RPPI
Revista Brasileira de Políticas Públicas e Internacionais

Family Capitals and student's performance on ENEM

Os Capitais Familiares e o desempenho dos alunos no ENEM

Leonardo Trajano Dias[1]
Hugo Ormond Vianna Sá Nogueira[2]
Caroline Ponce de Moraes[3]
Rodrigo Tosta Peres[4]

DOI: 10.22478/ufpb.2525-5584.2025v10n2.71309

Received in: 29/08/2024
Approved in: 19/08/2025

Abstract: Educational evaluation has been vastly researched since The Coleman Report was published in 1966. In order to establish a school environment marked by equality and quality, it is necessary for the government to identify the social and economic data that affect students' performance. This paper analyzed the data from the Mathematics section of ENEM (Brazil's High School National Exam) of students from public schools using both multiple linear regression models and multilevel models. This article's goal is to assess the impact of the family capitals (economic, cultural, and social capitals) on students' performance in the aforementioned exam. Our models used individual characteristics such as sex and race. Our results show that the economic capital had the most significant impact when compared to the other two. The Intraclass Correlation Effect was estimated to be between 11% and 12%. Thus, we conclude that the family capitals researched have a significant impact on the students' performance in the Mathematics section of ENEM.

Keywords: inequality; enem; public policies; multilevel models.

[1] Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ).Email: leotrajano97@gmail.com
[2] Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ). Email: hugo.vianog@gmail.com
[3] Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ). Email: caroline.moraes@cefet-rj.br
[4] Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ). Email: rodrigo.peres@cefet-rj.br

Resumo: Avaliação educacional é um tema muito estudado desde o relatório de Coleman, em 1966. Governantes devem ter acesso aos aspectos sociais e econômicos que afetam o desempenho dos alunos para produzir um ambiente com equidade e qualidade. Este trabalho analisou os dados de Matemática do ENEM de alunos de escolas públicas por meio de modelos de regressão múltipla e multiníveis. O objetivo é medir o impacto dos capitais familiares no desempenho. O capital econômico possui maior destaque perante os demais e deve-se levar em consideração variáveis como sexo e raça, e escolares. O efeito escola foi estimado entre 11 e 12%. Pode-se concluir que os capitais familiares impactam o desempenho em Matemática de alunos que se candidatam ao ensino superior no Brasil.

Palavras-chave: desigualdade; enem; políticas públicas; modelos multiníveis.

1. INTRODUCTION

Until half of the 20th century, the idea that providing access to a public and cost-free school would naturally fix social and economic inequities prevailed based on the assumption that, with equal conditions, those who managed to stand out would do so based on their own merit (Nogueira & Nogueira, 2002). However, this optimistic view, in which education would pave the way for a more just and correct society was proven false with the publication of James S. Coleman's report (Coleman, 1966), which showed that variables related to the position occupied by the student and their family in the social hierarchy influenced significatively in their school performance.

As the most interesting result in his report, Coleman stated that socioeconomic variables had more impact on school performance than school variables themselves. Thus, his work proved that students' performance did not only depend on individual capacity, but was also influenced by social, economic, and cultural aspects. Coleman and other authors described the concept of family capitals, which are basically the dimensions in which the student's family socioeconomic situation would influence their school performance. The following family capitals will be used in this work: economic, social, and cultural.

Internationally, since Coleman's report, there have been many studies on the relation between parents' education and their child's school performance, with some of them putting this aspect as the most important to determine school

performance (Björklund & Salvanes, 2011). However, it is highlighted that this trait exerts great influence via indirect channels that are related to the parents' level of education, with these being income (BJörklund & Salvanes, 2011), the number of children (BJörklund & Salvanes, 2011), domestic infrastructure (Curi & Menezes Filho, 2013), and the choice of the best schools (Hoyos, Espino, & García, 2012), besides the fact that parents can transmit to their children visions of reference, values, expectations, and culture (Jacobs & Harvey, 2005).

At the national level, studies have verified a positive relation between income and school performance (Reis & Ramos, 2011). Bonamino et al., (2010) affirm that, even though it is the source of all other forms of capital, economic capital is not enough to explain the relation between school performance and socioeconomic origin by itself, which also must be explained by using other forms of capital. It is worth noting that studies about educational assessment increased only later, mainly because of the low availability of databases with information about the students (Machado et al., 2008).

The lack of data presented in the past is no longer an issue nowadays, in which it is possible and viable to access and manipulate complex data of students, especially those who took part in some sort of large-scale exam, as there is usually a registration form, composed of sections focused on gathering socioeconomic information of the students. One of these exams is the ENEM (National High School Exam), which contains five parts (Nature Sciences and their Technologies, Languages and Codes and their Technologies, Mathematics and their Technologies, Human Sciences and their Technologies, and Essay), whose main objective is to serve as the entrance to higher education through three programs: Sisu, associated with public universities; Prouni, which grants partial and total scholarships, and Fies, which offers students loans.

Based on data from ENEM (INEP, 2022), some quantitative analyses have been made in recent literature. Discussions focused on inequality or in the relation between students' performance and socioeconomic variables show the importance of family capitals in the educational context. Recently, the relation between family income and maternal education in ENEM's performance has

been shown descriptively (Peres, Campos, & Moraes, 2025). The necessity to achieve equity in the Brazilian educational system through public policies outside of the educational scope was highlighted by Travitzki, Ferrão, and Couto (2016). Considering individual variables, the municipal equity in ENEM associated to sex and race has been discussed by Moraes et al. (2022a). Outside of the scope of family capitals, the school impact in ENEM has been calculated by Moraes et al. (2022b) and the differences between high and low achievers also has been studied by Moraes and Peres (2022). A discussion about ENEM not being an adequate exam to measure school quality has been presented by Travitzki (2013).

