the Central West regions of Brazil (Figure 3, AI and A.II). This result suggests a correlation with the average grades obtained by students in Enem's objective test and essay, to be proven in the regression calculation in section 4.2 below.

Differently, the variable "percentage of students whose mothers work in highly specialized jobs" (Figure 3, B.I and B.II) stands out in the municipalities in the North of the country, with a high number of municipalities in ranges 4 and 5.

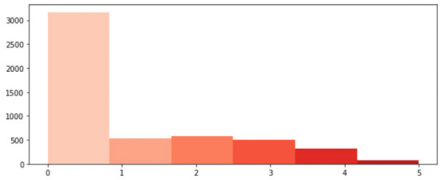
For the variable "percentage of schools with a canteen" (Figure 3, C.I and C.II), the North and Southeast regions stand out, with a large number of municipalities with a high rate of schools with canteens.

FIGURE 3 MAPS (LEFT) AND HISTOGRAMS WITH FREQUENCY DISTRIBUTIONS (RIGHT) OF INDEPENDENT VARIABLES (RANGE 0, LESS INTENSITY, TO RANGE 5, MORE INTENSITY)

A.I) Percentage of high-income families



A.II) Histogram of A.I

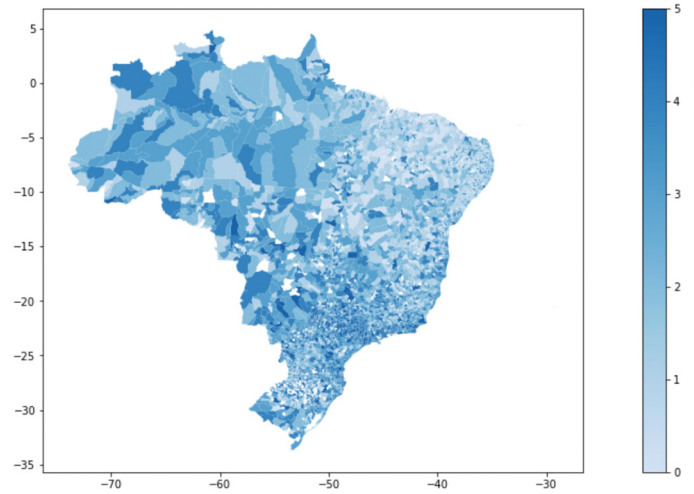


Percentage of high-income families:

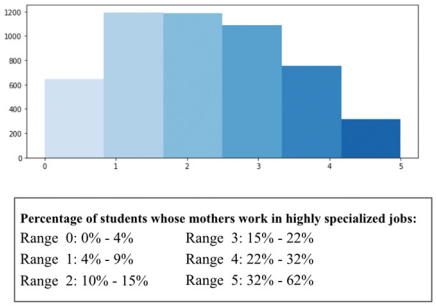
Range 0: 0% - 1%	Range 3: 4% - 6%
Range 1: 1% - 2%	Range 4: 6% - 11%
Range 2: 2% - 4%	Range 5: 11% - 26%

Continue

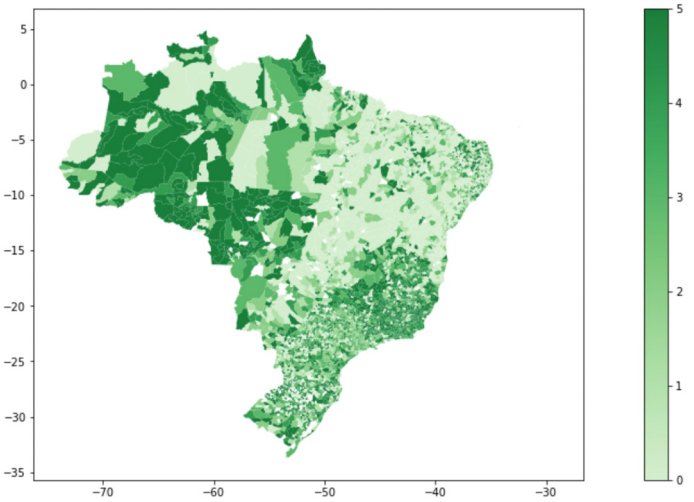
B.I) Percentage of students whose mothers work in highly specialized jobs



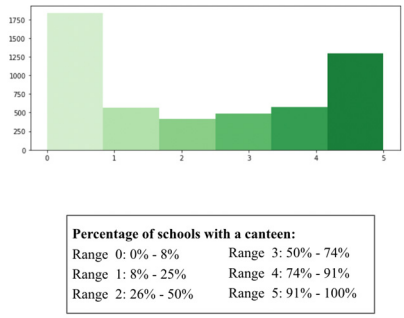
B.II) Histogram of B.I



C.I) Percentage of schools with a canteen



C.II) Histogram of C.I



Note: Maps elaborated based on the Jenks natural breaks classification method (Jenks, 1967).
Source: Elaborated by the authors.

Seeking to highlight a local example and understand how these variables behave at the municipality level, we observed the municipality of Itapipoca in the state of Ceará. Box 2 shows how this municipality is positioned in each of the variables described in Figures 2 and 3. Itapipoca falls into range 1 for the "average of grades students obtained in Enem's objective test and essay," i.e., among the lowest average scores in the country.

For the independent variables, the municipality is range 0 for "percentage of high-income families"; and in range 1, for "percentage of students whose mothers work in highly specialized jobs" and "percentage of schools with a canteen".

