FOLIA 407

Annales Universitatis Paedagogicae Cracoviensis

Studia ad Didacticam Mathematicae Pertinentia 16(2024)

ISSN 2080-9751
DOI 10.24917/20809751.16.4

Guilherme Stecca Marcom, Mayra Picoloto Papani

The Analysis of Mathematics Performance in the National High School Exam in Brazil*

Abstract. This study examines the performance disparity in Mathematics tests of the National High School Exam (ENEM) in Brazil in 2022. The ENEM, created in 1998 and restructured in 2009, not only assesses educational quality but also influences social standards and educational practices. Using microdata from candidates who completed high school in 2022 and participated without scoring zero in any section, the analysis includes 909,625 students from public and private schools across two geographical regions of Brazil. Statistical methods, including Cohen's distance, revealed that students from private schools consistently outperformed those from public schools across all regions. Additionally, male candidates tend to outperform females in Mathematics. The results emphasize the need for targeted interventions to mitigate disparities based on school type and gender, highlighting the importance of equitable and high-quality educational policies in Brazil.

Introduction

During the 1990s, a series of initiatives to evaluate the Brazilian educational system began. It created a sequence of tests in the federal sphere as Brazil Assessment (Prova Brasil), and National High School Exam (ENEM) among others. Thus, large-scale assessments have come to play a key role in Brazilian education. These examinations are intended to describe the situation of the various units of the federation and public policies for investment in education (Bruhn et al., 2016; Krasilchick, 2000). Global institutions such as the World Bank, the Inter-American Bank, and UNESCO use these indicators to foster and finance projects in education or other areas they support (Bruhn et al., 2016).

In 1998, the National High School Exam – Exame Nacional do Ensino Médio (ENEM) in Portuguese – was created with the initial proposal of evaluating the

*2020 Mathematics Subject Classification: *97B20, 97D60, 97C60*
Keywords and phrases: *mathematics teaching, ENEM, performance, socioeconomic status, gender*

profile of students completing high school, allowing public authorities to identify failures in the youth education process (Inep, 2009). Although the disclosure of the average score of each school was not part of the initial proposal of the exam, these results have become indicators of school quality. However, the analysis of student performance in ENEM, with the idea of improving the quality of education and learning, which would contribute to identifying school factors that influence student performance, is little explored, despite the recognition of the importance of these analyses.

The ENEM 2009 underwent a major transformation and is now known as the "New ENEM". The importance of assessments such as ENEM and the selection tests hold in Brazilian society, goes far beyond the measurement of education in the country. These large-scale assessments exercise a normative function that is often more powerful than official programs, textbooks, or curriculum proposals, having a great influence on primary and secondary education (Krasilchick, 2000, p. 90).

From this point of view, the importance of the New ENEM in promoting access to higher education and shaping the school environment highlights the need for social equity in its structure. This means that the candidate, regardless of their social situation, should not be prejudiced by a question on the test. Based on the need for equity in testing, regarding the gender and social situation of the candidate, some authors point out that there are marked differences in the performance of male and female candidates in assessments. In mathematics assessments, male candidates perform better than female candidates (Marcom & Aragão, 2022; Burgoa Etxaburu, 2017; Aylon & Liveh, 2013; Hyde & Linn, 2013). Authors such as Bourdieu (2002) point to an increase in reflection on the influence of socio-cultural factors on the results of ENEM. Recognizing that a question may favour men or women, private or public schools can provide valuable insights for action in different educational spaces. The disclosure of these results is extremely important, as it enables teachers and public administrators to take action to reduce inequality in large-scale assessments and education systems. From this point of view, the notion of cultural capital is an indispensable concept to explain the inequality of school performance of children and adolescents coming from different social classes (Bourdieu, 2002, p. 73).

To identify the impact of extracurricular and gender factors on the performance of basic education graduates who took the ENEM test in 2022, we will adopt a geographical perspective in the analysis, based on the concepts of large and small scale. Analyzing a phenomenon at different scales provides not only deeper insights but also the possibility of generalizations especially when working with different school groups whose unique characteristics may influence their performance.

Corrêa (2003) presents a study on urban space, using the concept of small scale as a cartographic ratio, for example, 1:5,000,000. Large scale would be the same cartographic ratio with a larger value than that of small scale, such as 1:10,000. In this work, we will use this concept of small scale to analyze the national results (Brazil) across the four areas of knowledge and writing in the ENEM test, and the concept of large scale to examine local results. We study the phenomenon at four levels of scale: the first level is Brazil (small scale), the second is the geographic

region, the third is the unit of federation (state), and the smallest level is the city (large scale).

Brazil is a continental country with big socioeconomic differences across its regions. For this analysis, we decided to analyze two contrasting regions: the richest, the Southeast (SE), and the poorest, the North (N). The Southeast region includes four states – São Paulo, Rio de Janeiro, Espirito Santo, and Minas Gerais – while the North region comprises seven states – Amazonas, Acre, Amapá, Pará, Rondonia, Roraima, and Tocantins. For the scale analysis, we use the most states from each region: São Paulo (SP) and Pará (PA). At the largest level, we compared two cities with similar populations: Belém, (in Pará, the North region) and Campinas (in São Paulo, the Southeast region).