In this context, this work will make use of multiple regression models and multilevel models to understand how the three family capitals (economic, social, and cultural), besides individual variables of the students (sex and race) and school variables (represented by the socioeconomic school level – Inse), influence on public school's students' performance in Mathematics in ENEM 2019.

2. FAMILY CAPITALS AND PERFORMANCE

Students' performance is affected by variables linked to the school they attend, but also by variables related to their personal lives, especially those related to their socioeconomic context. According to Alves and Soares (2007), in the context of school inequity, Bordieu and Coleman, two pioneers in this type of study, have used the term "capital" as a reference to social and cultural advantages that allow someone to be in a higher socioeconomic level. In this study, three types of capitals will be used: economic capital, social capital, and cultural capital.

The economic capital is represented by income, by the family's assets, and, thus, by their respective purchasing power. Although it has great relevance in the result obtained by a student in school, by itself, it cannot explain it fully (Bonamino et al., 2010). Besides, to Soares and Collares (2006), focusing only on economic capital can reinforce the pessimistic idea that there is no way to improve a student's performance in school unless their family's financial

situation is resolved first. There are at least two other types of capitals that are interesting to study: social and cultural. Social capital is related to social aspects of a student's life, especially those concerning their relationships at home and the exchanges from these relations. Cultural capital involves more elements, usually divided into some types: incorporated, which includes practices and customs that are learned continuously throughout life; objectified, related to the possibility of interpreting and understanding material aspects of culture, such as paintings, sculptures, and books; institutionalized, which involves academic titles (Bourdieu, 1979, p. 3).

It is interesting to think about the contribution of each of these capitals during the years a student stays in school. Economic capital grants access to expensive schools, which are more academically rigid, besides granting access to more complete materials, complementary courses, the possibility of participating in exchange programs and so many other advantages. Social capital, on the other hand, can provide exchanges between the entities of the relationships, exchanges those that can be material or symbolic, reinforcing the feeling of belonging to a group (Bonamino et al., 2010). Lastly, cultural capital can allow the adequate use of formal language, the capacity of interpreting images, besides other benefits that come from having academic titles, which are usually associated with higher income.

The three capitals presented here are related to each other in the sense that all of them derive from the student's family. According to Feijó (2019), parents with a higher degree of education have the capacity of providing a better educational life to their children, through investment of resources based on the idea that a good education can offer future benefits. Economic capital usually has a very strong relation to cultural capital since it is expected that parents with a better education will have a better income, which, in turn, makes them able to provide better socioeconomic conditions for their children, such as access to cultural and educational events. Social capital is a bit more independent than the others, since it is based on the relationships between the family members, which can be strong even if, for example, the economic capital

is low. The relation between cultural capital and social capital is a bit more significative: with no strong social relationships, cultural capital cannot exert the influence it should exert, even if the parents have a high educational degree (Coleman, 1988, p. 110; Bonamino et al., 2010, p. 491).

It becomes knowable that each capital has its individual contribution in a student's school performance. However, it is important to highlight that having a combination of two or even three of these capitals usually results in an even more impactful result in school performance. For example, in the work of Jacob and Harvey's (2005), parents with high cultural and economic capitals were more worried with their children's education: they were able to talk about their children's results in school; on the other hand, parents whose children were in schools considered a bit worse knew little or nothing about their children's results and claimed that it was not something important to know.

In a more empirical and quantitative analysis, social capital can be represented by the size of the student's family. Feijó (2019) affirms that there is a trade-off between the student's family size and the student's years of study. Having more children or simply having a bigger family demands more usage of resources in other areas, which explains the trade-off. Besides, an environment with many people living together, especially in the hypothesis of having low economic capital, can generate more issues for students trying to concentrate on their studies.

An aspect that must be highlighted is the difficulty in measuring social and cultural capitals, especially considering that, in a quantitative study such as this, variables associated with them will be added to models to predict a student's result. In this study, economic capital will be represented by monthly income range; social capital will be represented by the number of people residing with the student; cultural capital will be represented by the mother's degree of education. All this data was obtained in the socioeconomic form that students who took ENEM must fill annually.

3. METHODOLOGY

The present study made use of two sources of data obtained through the website of Inep (National Institute of Educational Studies and Research Anísio Teixeira):

The first is composed of the microdata of ENEM 2019, which provided students' grades and registration information, obtained through the form that the students must answer when they register for the exam. This information is essential to this study.

The second one refers to the Inse (socioeconomic school level) of 2019, which provided information concerning schools, including the socioeconomic level of each one of them.

For the treatment of the data and creation of the database for the treated data, both used in the descriptive analysis and inferences, the software *Microsoft SQL Server Management Studio* (SSMS) (Microsoft, 2018) was used.

It is important to highlight that this study is essentially transversal, since it does not assess students' progress over time.

For the data treatment, firstly both sources were extracted to SSMS, resulting in two tables with raw data, with one containing ENEM 2019 microdata and the other containing data from Inse 2019, which is the indicator of a school's socioeconomic level.

Afterwards, a preliminary treatment was applied to the data, for typing adjustment, character adjustment, and removal of possible unnecessary spaces in records. Also, columns that would not be used in this study were removed. Thus, a second group of tables was created, one with treated data of ENEM 2019 and the other with treated data of Inse 2019. They contained the following information:

- ENEM 2019: Registration number (which was used later to count the number of registered students), student's city, student's federative unit, student's sex, student's race/color code, school code, mother's education code, father's education code, number of people residing with the student, monthly income range code, and internet access code.

- Inse 2019: School code, city of the school code, federative unit of the school, school area code, administrative dependency of the school code, school location code, absolute value of Inse, and Inse classification.

It is worth noting that, for the descriptive analysis, Inse classification was used, which is a categorization based on the absolute value of Inse (utilized in the statistics models), as shown in Table 01.