BOX 2 **RESULTS AND RANGE OF VARIABLES PRESENTED IN MAPS OF FIGURES 2 AND 3 FOR THE CITY OF ITAPIPOCA (CEARÁ).**

Variable	Result	Range
Average grade of the objective test	481.9	1
Average grade of the essay	417.8	1
% of students from high-income households	0.4%	0
% of students whose mothers work in highly specialized jobs	8.%	1
% of schools with a canteen	24.3%	1

Source: Elaborated by the authors.

4.2. Multiple Linear Regression and Selection of Variables

Box 3 presents the adjusted R^2, obtained in each of the four multiple linear regression models – one for each of the dependent variables – after using the variable selection methods explained above.

BOX 3 **R2 OF MULTIPLE LINEAR REGRESSION AFTER APPLYING LASSO AND EXCLUDING MULTICOLLINEAR VARIABLES VIA FIG**

Model	R^2 adjusted
Average grade of the objective test	0.68
Variance grade of the objective test	0.37
Average grade of the essay	0.37
Variance grade of the essay	0.10

Source: Elaborated by the authors.

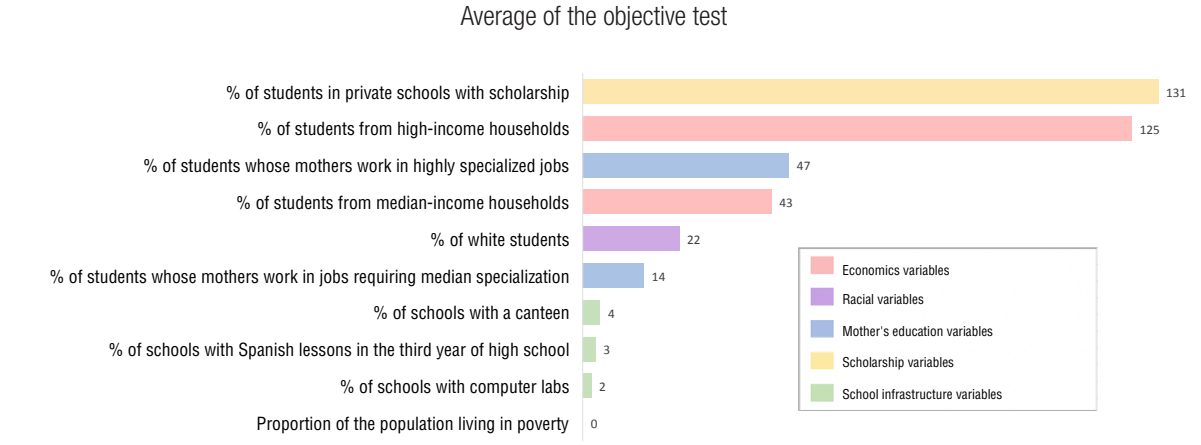
The average of the objective test is explained by the selected independent variables (68%), while its variance is moderately explained (37%). In other words, the socioeconomic variables selected by the model have a significant impact on the students' performance per municipality, represented by the average. The variables have less impact on the dispersion of students' grades per municipality represented by the variance.

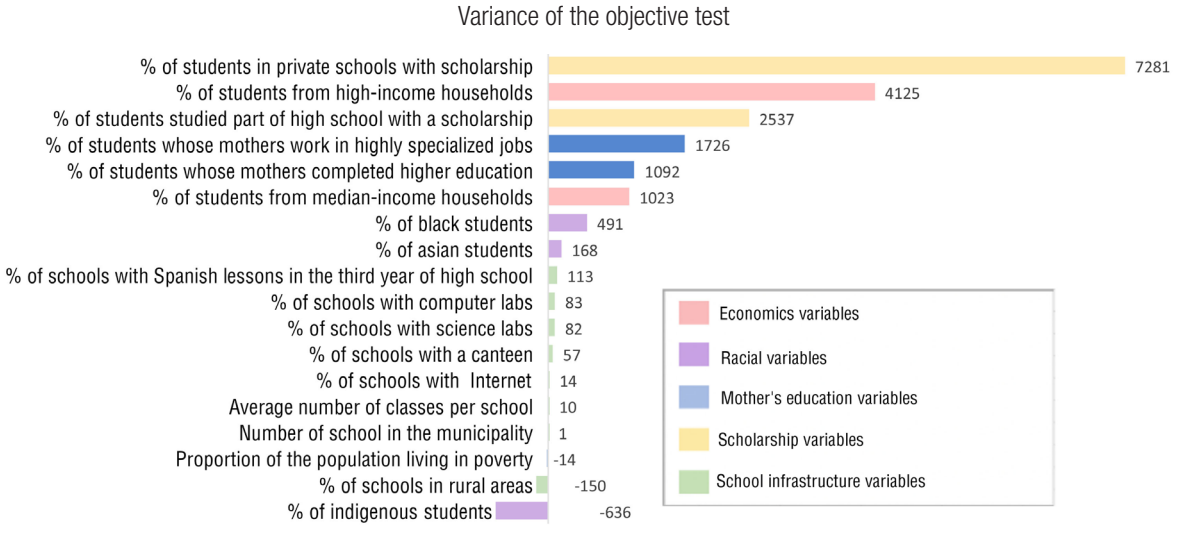
For the essay, the average is partly explained by the selected variables (37%), while the variance is poorly explained (10%). Thus, the economic variables have a relative impact on students'

performance per municipality, represented by the average, and little impact on the dispersion of student grades per municipality, represented by the variance. There is an indication that other factors also influence the variance of the essay grade. No analysis was performed on the importance of independent variables for the dispersion of essay grades due to the low R^2 found (Box 3).