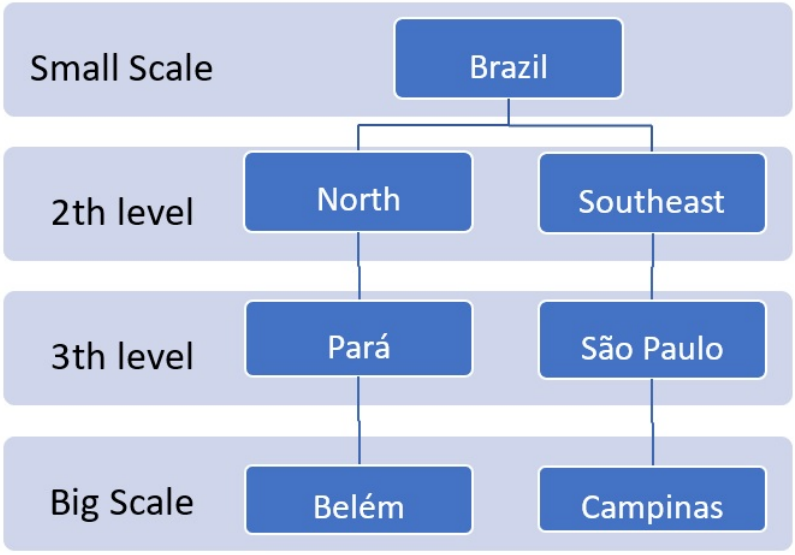

Figure 1: Diagram of the scale-based analysis of mathematics performance in the ENEM

Thus, this study aims to present and analyze the performance results of candidates completing basic education in the ENEM test in 2022, across different geographic scales. The data used for this analysis were provided by the National Institute of Educational Studies and Research Anísio Teixeira (INEP) and processed using the statistical software SPSS 22. These data will be presented as follows: first, the overall scores of students for each area of knowledge plus the overall and writing; second, the scores according to the type of educational institution attended (public or private); and finally, the scores disaggregated by gender of the candidate, in association with the type of educational institution.

According to Inep (2009), the ENEM establishes general and specific competencies and skills in each area of knowledge assessed in the exam. In the case of Mathematics, the ENEM Reference Matrix presents seven competencies and 30

skills, distributed across the different content of areas taught Basic Education in Brazil (which includes both Elementary and Secondary Education).

Out-of-class relations and the teaching of mathematics

ENEM candidates complete a socioeconomic questionnaire, made available by INEP online, which is used by researchers from different areas. This set of information makes it possible to analyze the performance of candidates according to their socioeconomic and gender conditions. Research relates socioeconomic status to academic performance became known worldwide through the Coleman, Plowden, and INED studies, the latter being a French government body for demographic studies, between the 1960s and 1970s. In France, the work of Pierre Bourdieu and Jean-Claude Passeron contributed to the development of the Theory of Reproduction, which argues that schools reinforce the social inequalities present in society (Bourdieu & Passeron, 2008).

We will base our study on the investigation of two key socioeconomic constructs: *symbolic violence* – a form of violence that occurs in an almost invisible way – and *cultural capital* – which is a set of actions that help explain performance differences among students from different social classes. According to Bourdieu and Passeron (2008), the relationships that occur within school environments are governed by the relations present in symbolic violence. "All power of symbolic violence, that is, all power that comes to impose meanings and to impose them as legitimate" (Bourdieu & Passeron, 2008, p. 25). In this way, symbolic violence is primarily exercised through social structures such as the educational system, the media, and other institutions that help perpetuate power relations. These institutions contribute to the naturalization of social inequalities by presenting the conditions imposed by dominant groups as "normal" or "just".

This process of symbolic violence affects all students, but its impact is more strongly felt by individuals who do not belong to the dominant social classes and are distant from the dominant culture – one that is legitimized by the school. This distance forces individuals to make greater efforts to acieve academic, as they are less familiar with these symbols that are taken for granted within the school environment (Bourdieu & Passeron, 2008). Through symbolic violence, schools reproduce social structures, which influence students learning and result in performance disparities between groups, particularly in large-scale assessments. Such outcomes help to understand why such variables cannot be excluded from analyses of both school performance and results of national exams such as the ENEM. Since such variables strongly influence the school life of individuals, most classrooms are composed of people who do not begin from the same economic, social, and knowledge levels.

One way to explain these results is through the concept of cultural capital, as proposed by Bourdieu (2002) in his theory of the sociology of education. According to the author "[...] first imposed itself as an indispensable hypothesis to account for the inequality of school performance of children coming from different social classes [...]" (Bourdieu, 2002, p. 73). For Bourdieu, cultural capital is a series of resources that are incorporated from the relations that the individual has with their

surrounding environment (Bourdieu, 2002). This environment transmits a series of implicit and deeply internalized values – which can be understood as synonymous with capital – which contribute to defining the attitudes towards the cultural capital itself and the educational institution (Bourdieu, 2002). This environment can be defined by the social relationships to which students are exposed and the level of cultural capital present within those relationships. In this context, the family is recognized as the main source of the cultural capital of students.