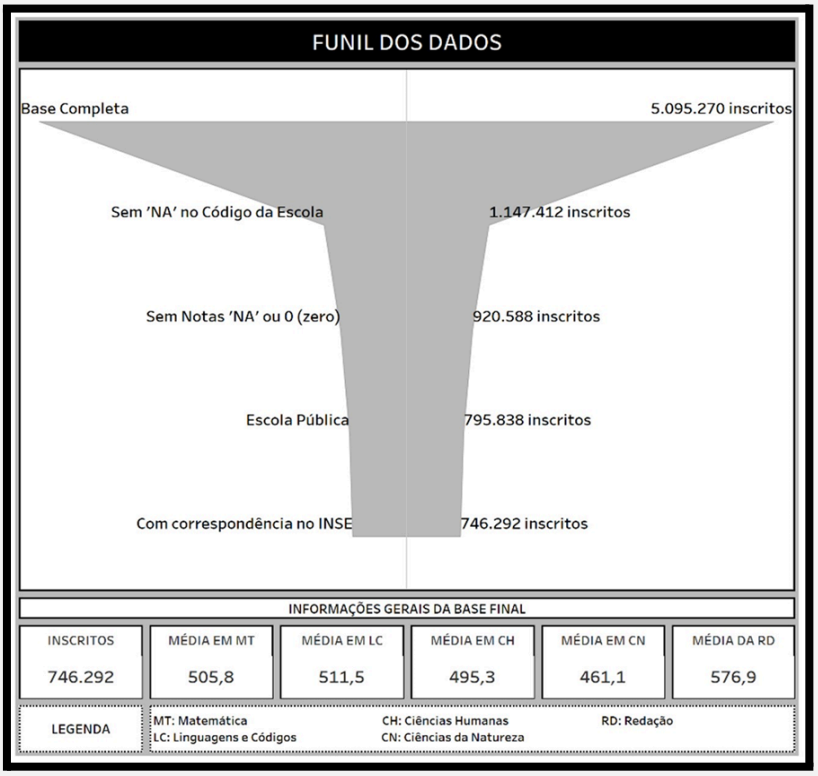
Table 01: Inse Classification

Inse Classification	Absolute value of Inse
I	Up to 3,00
II	3,00 to 4,00
III	4,00 to 4,50
IV	4,50 to 5,00
V	5,00 to 5,50
VI	5,50 to 6,00
VII	6,00 to 7,00
VIII	7,00 or more

In a third moment, to obtain the final tables used to elaborate the descriptive analysis, a filter was applied to the data and the treatment of fields that were initially coded were converted into descriptive fields, to make future analysis easier to make. The only field that did not undergo this process was the field 'school code' because it was later used to unite the tables.

For the ENEM 2019 data, four filters were applied, as shown, together with a summary of the final database in Figure 01, which contains a funnel that details all filters applied in this process.

Figure 01: Data funnel



The initial base contained a total of 5,095,270 registered students. After removing all registered students with no school data, which is vital information for this work, the new base contained 1,147,412 registered students. From this new base, it was decided to remove students with null or zero grades, which represents absent students or students that were eliminated, further reducing the database to 920,588. In a third filtering, students that were not from public schools were also removed, which reduced the base to 795,838 students. Lastly, students from schools that did not have an Inse 2019 value were also removed, which resulted in the final database composed of 746,292 students. In this final base, students had the highest average in Language and Codes and lowest average in Nature Sciences.

For the Inse 2019 database, the only filtering done involved keeping only information of schools that had correspondence in the ENEM database or, in other words, only records of schools that had at least one student among the 746,292 students of the final base. Due to this, from an initial total value of 68,868 schools, the number was reduced to 19,613.

With the two final tables (ENEM 2019 and Inse 2019), the next step consisted of exporting a CSV file so that later a consolidated source could be created, containing both tables on the Tableau software (Tableau, 2020). This source contained information about the 746,292 registered students and the Inse information of their respective schools. To do so, the JOIN function from Tableau was used (Tableau, 2020).

Some decisions related to the database were made to utilize the data in the statistics models. Therefore, another treatment was applied, this time with the aid of Microsoft SQL Management Server (Microsoft, 2018). The variables that were related to the first level, the students' level, were divided into two groups:

- Individual variables, composed of the student's sex and race;
- Family capitals: social, economic, and cultural, represented, respectively, by the number of people residing with the student, by the family's monthly income, and by the student's mother's education.

As the next paragraphs will describe, the main procedure done was the creation of binary variables from the categorical variables so that these could be added to the models built in *RStudio* (Rstudio, 2021) (the dummy variables), which only accept numerical variables.

The categorical nominal variable "sex" was transformed into a binary variable, being "1" the value attributed if the student was a boy and "0" if the student was a girl so that the impact of being a boy in the result could be measured. This choice was made because men have historically achieved higher grades in Mathematics than women, as analyzed by Moraes (2021).

The categorical nominal variable "race" was transformed into a binary variable, being "1" the value attributed if the student had self-reported as white or yellow, and "0" in any other case so that the impact of being white or yellow in the result could be measured. This choice was made because these races, white and yellow, have historically achieved higher grades in Mathematics than the other races, as it can be seen stated in the works of Moraes (2021) and Travitzki, Ferrão and Couto (2016).

The categorical ordinal variable "income" was initially transformed into a continuous variable, by using the midpoint of each category. Afterwards, a binary variable was created, which received the value of "1" if the student's family monthly income was higher or equal to R$ 3493,005 and "0" if it was not. This choice was because the quantile 90 is represented by that specific value of monthly income, which includes 111,363 out of the 746,292 entries (approximately 15% of students due to draws in the values because of the use of the midpoint). This was done so that the impact of having at least that family monthly income could be measured in the student's result.