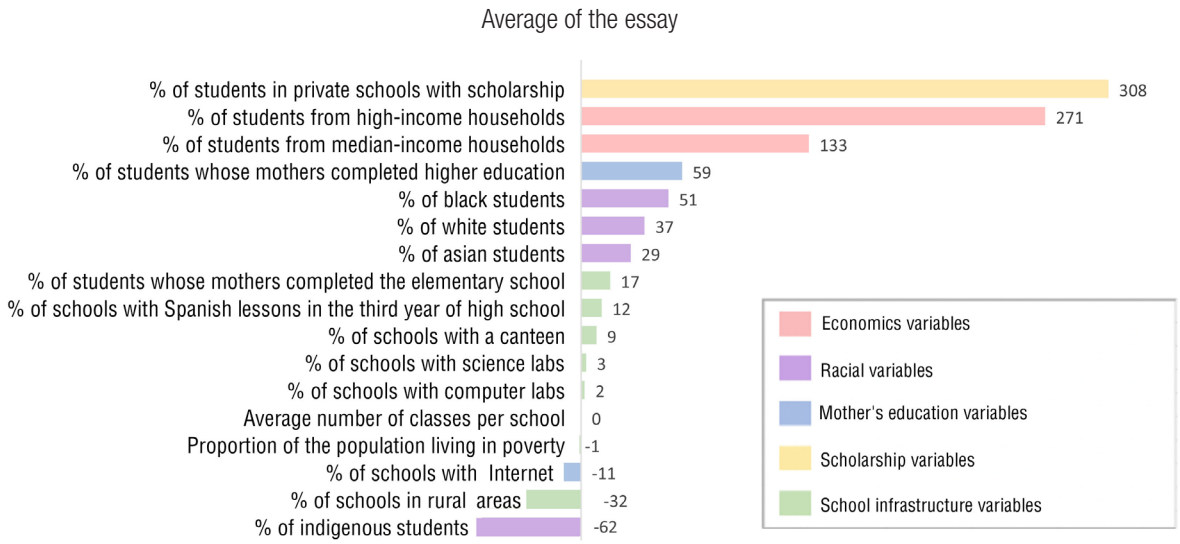
Based on these results, we explored how each independent variable influences the performance and dispersion of grades per municipality. Graph 1 (A to C) shows the importance of each feature for each dependent variable (except for the essay test due to the low R^2).

GRAPH 1 IMPORTANCE OF SELECTED VARIABLES FOR THE AVERAGE AND VARIANCE OF THE GRADE OBTAINED IN THE OBJECTIVE TEST AND THE ESSAY





Continue



Source: Elaborated by the authors.

For the objective test, the "percentage of students in private schools with scholarships" is relevant for the average performance and dispersion of grades in the municipality (Graph 1, A and B). This percentage positively influences student grades while increasing dispersion. Some socioeconomic variables have a high impact on the objective test. In order of relevance, these variables are economic (students' grades and dispersion are higher in municipalities with the highest percentage of high-income families); the students' mother's education (students' grades are higher in municipalities with more educated mothers and mothers working in highly specialized jobs); race (students' grades are higher in municipalities with the highest percentage of white students, and dispersion is higher in municipalities with the highest percentage of black students).

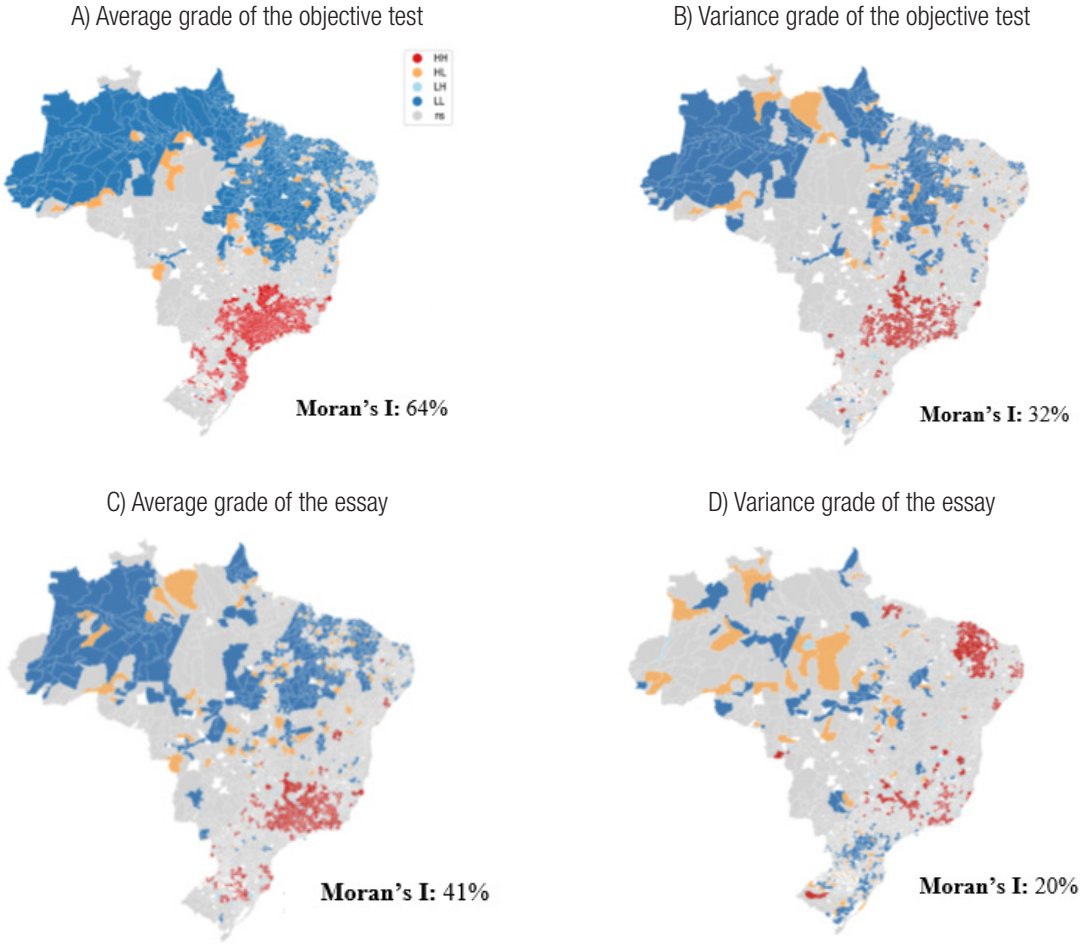
For the Enem essay, the "percentage of students with scholarships in a private school" is relevant for the students in Enem per municipality (Graph 1, C). In addition, some socioeconomic variables impact the essay grade. In order of relevance, they are: economic variables (students' grades are higher in municipalities with a higher percentage of high-income families); race (students' grades are lower in municipalities with the highest percentage of indigenous students); and the students' mother's education (students' grades are higher in municipalities with more educated mothers and mothers working in highly specialized jobs.

Both for the objective test and for the essay, other variables affect the grade, but to a lesser degree. They refer to the school infrastructure, such as the percentage of schools with a canteen or schools with computer labs (the higher the percentage, the better the performance and the higher the dispersion), and the percentage of schools with Spanish lessons in the third year of high school (the higher the percentage, the better performance and the higher the dispersion).

4.3. Neighborhood Analysis and Moran's I

We used the Queen methodology of order 1 to assess the neighboring correlations for the average and variance for the grades of both the objective test and the essay. Figure 4 (A-D) shows the Lisa Maps of the variables and Moran's I for each of them.

FIGURE 4 LISA MAPS OF EACH DEPENDENT VARIABLE



Note: In all maps HH = high-high; HL = high-low; LH = low-high; LL = low-low; NS = not significant
Source: Elaborated by the authors.

The Lisa Maps (Figure 4, A to D) demonstrate the places with a concentration of high grades/variances (HH – in red), low grades/variances (LL – in blue), and transition of grades/variance concentration regions high and low (HL and LH – orange and light blue, respectively). The Lisa Maps evidence a concentration of higher grades in the South and Southeast regions (HH – red) and lower grades in the North and Northeast regions (LL – blue), both for the objective and essay grades. Figure 4 (B and D) shows that the dispersion of student grades for the objective test is greater in the Southeast (HH – red). The dispersion of essay grades is greater in the Southeast and Northeast (HH – red).

Figure 4 shows spatial autocorrelation for the averages and variances high for the average (64% of spatial explanation) and high for the variance (32% of spatial explanation) of the objective test's grade. There is a high spatial autocorrelation for the average of the essay grade (41% of spatial explanation) and moderate for its variance (20% of spatial explanation).

4.4. Spacial Regression vs Linear Regression

The next step was to test the spatial regression models using the STLS method described in section 3.3.1. The same variables selected by the multiple linear regression method were used. Box 4 compares the adjusted R^2 of the multiple linear regression and the adjusted R^2 of the spatial regression.

BOX 4 R^2 OF MULTIPLE LINEAR REGRESSION VS SPATIAL REGRESSION

Model	Multiple linear regression	Spatial regression	Absolute increase of the spatial component	Relative increase of the spatial component
Average grade of the objective test	0.68	0.70	0.02	3.0%
Variance grade of the objective test	0.37	0.39	0.02	5.2%
Average grade of the essay	0.37	0.42	0.05	13.8%
Variance grade of the essay	0.10	0.19	0.09	92.6%

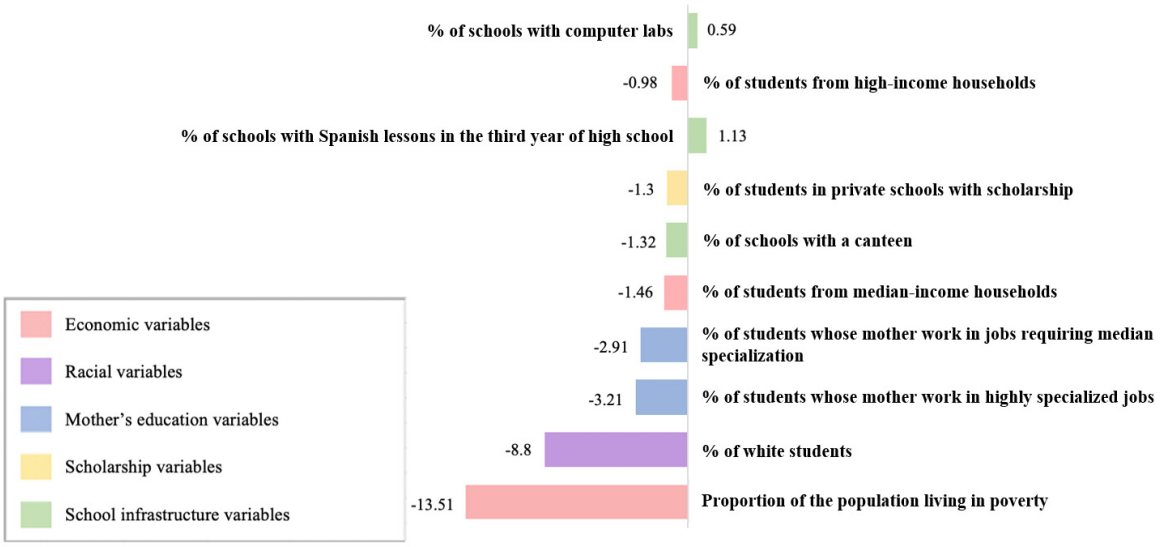
Source: Elaborated by the authors.