The categorical ordinal variable "mother's education" was transformed into a binary variable, with the value "1" being attributed if the mother at least had completed high school, and "0" if that was not the case, in such a way that it would be possible to measure the impact of the student's mother education in their result. This was done because previous studies have concluded that this is a factor that exerts impact on children's grades (Moraes, 2021).

The discreet quantitative variable "number of people that reside with the student" was transformed into a binary variable, with the value "1" being attributed if that number was less or equal to 4, and "0" if it was not. This was made because this variable's distribution showed that the option with the highest frequency was "four people", which registered 268,009 responses (approximately 35.91% of total responses). By turning this variable into a binary variable, its new distribution of frequencies was the following: 509,295 students (68.24%) with "maximum of 4 people" and 236,997 students (31.76%) with "5 or more people". With this procedure, it will be possible to measure the impact of living with a number less than or equal to the original mean four of the variable in a student's result. This is important because it is believed that the less people that live with a student, the better conditions this student will have to study, which would lead to higher grades.

For the multilevel models, only one variable was used for the level associated to schools, which was the "absolute value of Inse" as it varies according to the school the student attends.

Firstly, the idea was to use models with the exclusive data of each variable. However, after studying Feijo's (2019) work, the authors were inspired to make use of similar models to those used by her. These models make use of binary variables in such a way that it becomes possible to measure the impact of having only the first capital, only the second capital, and of having both.

The models, represented by the capital letter M, were divided into three types: 1) M1 up to M3: multiple linear regressions with interactions between the family capitals, two at a time; 2) M4 up to M6: multiple linear regressions with the family capitals clustered two at a time with the addition of students' individual variables (sex and race); 3) M7 up to M9: multilevel models with two levels (student and school), controlled by the absolute value of Inse.

Models M1 to M3 will be the first to be described. M1 includes mother's education and number of people that reside with the student. M2 includes mother's education and family's monthly income. M3 includes number of people that reside with the student and family's monthly income. The three models are represented, respectively, by the equations 1, 2, and 3 below:

$$Y_i = \beta_0 + \beta_1\left(1 - Motheredu_i\right)People_i + \beta_2\left(1 - People_i\right)Motheredu_i + \beta_3\left(People_i\right)\left(Motheredu_i\right) + \varepsilon_i \quad \textbf{(1)}$$

$$Y_i = \beta_0 + \beta_1\left(1 - Motheredu_i\right)Income_i + \beta_2\left(1 - Income_i\right)Esc_i + \beta_3\left(Qtde_i\right)\left(Esc_i\right) + \varepsilon_i \quad \textbf{(2)}$$

$$Y_i = \beta_0 + \beta_1\left(1 - People_i\right)Income_i + \beta_2\left(1 - Income_i\right)People_i + \beta_3\left(Income_i\right)\left(People_i\right) + \varepsilon_i \quad \textbf{(3)}$$

Where Y_i éis the grade achieved by the student i in Mathematics in ENEM 2019, β_0 is the intercept of each regression model, $Motheredu_i$ is the binary variable that assumes the value of "1" if the student i's mother education is at least completed high school and "0" if it is not, $People_i$ is the binary variable that assumes the value of "1" if the student resides with up to 3 people and "0" if

they do not, $Income_i$ is the binary variable that assumed "1" if the student's i family monthly income is equal to or higher than R$ 3.493,005 and "0" otherwise and ε_i represents the associated error of each model. The remaining β's are the regression coefficients associated to each part of the regression, which will determine how important and related to Y_i that part is.

There are three interactions between the family capitals in each one of the equations above. In each model, there are three conditions being tested: 1) if the student only has the first family capital; 2) if the student only has the second family capital; 3) if the student has both. Using these models that contain interactions makes it possible to measure the individual effect of each variable of interest and the effect of the combinations of two and even three family capitals.

In M1, the term "$\beta_1\left(1 - Motheredu_i\right)Income_i$" measures the impact in a student's grade whose family's monthly income is above the established cutoff point and whose mother's education is below complete high school. On the other hand, the term "$\beta_2\left(1 - Income_i\right)Motheredu_i$" measures the impact in a student's grade whose family's monthly income is below the established cutoff point and whose mother's education is complete high school or more. Lastly, the term "$\beta_3\left(Income_i\right)\left(Motheredu_i\right)$" measures the impact in a student's grade whose family's monthly income is above the established cutoff point and whose mother's education is complete high school or more.

In M2, the term "$\beta_1\left(1 - Motheredu_i\right)People_i$" measures the impact in the grade of a student that lives with a number below the cutoff point and whose mother's education is below complete high school. On the other hand, the term "$\beta_2\left(1 - People_i\right)Motheredu_i$" measures the impact in a student's grade whose mother's education is complete high school or more and that resides with a number of people above the cutoff point. Lastly, the term "$\beta_3\left(People_i\right)\left(Motheredu_i\right)$" measures the impact in a student's grade whose

mother's education is complete high school or more and that resides with a number of people equal to or above the cutoff point.

In M3, the term "$\beta_1\left(1 - People_i\right)Income_i$" measures the impact in a student's grade whose family's Monthly income is above the cutoff point and whose residence has a number of people above the cutoff point. On the other hand, the term "$\beta_2\left(1 - Income_i\right)People_i$" measures the impact in a student's grade that resides with a number of people equal to or below the cutoff point and whose family's monthly income is below the cutoff point. Lastly, the term "$\beta_3\left(Income_i\right)\left(People_i\right)$" measures the impact in a student's grade whose family's monthly income is above the cutoff point and that resides with a number of people below ore qual to the cutoff point.

The second group of models includes M4, M5, and M6, which are like M1, M2, and M3. The main difference is the addition of students' individual variables sex and race, represented, respectively, by Sex_i e $Race_i$ in the models. Therefore, this second group of models has more explanatory variables than the first.

The order of the two-by-two combinations of the family capitals was kept the same as the one used in the first group of models to allow the comparison in the Results section easier to make and understand. Therefore, M4 includes mother's education and the number of people that reside in the same home; M5 includes mother's education and family's monthly income; and M6 includes the number of people that reside in the same home and family's monthly income. These three models are represented, respectively, by equations 4, 5, and 6:

$$Y_i = \beta_0 + \beta_1\left(1 - Motheredu_i\right)People_i + \beta_2\left(1 - People_i\right)Motheredu_i + \beta_3\left(People_i\right)\left(Motheredu_i\right) + \beta_4\left(Sex_i\right) + \beta_5\left(Race_i\right) + \varepsilon_i \quad \textbf{(4)}$$

$$Y_i = \beta_0 + \beta_1\left(1 - Motheredy_i\right)Income_i + \beta_2\left(1 - Income_i\right)Motheredu_i + \beta_3\left(Income_i\right)\left(Motheredu_i\right) + \beta_4\left(Sex_i\right) + \beta_5\left(Race_i\right) + \varepsilon_i \quad \textbf{(5)}$$

$$Y_i = \beta_0 + \beta_1\left(1 - People_i\right)Income_i + \beta_2\left(1 - Income_i\right)People_i + \beta_3\left(Income_i\right)\left(People_i\right) + \beta_4\left(Sex_i\right) + \beta_5\left(Race_i\right) + \varepsilon_i \quad \textbf{(6)}$$

Where Y_i is the grade achieved by the student *i* in Mathematics in ENEM 2019, β_0 is the intercept of each model of regression, and ε_i represents the error associated to each model.

Similarly to what was discussed for the models of the first group, once again there are two-on-two interactions between the family capitals in each equation. However, in this second group of models, it is possible to quantify the impact of being a boy or not and of a student self-declaring themselves as white or yellow in their grade in Mathematics.

The third and last group of models developed in this study includes multilevel models. They are highly recommended when it is not possible to guarantee total independence of the data, such as the case of students that go to the same school and tend to have similar socioeconomic conditions. Thus, models M7, M8, and M9 have two levels, with these being: level I) Students; level II) Schools. The choice of these levels is like other studies in literature, such as Moraes (2021), who used the levels student, school, and city to measure the impact of the school effect in ENEM 2017. To characterize students, there are individual variables and variables related to the family capitals. To characterize the school, there is Inse. The clustering of the students was made using the school code.

Once again, the order of the two-by-two combinations was respected so that the comparison between the regression coefficients in the results sessions can be made more evident. Therefore, M7 includes mother's education and the number of people that reside in the same home; M8 includes mother's education and family's monthly income; and M9 includes the number of people that reside in the same home and family's monthly income. The three models are represented, respectively, by equations 7, 8, and 9:

$$Y_{ij} = \beta_{0j} + \beta_1\left(1 - Motheredu_{ij}\right)People_{ij} + \beta_2\left(1 - People_{ij}\right)Motheredu_{ij} + \beta_3\left(People_{ij}\right)\left(Motheredu_{ij}\right) + \beta_4\left(Sex_{ij}\right) + \beta_5\left(Race_{ij}\right) + \varepsilon_{ij} \quad \textbf{(7)}$$

$$Y_{ij} = \beta_{0j} + \beta_1\left(1 - Motheredu_{ij}\right)Income_{ij} + \beta_2\left(1 - Income_{ij}\right)Motheredu_{ij} + \beta_3\left(Income_{ij}\right)\left(Motheredu_{ij}\right) + \beta_4\left(Sex_{ij}\right) + \beta_5\left(Race_{ij}\right) + \varepsilon_{ij} \quad \textbf{(8)}$$

$$Y_{ij} = \beta_{0j} + \beta_1\left(1 - People_{ij}\right)Income_{ij} + \beta_2\left(1 - Income_{ij}\right)People_{ij} + \beta_3\left(Income_{ij}\right)\left(People_{ij}\right) + \beta_4\left(Sex_{ij}\right) + \beta_5\left(Race_{ij}\right) + \varepsilon_{ij} \quad \textbf{(9)}$$

Where:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\left(Inse_j\right) + u_{0j} \quad \textbf{(10)}$$

Where Y_{ij} is the grade achieved by the student i that goes to the school j in Mathematics in ENEM 2019, β_{0j} measures the impact of the school in the student's grade, ε_{ij} is the error associated to each model at the student level, and u_{0j} is the error associated to the model at the school level. It is worth noting that students that study in schools with a high value of β_{0j} have advantages when compared to students that study in schools with a low value of β_{0j} regardless of their individual socioeconomic conditions.

In this group of models, it is important to see that the absolute value of Inse plays a major role, since it represents the only explanatory variable at the school level. The coefficient associated with it represents how impactful the socioeconomic level of the students is in their grades.

Due to the presence of a second level, the school, in these multilevel models, it becomes possible to measure the impact that the school has in the performance of students that attend public schools in the Mathematics section of ENEM 2019. This impact is measured through the intraclass correlation coefficient (ICC), which can be calculated through the following formula, according to Hox, Moerbeek and Schoot (2010):

$$\rho = \frac{\sigma^2_{\mu 0}}{\sigma^2_{\mu 0} + \sigma^2_{\varepsilon}} \quad \textbf{(11)}$$

Where represents the residual variance at the school level, related to u_{0j} and σ^2_{ε} represents the residual variance at the student level, related to ε_{ij}. Therefore, the result of ICC, calculated through equation (11) shows how much variability between schools can explain the students' result in the Mathematics section of ENEM 2019. Therefore, the higher the ICC, the more the students' result is related to the school they attended during high school.

A p-value at a significance level of 1% was used to confirm the significance of each of the variables used in the models described in this section. In all the models the condition was confirmed ($p < 0,01$), which reassures that the choice of the variables was accurate.