The addition of the spatial component has a marginal increase in the correlation for the average and variance of the grade of the objective test. This result indicates that the variables defined in the linear model represent a strong spatial relationship. For the average and variance of the essay grade, the addition of the spatial component has a relatively larger increase (13.8% for the average and 92.6% for the variance). Thus, some regional factors, different from socioeconomic factors, influence the performance and dispersion of grades.

4.5. Local explanation

Finally, the LIME method was used to obtain a more directive understanding of the analysis of the variables that have greater explanatory power per municipality.

GRAPH 2 **IMPACT OF EACH VARIABLE ON STUDENTS' PERFORMANCE IN THE OBJECTIVE TEST OF ENEM IN THE MUNICIPALITY OF ITAPIPOCA (STATE OF CEARÁ), ACCORDING TO THE LIME METHOD**



Source: Elaborated by the authors.

Graph 2 shows the output generated via Lime regarding the average test grade for the city of Itapipoca (state of Ceará), which was 481.9, compared to 507.7 for the average for Brazilian municipalities. As suggested by the exploratory analysis (Section 4.1), the poor results for the variables in the categories economic, race, and mothers' education lead to grades below the national average. In addition, the low results in variables in the categories scholarship and school infrastructure – in the case of the variables percentage of schools with a canteen and schools with Internet – also lead to grades below the national average.

5. DISCUSSION

The results obtained corroborated previous studies on socio-cultural and economic issues as the core of inequalities regarding student performance (Peis, 2011). Thus, the higher the socio-economic status, the higher the student performance in school assessments (Travitski, 2013).

The data obtained reveal that economic factors such as family income and mothers' education (Graph 1, AC) have a strong relationship with student performance, reflecting the dependence between socioeconomic origin and educational outcomes (Bonamino, Alves & Franco, 2010). These aspects are in line with the perspectives of Bourdieu, who argued that some social groups, because of their conditions of existence (capital, habitus, and field), have a history marked by material pressures and temporal urgency, distancing themselves from educational success (1998). Families with more substantial economic and cultural capital tend to invest in their children's school experiences, consequently reproducing the structural order of classes and maintaining power (Lima, 2015; Peis, 2011).

A controversial result is that the percentage of students in private schools with scholarships increases the average grade per municipality. For Soares, Batista, Alves, and Teixeira (2002), social heterogeneity could improve student performance of poorer social strata. Therefore, from this perspective, it would be interesting for the municipality to invest in scholarships. However, it is necessary to carefully reflect on these proposals, as they can be doomed to compensatory programs and, consequently, reinforce the social deficit theory (Ratz, 2019). In addition, this percentage of students with scholarships can intensify inequalities in school performance, reflected in the dispersion of grades indicated by the statistical analyses carried out here.

Another relevant variable for performance in Enem was race. It strongly influences cultural and economic capital, as Brazil is marked by a historical context of racial exclusion, in which pardo and black people have lower incomes than their white counterparts. This explains the superior performance of white students compared to other racial groups (Heringer, 2002). The findings corroborate scientific studies in the area, showing that the educational process often reinforces performance inequality between white and pardo/black students in favor of white students (Basso et al., 2012).

The geographic/spatial variable is another important factor, with students from the South/Southeast regions presenting the best performance. Our data corroborate the results found by Santos (2018), for whom the performance reflects the socioeconomic inequalities of the country's different regions. Thus, Enem becomes a controversial tool since, while increasing competitiveness, it also incorporates the socioeconomic differences of each region, a clear problem in a country with notable regional inequality (Mello, Medeiros, Paiva & Simões, 2014).

This spatial variable is a relevant component when analyzing student essay grades. The results indicate that regional issues, apparently beyond socioeconomic variables, influence essay grades (Box 4). This influence is evident in the low performance of indigenous and rural populations (Graph 1, C). These data may be related to cultural influences specific to different regions, particularly in terms of language.

There is a relational dynamic between linguistic improvement and the sociocultural environment: the use of different functions of language and different levels of speech, the transfer of oral to written language – from the phonological to the orthographic system. Thus, it is necessary to consider that sociocultural factors are closely linked to orality, which greatly influences writing (Batista-Santos & Pereira, 2017). Batista-Santos and Pereira (2017) argue that the linguistic repertoire refers to writing, implying a greater or lesser incidence of inadequacies and marks of orality (i.e., the cultural origin). Consequently, access to linguistic capital is a key determinant of essay grades.

For Bourdieu, writing is the foundation of students' cultural capital. It "allows the accumulation of culture until then preserved in the incorporated state and, correlatively, the primitive accumulation of cultural capital" (Bourdieu, 2009, p. 209, our translation). Therefore, those who are farther away from the hegemonic language have difficulty apprehending school, oral, or written language. Thus, an assessment in which the hegemonic language is the main component tends to intensify inequalities.

We observed that the influence of regionality on essay grades reflects the theory of symbolic school violence developed by Bourdieu (1992). This phenomenon happens because educational assessment privileges the dominant culture – standard language, part of the linguistic capital of socioeconomically privileged regions, such as Brazil's South and Southeast regions.

These findings corroborate assertions from previous research on the notion of school failure based on the context of economic and social inequalities in Brazil (Louzano, 2013). Despite recent progress, poverty and inequality remain one of Brazil's central problems: "Educational opportunities are not equally distributed; school failure affects students from different economic, social, and ethnic-racial groups in different ways" (Louzano, 2013, p. 112, our translation).