4. RESULTS

Tables 02, 03, and 04 were created to allow comparison between students that had each of the family capitals, combined two-on-twos. It is observed that the averages of the grade obtained in Mathematics in ENEM 2019 are different between these groups. With the aim of confirming this fact, two statistical tests were performed: test-t for two samples and non-parametric test of Mann-Whitney. Test-t compares if two means of independent groups are different and Mann-Whitney's test compares if the medians of two independent groups are different. Both provided a p-value below 0,01, which indicates that the hypothesis that the averages are the same should be rejected – in other words, they have distinct values. This result stimulates the creation of models using interactions between variables that represent the three family capitals.

Table 02: Average grade (Family's Monthly income and Mother's education)

	Income below quantile 900	**Income above quantile 90**
Mother with H. S.	485,0	535,7
Mother without H. S.	509,7	569,9

Table 03: Average grade (Mother's education and People that live in the same home)

	Up to 4 people	**5 or more people**
Mother with H. S.	492,7	481,1
Mother without H. S.	527,1	514,3

Table 04: Average grade (Family Monthly income and People that live in the same home)

	Up to 4 people	**5 or more people**
Income below quantile 90	500,5	486,6
Income above quantile 90	566,0	551,4

Table 05 presents the results of the coefficients of the two first groups of models (from M1 to M6, as described previously). Note that the family capitals were identified as EC for economic capital, SC for social capital, and CC for cultural capital. Models M1 to M3 were estimated without considering the individual variables race and sex. On the other hand, in models M4 to M6, these variables were added. Mother's education is represented by "Esc." and number of people that reside in the same home is represented by "Qtde". All variables in the models were considered significative ($p < 0,01$).

Table 05: M1 a M3 (family capitals' influence without considering individual variables). M4 a M5 (family capitals' influence considering individual variables)

	Without individual variables			**Without individual variables**		
Model	M1	M2	M3	M4	M5	M6
Intercept	481,08	484,97	486,62	461,32	466,06	467,98
Race	-	-	-	27,33	22,32	22,85
Sex	-	-	-	34,64	32,75	33,32
Motheredu (CC)	33,22	24,71	-	28,80	21,96	-
People (SC)	11,57	-	13,89	8,38	-	11,10
Income (EC)	-	50,72	64,73	-	41,87	55,98
Income*Motheredu.	-	84,93	-	-	75,14	-
Income*People	-	-	79,42	-	-	68,34
Motheredu*People	45,99	-	-	38,84	-	-

Based on these results, it is possible to notice that the economic capital (Income) has the highest coefficients, followed by cultural capital (Motheredu), and, lastly, as the least influent, capital social capital (People). As for the combinations, income with mother's education has the highest coefficient, followed by income and number of people that reside in the same home and lastly mother's education and number of people that reside in the same home.

This pattern was observed in every model analyzed, with only small variations in terms of magnitude of the value of the coefficients. When individual variables were added, the magnitude of individual family capitals was reduced.

Economic capital's protagonism is such that it exceeds even the impact of individual variables sex and race, which were bigger than the impacts of social and cultural capitals in models M4, M5, and M6.

To interpret the coefficients presented in Table 5, consider the characterization of M1. Based on the estimated coefficients, it can be affirmed that a student whose mother has at least completed high school achieves, on average, 33.22 more points than a student whose mother has not completed high school. On the other hand, a student that lives with 4 or less people in their home achieves, on average, 11.57 more points than a student that lives with

more than 4 people. Lastly, a student whose mother has at least completed high school and that lives with 4 or less people in their home achieves, on average, 45.99 more points than a student whose mother hasn't completed high school and that lives with more than 4 people.

Similarly, if we analyze a model from the second group, M5, per example, it can be affirmed that a white or yellow student achieves, on average, 22.32 more points than a student of another race. A student that is a boy achieves, on average, 32.75 more points than a girl. A student whose mother has completed high school achieves, on average, 21.96 more points than a student whose mother has not completed high school. A student whose family's monthly income is higher or equal to the cutoff point achieves, on average, 41.87 more points than a student whose family's monthly income is below the cutoff point. Lastly, a student whose family's Monthly income is higher or equal to the cutoff point and whose mother has at least completed high school achieves, on average, 75.14 more points than a student whose mother has not completed high school and whose family's monthly income is below the cutoff point.

Table 06 presents the results of the multilevel models with the control of individual variables race and sex and the socioeconomic level of the schools. All variables were considered significative ($p < 0,01$).

It is possible to observe that, once again, the pattern of the magnitude of each coefficient was the same. However, it is also worth noting that, even though the variable sex suffered a decrease, it was a very small decrease related to the other variables. The socioeconomic level of the school in a student's grade is also very impactful.

With the results of the multilevel models, it was possible, through the residual variance of each level, to calculate the ICC, which is equivalent to the school effect in each model, as shown in Table 07.

Table 06: M7 a M9 (Influence of the family capitals with the addition of individual variables and control with the socioeconomic school level)

	Multilevels		
Model	M7	M8	M9
Intercept	299,00	316,69	313,92
Race	11,99	10,97	11,44
Sex	32,88	31,95	32,33
Motheredu (CC)	16,67	13,50	-
People (SC)	2,32	-	3,95
Income (EC)	-	21,32	28,82
Income*Motheredu	-	40,83	-
Income*People	-	-	32,79
Motheredu*People	18,95	-	-
Inse	35,64	32,03	33,18

Table 07: Quantification of the school effect for models M7 to

	Multilevels		
Model	M7	M8	M9
Residual Var - School Escola	836,6	790,8	806,0
Residual Var - Student	6386,6	6324,8	6364,0
ICC (school effect)	11,58%	11,11%	1,24%

Therefore, it was verified that the contribution of the variable that represents the school effect (Inse) in a student's grade ranged from 11 to 12% for the 3 models developed in this study. This means that from 11 to 12% of public-school students' performance in the Mathematics section of ENEM 2019 that did not obtain zero in any of the sections can be explained by the school that they go to. Although apparently low, these percentages agree with other studies from literature that used the school socioeconomic level as a control variable, such as Moraes (2021), who estimated ICC as 9% for students that attended public schools, although the author used some more variables in her study.