In Brazil, it is well-known that socioeconomic, racial, ethnic, and cultural issues have a historical and cumulative origin, and these issues are addressed and overcome through constant social and civil struggles. However, these issues need to be studied deeper in the school environment. One way to break the stigma of school as a reproducer of inequalities is to provide students from underprivileged classes with "powerful knowledge" (Young, 2007), creating didactic opportunities for students to appropriate sophisticated knowledge to live in society. According to Bourdieu (1998), this is the "praxiological" knowledge, which is the source for the subject/social agent's awareness in changing society.

Therefore, education guidelines, policies, and professionals have to overcome the determinist and finalist view of the circumstances lived by the population from different classes. It is crucial to ponder "what are the conditions of social structure and transmission of knowledge in school education that help to overcome or maintain the inequalities related to the access to scientific knowledge by working-class students" (Ratz, 2019, p. 17, our translation).

The findings obtained in this study point out the need to reflect on the possibilities of using Enem as a tool to promote public policies, both for basic and higher education. This means addressing the fact that Brazil is the eighth most unequal country in the world, according to the Gini index (Costa & Gartner, 2017; Indexmundi, 2017).

The inclusive economy, highlighted in Cowell's (2011) studies on inequality measurement, points out the possibility of reducing or eliminating the exclusionary perspective in society caused by factors such as income discrepancies. Therefore, our results emphasize the need for public policies focused on remedial actions (Mello et al., 2014) to democratize population access and permanence in schools. Once the elements necessary to implement such actions are identified, a more specific look at the school environment is required, from the school administration to its educational boards and teachers' pedagogical activities (Louzano, 2019).

Also, the study pointed out other determining factors influencing grades, such as the schools' infrastructure. As shown in the models presented in this research, infrastructure – such as having canteens and laboratories – positively influences educational performance, corroborating studies on this matter (Araújo, 2019). Therefore, when considering that the Brazilian education law (Lei nº 9.394, de 20 de dezembro de 1996) refers to a national assessment exam such as Enem as a source of data that subsidize the improvement of education in the country, the findings on school infrastructure help to reflect on public administration and policies in the field.

This study is relevant to building perspectives for public policies at the municipal level, allowing to map demands that can reduce inequalities and associate them with educational progress. A study conducted in the Brazilian state of Minas Gerais, Caetano, Ávila, and Tavares (2017, p. 913, our translation) did not find an empirical relationship between spending on education and quality, which could indicate poorly managed resources.

> The fact that no empirical relation between the investments in education and the quality of education was found does not dismiss the importance of financing municipal educational services. The resources might be poorly managed due to a possible inefficiency of the public sector. Should that be the case, public management should be more attentive to the management of the available funds and concerned with the return and the efficiency of the services offered.

This study supports this perspective, suggesting the importance of paying attention to the actions for improving educational performance at the municipal level, aiming at greater efficiency in public spending.

6. CONCLUSION

This study sought to identify the variables with the greatest impact on student performance in Enem per municipality, examining the grades obtained in the objective test and the essay. We used data from Enem 2018 (microdata and School Census) referring to municipalities from all over the country, assessing regional differences, and exploring the variables (related to social context and school infrastructure) with the greatest impact on student performance. Statistical models and geospatial analysis were used on the Enem 2018 microdata (Inep, 2018b), School Census 2018 microdata (Inep, 2018a), and IBGE statistics (2018), applying statistical techniques such as thematic maps, Lasso, Moran's I, STLS, and LIME.

The results obtained point to a set of variables with significant explanatory power for the objective test. There is a strong indication that mothers' education, student race, average family income, and the percentage of students in private schools with scholarships are relevant variables for performance and dispersion of grades in the objective test. Although less significantly, the existence of a canteen and science laboratories in schools and Spanish lessons in the third year of high school also positively affected performance and dispersion of the objective test.

Regarding the Enem essay, the variables affecting students' performance are the same observed for the objective test, but to a lesser extent. With the addition of the spatial component, the model's explanation capacity increases, revealing the influence of other regional factors beyond socioeconomic conditions. We notice that the variables related to the macro-socioeconomic context (economic, race, and mother's education) have a greater impact than those related to school infrastructure, both for the objective test and the essay.

The results corroborate studies on the sociology of education by reflecting educational inequality regarding student performance. The findings are connected to Bourdieu's (1998) concepts of cultural capital, as student income depends on their social origin.

Studies that integrate sociology of education with statistical and spatial analysis are crucial to expanding the comprehension of school performance inequalities and allow for a critical and in-depth reflection to promote educational guidelines, policies, and actions that break such symbolic violence. Therefore, we move from reproduction to transformation (Pies, 2011). Thus, this study offers subsidies to reflect on the educational inequalities reproduced by Enem.

This article also proved to be relevant for managerial purposes by understanding the variables with the greatest impact on student's performance at the municipal level to identify the aspects that should be addressed in policies to reduce inequalities and advance educational development.

REFERENCES

Alves, M. T. G., & Soares, J. F. (2007). Efeito-escola e estratificação escolar: o impacto da composição de turmas por nível de habilidade dos alunos. *Educação em Revista*, *45*, 25-59.

Anselin, L. (2013). *Spatial econometrics*: methods and models. Santa Barbara, CA: Springer Science & Business Media.

Araújo, D. L. (2019). *Determinantes do desempenho no Enem dos concluintes do ensino médio no município de Viçosa-MG* (Master Thesis). Universidade Federal de Viçosa, MG.

Arias, D. R. M. (2009). Usos, aplicaciones y problemas de los modelos de valor añadido en educación. *Revista de Educación*, *348*, 217-252.

Basso, A. C. M., Rodrigues, C. F., Brooke, D. A. L., Vignoli, D. A., Candian, J. F., ... Rezende, W. S. (2012). Desigualdade de desempenho e raça: uma análise a partir do Paebes 2009. *Estudos em Avaliação Educacional*, *23*(51), 40-56.

Batista-Santos, D. O., & Pereira, E. B. (2017). A influência da oralidade na escrita: reflexões e desafios no ensino de língua materna. *Desafios – Revista Interdisciplinar da Universidade Federal do Tocantins*, *4*(2), 167-184.

Blini, B. A. Q., & Backes, J. L. (2019). As diferenças culturais no contexto da avaliação em larga escala e de desigualdade social. *Atos de Pesquisa em Educação*, *14*(2), 589-617.

Bonamino, A., Alves, F., Franco, C., & Cazelli, S. (2010). Os efeitos das diferentes formas de capital no desempenho escolar: um estudo à luz de Bourdieu e de Coleman. *Revista Brasileira de Educação*, *15*(45), 487-499.

Bourdieu, P. (1992). *A reprodução*. Rio de Janeiro, RJ: Francisco Alves.

Bourdieu, P. (1998). *Escritos de educação*. Petrópolis, RJ: Vozes.

Bourdieu, P. (2008). *Razões práticas: sobre a teoria da ação*. Campinas, SP: Papirus.

Bourdieu, P. (2009). *O senso prático*. Petrópolis, RJ: Vozes.

Bourdieu, P., Passeron, J. C., & Silva, C. P. G. (1975). *A reprodução: elementos para uma teoria do sistema de ensino*. Petrópolis, RJ: Vozes.

Caetano, C. C. R., Ávila, L. A. C. D., & Tavares, M. (2017). A relação entre as transferências governamentais, a arrecadação tributária própria e o índice de educação dos municípios do estado de Minas Gerais. *Revista de Administração Pública*, *51*(5), 897-916.

Constituição da República Federativa do Brasil de 1988. (1988). Brasília, DF. Retrieved from http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

Costa, G. P. C. L. D., & Gartner, I. R. (2017). O efeito da função orçamentária alocativa na redução da desigualdade de renda no Brasil: uma análise dos gastos em educação e saúde no período de 1995 a 2012. *Revista de Administração Pública*, *51*(2), 264-293.

Cowell, F. A. (2011). *Measuring inequality* (3a ed.). Oxford, UK: Oxford University Press.

Cury, C. R. J. (2013). Sistema Nacional de Educação: uma reflexão provocativa ao debate. *Sistema Nacional de Educação*, *2*, 34.

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, *87*(417), 178-183.

Freire, P. (1994). *Pedagogia do oprimido* (23a ed.). São Paulo, SP: Paz e Terra.

Gremaud, A. P., Felício, F., & Biondi, R. L. (2007). *Indicador de efeito escola: uma metodologia para a identificação dos sucessos escolares a partir dos dados da Prova Brasil*. Brasilia, DF: Inep.

Hartman, J. E. (1999). Accountability, testing, and politics. *Profession*, 125-136. Retrieved from http://www.jstor.org/stable/25595678

Heringer, R. (2002). Desigualdades raciais no Brasil: síntese de indicadores e desafios no campo das políticas públicas. *Cadernos de Saúde pública*, *18*, S57-S65.

Index Mundi. (2017). *Gini Index (World Bank Estimate): country ranking*. Retrieved from www.indexmundi.com/facts/indicators/SI.POV.GINI/rankings

Instituto Brasileiro de Geografia e Estatística. (2018). *Estatísticas*. Retrieved from www.ibge.gov.br/estatisticas/downloads-estatisticas.html

Instituto Brasileiro de Geografia e Estatística. (2019). *Estrutura territorial do território brasileiro*. Retrieved from https://geoftp.ibge.gov.br/organizacao_do_territorio/estrutura_territorial/divisao_territorial/

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2009). *Apresentação do Exame Nacional do Ensino Médio (Enem)*. Retrieved from www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018a). *Microdados do Censo Escolar*. Retrieved from http://inep.gov.br/web/guest/microdados.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018b). *Microdados do Enem*. Retrieved from http://inep.gov.br/web/guest/microdados

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018c). *Pisa 2018 revela baixo desempenho escolar em leitura, matemática e ciências no Brasil*. Retrieved from http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/pisa-2018-revela-baixo-desempenho-escolar-em-leitura-matematica-e-ciencias-no-brasil/21206

Jenks, G. F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, *7*, 186-190.

Kerstenetzky, C. L. (2006). Políticas sociais: focalização ou universalização? *Brazilian Journal of Political Economy*, *26*(4), 564-574.

Lei nº 9.394, de 20 de dezembro de 1996. (1996). Estabelece as diretrizes e bases da educação nacional. Brasília, DF. Retrieved from http://www.planalto.gov.br/ccivil_03/leis/l9394.htm

Libâneo, J. C. (2012). O dualismo perverso da escola pública brasileira: escola do conhecimento para os ricos, escola do acolhimento social para os pobres. *Educação e Pesquisa*, *38*(1), 13-28.

Lima, P. J. (2015). Crítica sociológica do Exame Nacional do Ensino Médio: uma análise bourdiana. *Encontro Nacional de Pesquisa em Educação em Ciências*, *10*, 1-8.