5. DISCUSSION

The present work verified a strong direct influence of the family capitals in the students' results in ENEM. It is worth nothing that, according to Feijo (2019), parents' education influences their children through four indirect channels, which are financial resources, family size, house infrastructure, and school quality. Thus, the direct influence of parents' education is related to the impact on their children's school performance through the transmission of reference vision, values, expectations, culture (Björklund & Salvanes, 2011), parental involvement (Castro et al., 2015), genetics (Plug, 2004), among others.

It was observed that the economic capital, represented by family's monthly income, had the highest influence, followed by cultural capital, represented by the student's mother education and, lastly, with the least influence, social capital, represented by the number of people that reside with the student.

This pattern was observed in all models, even though as individual variables were added (in models M4 to M6) and multilevel models were introduced, the magnitude of the coefficients, or in other words, the impact of the family capitals, decreased.

The protagonism of the economic capital was not unexpected just as it was not something unknow in literature. It is plausible to expect that families with higher income can provide excellent studying conditions, besides adequate daily care (Bonamino et al., 2010). However, economic capital cannot be the only aspect that influences a student's performance. In their work, Bonamino et al. (2010), affirm that, even though it is the source of all other forms of capitals, economic capital has limitations in explaining the relation between school performance and socioeconomic origin. This led sociologists such as Bourdieu and Coleman to consider other forms of capital that interact and contribute directly with the economic capital to explain that relation.

The relation between the forms of capital and the idea that a better result is expected when a combination of two or more family capitals is present was also not unknown in literature. Feijó, (2019) highlights that there are other

factors correlated to the parents' education and that favor student's performance. As examples, the relationship between parents' education with family's monthly income (Björklund & Salvanes, 2011), with the smaller number of children (BJörklund & Salvanes, 2011), and with the choice of better schools for the children (Hoyos, Espino, & García, 2012).

Alves (2010), for example, highlights that the choice of a school by the parents is extremely influenced by the volume and structure of economic, cultural and social capitals, where inequality, both in the form and also in the possession of such capitals, affects directly in the criteria of the choice of a school, posing limits, especially financially, to that choice. It is worth mentioning that the study of Alves (2010) was made considering both public and private schools, while this work focused only on public schools.

Aside from results that were already known by literature, this work used data of a large-scale exam, ENEM, differently from many other national works that use SAEB (Sistema de Avaliação da Educação Básica) as their main source.

Besides, this study measured the influence of each variable in the grade achieved on ENEM, proving the importance of some variables in comparison to others in the student's final grade.

Income and mother's education, representing, respectively, economic and cultural capitals, had the highest impact, followed by income and the number of people that reside with the student, representing, respectively, economic and social capitals, and lastly mother's education and the number of people that reside in the same home. The pattern of decrease in the magnitude of these types of variables in the models developed were also shown.

In M7, M8, and M9, models with the introduction of the school level, by using the residual variance of the student and school levels, it was possible to calculate the ICC which, for this study, is the school effect. Moraes (2021) affirms that the school effect and the family effect (complimentary to the school effect) are defined, respectively, by the contributions of the school and the family in the student's grade. Calculating the school effect has the objective of

characterizing the contribution of the school to a student's performance through the residual variances in each level.

For models M7 to M9, the school effect was estimated to be between 11 and 12%, which means that the family effect was between 88 and 89%. This result shows that a student's performance can be explained by a diversity of factors, many of which are outside the school context.

Considering that this study was elaborated during the pandemic of covid-19 in Brazil and in the world, but made use of data from ENEM 2019, an event that happened before the pandemic, the authors believe that the impacts observed here can become even greater. In a work that discussed family variables, Moraes et al. (2021) highlight the restrictions and their impacts as a result from the pandemic, being one of them the closing of many schools, which directly affected students' performance, especially to the poor students, increasing the pre-existing social abyss and once again highlighting the importance of social and economic factors.

6. CONCLUSION

This study consisted of a transversal study in the subject of educational evaluation, focusing on measuring the impact of family capitals in public school students' grade in the Mathematics section of ENEM 2019. As expected, economic capital had the highest impact, being followed by cultural capital and social capital.

Besides the descriptive analyses made to better understand the database, multilevel models were used so that it became possible to measure the contribution of the school in the students' performance. ICC, which measures the school effect, oscillated between 11 and 12% in the three developed models, showing that, in fact, the school a student goes influences their results. However, the remaining 88 to 89% come from factors that are external to the school context, which highlights that socioeconomic inequity continues to be a strong obstacle in one of education's main functions: that of being a source of transformation and social mobility.

This work contributes scientifically to the education evaluation literature by making use data from the last edition of ENEM before the start of the pandemic and by focusing on students from public schools. As suggestions for future studies, students from private schools could be included in the database in a way that a comparison between public and private schools could be made. Besides, it would be interesting to make use of more variables to represent the family capitals, especially the social capital (for instance: if there is a reserved space at home for studying) and cultural capital (for instance: father's education). Lastly, making use of random regression coefficients and not just a random intercept could shed more light on the impact of the family capitals controlled by the socioeconomic level.

REFERENCES

Alves, F. (2010). Escolhas familiares, estratificação educacional e desempenho escolar: quais as relações? *Dados - Revista de Ciências Sociais*, 53(2), 447-468.