Lima, P. J., Ostermann, F., & Rezende, F. (2013). Análise dos condicionantes sociais do sucesso acadêmico em cursos de graduação em física à luz da sociologia de Bourdieu. *Ensaio – Pesquisa em Educação em Ciências*, *15*(1), 113-129.

Lobo, G. D., Cassuce, F. C. C., & Cirino, J. F. (2017). Avaliação do desempenho escolar dos estudantes da região nordeste que realizaram o Enem: uma análise com modelos hierárquicos. *Revista Espacios*, *38*(5), 12.

Louzano, P. (2013). Fracasso escolar: evolução das oportunidades educacionais de estudantes de diferentes grupos raciais. *Cadernos Cenpec*, *3*(1), 111-133.

Louzano, P., Freitas, P. F., Santos, A. F., Ribeiro, V. M., & Gusmão, J. B. (2019). Implementação de políticas educacionais: elementos para o debate e contribuições para o campo. *Cadernos Cenpec*, *8*(2).

Mello, R. D. N., Medeiros, H. A. V., Paiva, F. S., & Simões, J. L. (2014). O impacto do Enem nas políticas de democratização do acesso ao ensino superior brasileiro. *Comunicações*, *21*(3), 109-123.

Menezes, N. F., & Pazello, E. (2007). Do teachers' wages matter for proficiency? Evidence from a funding reform in Brazil. *Economics of Education Review*, *26*(6), 660-672.

Moran, P. A P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society – Series B (Methodological)*, *10*(2), 243-251.

Muzzeti, L. R. (1999). Consenso ou conflito: contribuições das teorias sociológicas de Émile Durkheim e de Pierre Bourdieu. *Boletim do Departamento de Didática*, *15*, 43-62.

Nogueira, C. M. M., & Nogueira, M. A. (2002). A sociologia da educação de Pierre Bourdieu: limites e contribuições. *Educação & Sociedade*, *23*(78), 15-35.

Perrenoud, P. (2003). Sucesso na escola: só o currículo, nada mais que o currículo! *Cadernos de Pesquisa*, *119*, 9-27.

Pies, N. G. (2011). *Capital cultural e educação em Bourdieu* (Doctoral Dissertation). Universidade de Passo Fundo, Passo Fundo, RS.

Python Software Foundation. (2020). *Python Language Reference: version 3.7.4*. Retrieved from www.python.org

R Development Core Team. (2020). *R: a language and environment for statistical computing*. Vienna,

Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Ratz, S. V. S. (2019). *Influência do nível de exigência conceitual da prática pedagógica de uma professora no desempenho de alunos socialmente diferenciados em uma sequência didática investigativa sobre ecologia de morcegos* (Doctoral Dissertation). Universidade de São Paulo, São Paulo, SP.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22º ACM SIGKDD international conference on knowledge discovery and data mining*, San Francisco California.

Silva, F. G. M. (2017). *Desempenho dos estudantes no Enade: influenciadores ambientais e organizacionais sob a luz da sociologia da educação de Bourdieu* (Master Thesis). Universidade Salvador, Salvador, BA.

Silva, J. F. B. A., Rebouças, S. M. D. P., Abreu, M. C. S. D., & Ribeiro, M. D. C. R. (2018). Construção de um índice de desenvolvimento sustentável e análise espacial das desigualdades nos municípios cearenses. *Revista de Administração Pública*, *52*(1), 149-168.

Soares, J. F. (2004). O efeito da escola no desempenho cognitivo de seus alunos. Reice – *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *2*(2), 83-104.

Soares, J. F., Batista, J. R., Alves, M. T. G., & Teixeira, M. C. A. (2002). *Fatores explicativos do desempenho em língua portuguesa e matemática: a evidência do Saeb-99* (Relatório Técnico). Belo Horizonte, MG: Game/LME/Proav.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society – Series B (Methodological)*, *36*(2), 111-133.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society – Series B (Methodological)*, *58*(1), 267-288.

Torres, H. G., Bichir, R. M., Gomes, S., & Carpim, T. R. P. (2008). Educação na periferia de São Paulo: ou como pensar as desigualdades educacionais? In L. C. Q. Ribeiro, & R. Kaztman (Orgs.), *A Cidade contra a Escola? Segregação urbana e desigualdades educacionais em grandes cidades da América Latina* (pp. 59-90). Rio de Janeiro, RJ: Letra Capital.

Travitzki, R. (2013). *Enem: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar* (Doctoral Dissertation). Universidade de São Paulo, São Paulo, SP.

Young, M. F. D. (2007). *Bringing knowledge back in: from social constructivism to social realism in the sociology of education*. London, UK: Routledge.

Rafael Oliveira Melo

https://orcid.org/0000-0002-4570-1375
Post-graduate in business analytics and big data from Fundação Getulio Vargas (FGV Business Education).
E-mail: rafaelmelo353@gmail.com

Anne Caroline de Freitas

https://orcid.org/0000-0002-0898-074X
Ph.D. student of science teaching at the Institute of Biosciences of University of São Paulo (USP).
E-mail: annefreitas23@gmail.com

Eduardo de Rezende Francisco

https://orcid.org/0000-0001-8895-2089
Ph.D. in business administration from Fundação Getulio Vargas (FGV EAESP); Vice-Coordinator of the Graduate Program in Administration from Fundação Getulio Vargas (FGV EASP).
E-mail: eduardo.francisco@fgv.br

Marcelo Tadeu Motokane

https://orcid.org/0000-0002-8597-6832
Ph.D. in education from the University of São Paulo (USP); Professor of the Department of Biology of the Faculty of Philosophy, Sciences, and Language at the Ribeirão Preto Campus of the University of São Paulo (USP). E-mail: mtmotokane@ffclrp.usp.br