Alves, M. T. G., Soares, J. F. (2007). As pesquisas sobre o efeito das escolas: contribuições metodológicas para a sociologia da educação. *Sociedade e Estado (UnB. Impresso)*, 22, 435-473.

Björklund, A. & Salvanes, K. (2011). Education and family background: Mechanisms and policies. In Hanusek, E. A., Machin, S., & Woessman, L (Orgs.). *Handbook of the Economics of Education*. (pp. 201-247). San Diego: North Holland/Elsevier.

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. *Rev. Bras. Educ. [online]*, 15(45), 487-499.

Bourdieu, P. (1979). Les trois états du capital culturel. *Actes de la Recherche en Sciences Sociales*, 30, 3-6

Castro, M., Expósito-Casas, E., López-Martín, E., Lizasoain, L., Navarro-Asencio, E., & Gaviria, J. L. (2015). Parental involvement on student academic achievement: A meta-analysis. *Educational research review*, 14, 33-46.

Coleman, J. S. (1966). *Equality of educational opportunity.* Washington, DC: US Government Printing Office.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, 95-120.

Curi, A. Z., & Menezes Filho, N. A. (2013). Mensalidade Escolar, Background Familiar e os Resultados do Exame Nacional do Ensino Médio (ENEM). *Pesquisa e Planejamento Econômico* (Rio de Janeiro), 43, 223-254.

Feijó, J. R. (2019). *Ensaios sobre o desempenho dos estudantes no ENEM 2017*. (Tese de Doutorado). Universidade Federal do Ceará, Faculdade de Economia, Fortaleza, Brasil.

Hox, J. J., Moerbeek, M, & Schoot, R. V. de. (2010). *Multilevel Analysis: Techniques and Applications.* (2.ed.). New York: Routledge.

Hoyos, R. E., Espino, J. M., & García, V. (2012) Determinantes del logro escolar en México. Primeros resultados utilizando la prueba ENLACE media superior. *El trimestre económico*, 79(316), 783-811.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *ENEM - Exame Nacional do Ensino Médio, 2022*. Disponível em: https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem.

Jacobs, N., & Harvey, D. (2005). Do parents make a difference to children's academic achievement? Differences between parents of higher and lower achieving students. *Educational Studies*, 31, 431-448. doi:10.1080/03055690500415746

Machado, A., Moro, S., Martins, L., & Rios, J. (2008). Qualidade do ensino em Matemática: determinantes do desempenho de alunos em escolas públicas estaduais mineiras. *Revista ANPEC*, 9, 23-46.

Microsoft SQL: Management Server. Versão 18.9.2. Microsoft Corporation, 2018. Disponível em: https://learn.microsoft.com/pt-br/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver16.

Moraes, C. P. de. (2021). *Avaliação educacional no ENEM: Um estudo sobre qualidade e equidade.* (Tese de doutorado), Universidade Federal do Rio de Janeiro, Programa de Engenharia de Sistemas e Computação, Rio de Janeiro, Brasil.

Moraes, C. P. de, & Peres, R. T. (2022). Reflexões sobre diferenças de desempenho no ENEM: Uma análise socioeconômica e escolar do Sudeste do Brasil. *Jornal de Políticas Educacionais*, 16, e85377. https://doi.org/10.5380/jpe.v16i0.85377

Moraes, C. P. de, Peres, R. T., Barbosa, M. T. S., & Pedreira, C. E. (2022a). Equity and performance in the National High School Exam: A study on sex and

race in Brazilian municipalities*. Education Policy Analysis Archives*, 30(68). https://doi.org/10.14507/epaa.30.6971

Moraes, C. P. de, Peres, R. T., Barbosa, T., & Pedreira, C. (2022b). Efeito escola a partir de indicadores educacionais: análise entre escolas públicas e privadas no ENEM. *Revista Meta: Avaliação*, 14(42), 67-93. doi:http://dx.doi.org/10.22347/2175-2753v14i42.3625

Moraes, C. P. de., Peres, R. T., & Pedreira, C. E. (2021). Eficácia escolar e variáveis familiares em tempos de pandemia: um estudo a partir de dados do ENEM*. Interfaces da Educação*, 12(35), 635-658.

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

Peres, R. T., Campos, L. B. A., & Moraes, C. P. de. (2025). Considerações acerca dos impactos da renda familiar e da escolaridade materna no desempenho em Matemática no ENEM 2022. *Revista Educação E Políticas Em Debate*, 14(2), 1-15. https://doi.org/10.14393/REPOD-v14n2a2025-75359

Plug, E. (2004). Estimating the effect of mother's schooling on children's schooling using a sample of adoptees. *American Economic Review*, 94(1), 358-368.

Reis, M. C., & Ramos, L. (2011). Escolaridade dos pais, desempenho no mercado de trabalho e desigualdade de rendimentos. *Revista Brasileira de Economia (Impresso)*, 65, 177-205.

Rstudio: Software. Versão 4.1.2, RStudio Corporation, 2021. Disponível em: https://posit.co/download/rstudio-desktop/.

Soares, J. F., & Collares, A. C. M. (2006). Recursos familiares e o desempenho cognitivo dos alunos do ensino básico brasileiro. *Dados*, 49(3), 615–650.

Tableau: Software. Versão 20.2.17. Salesforce Company, 2020. Disponível em: https://www.tableau.com/pt-br/support/releases/desktop/2020.2.17.

Travitzki, R. (2013*). ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade* escolar. (Tese de Doutorado), Universidade de São Paulo, Faculdade de Educação, São Paulo, Brasil.

Travitzki, R., Ferrão, M. E., & Couto, A. P. (2016). Desigualdades educacionais e socioeconómicas na população brasileira pré-universitária: Uma visão a partir da análise de dados do ENEM. *Arquivos Analíticos de Políticas Educativas*, 24(74).