Performance differences are also observed when comparing the gender of the individuals. These differences have been documented in research on academic performance both within schools and in large-scale assessments such as ENEM and PISA, at national and international levels, particularly in the area of Natural Sciences (Marcom & Aragão, 2022; Burgoa Etxaburu, 2017; Aylon & Liveh, 2013; Hyde & Linn, 2013). This research has shown how there are significant gender-associated differences in performance within the school environment and during assessments, particularly in Mathematics. The results show a better performance of males in these areas compared to the performance of females.

To understand these results, we can analyze them from the perspective of the sociology of education, drawing again on the work of Pierre Bourdieu, who seeks to describe how society builds, reproduces, and transmits certain social standards – primarily through symbolic violence and cultural capital, as described above.

For Bourdieu, relations of domination are not easily perceptible, because as the author himself emphasizes the lasting effects that the social order exerts on women (Bourdieu, 2010). More explicitly, it is necessary to understand how symbolic relations present in society are transmitted and how they shape physical, cognitive, and psychological characteristics of both men and women.

In the context of gender relations, Bourdieu shows how the model of social reproduction through cultural capital differs for boys and girls, an example of which is an emphasis on literary culture by women and scientific culture by men (Bourdieu, 2007). Since childhood, boys are socialized with greater emphasis on certain cultural characteristics and skills related to Natural Sciences and Mathematics, while girls receive more encouragement in the area of Languages. This means that the symbolic framework associated with Mathematics is more easily recognized by boys, while girls are socially and culturally distanced from these symbols and values. In this way, girls are more susceptible to the effects of the symbolic violence within the domain of Mathematics.

Methodology

Scalar tool

Scalar dimensions are important tools for analyzing individual performances. They can be described using two concepts: horizontality and verticality. The horizontal dimension emphasizes the relationships between parts and the whole, contexualizing a phenomenon, object, or aspect spatially and temporally, and comparing it with others from its locality. It highlights particularities, singularities, and relationships in the search for patterns (Compiani, 2007, p. 34). Thus, the

concept of horizontality describes the particular aspects of the analyzed object, taking into account the relationships between the parts and the whole. In our work, this means identifying differences in performance for each geographical region characterized in this study.

The other dimension of analysis, the vertical, is characterized by Compiani (2007) as the dimension that emphasizes the procedural nature of phenomena and objects. It examines the different contexts of horizontality, seeking to explain them as wholes or sets, based on properties and towards decontextualization and generalization, where time and space are "erased". The vertical dimension reflects the relationships between the parts and the whole, thus explaining the phenomenon globally (Afonsin, 2021), which allows for the generation of generalized results. In our research, the vertical dimension reflects the generalization of our results, representing national characteristics associated with mathematics teaching.

In similar studies that compare the performance of different students on large-scale assessment tasks (Marcom & Kleinke, 2021; Marcom & Aragão, 2022), the concepts of scale are employed to understand the performance of each group within its context (particularities) and to subsequently generalize the results (phenomenon). In this way, employing this practice of scalar analysis allows for the observation and examination of data across various analytical dimensions. This approach facilitates a broader perspective on the results, seeking different ways to describe them.

Statistical tool

Measuring differences in performance enables comparisons and analysis of the various groups that make up a sample. This measurement is one of the main tools most commonly used in evaluation research. Cohen's distance (Cohen, 1988) makes it possible to measure the statistical distance between samples in standard deviation units (Cohen, 1988). We chose to use this test because of its simplicity and ease of interpretation. Thus, to measure the statistical distance between two groups, we can use the formula (1) described below:

$$d = \frac{A_f - A_r}{SD_{rf}} \tag{1}$$

where d is what is called Cohen's distance; A_r and A_f are the means of the reference and focal sample, respectively. SD_{rf} is the standard deviation of the total population formed both the focal and reference groups. The values of d can be either positive or negative, depending on the focal means of the and reference group. Thus, a positive d value indicates that the focus group performs better than the reference group. If a d value is negative, it means that the reference group performs better than the focal group.

To determine whether the difference between the compared groups is significant, a metric was established to interpret the values of d. When the value is up to 0.01, the difference is not statistically significant, indicating no meaningful difference between the groups. When the d value is between 0.01 and 0.35, the difference is small, but present. When the d values are between 0.35 and 0.65, the

difference is moderate, while values above 0.65 are considered large differences. It is important to note that this scale is valid for both positive and negative d values.

Results and discussion

For this study, we used the microdata of ENEM 2022, available on the INEP website , as our primary source of information. The sample was selected based on the following criteria: students who completed basic education in the same year as the ENEM 2022; who participated in the both days of testing; and who did not score zero on any of the tests or the essay. In total, 909,625 candidates from both public and private schools were analyzed.

Table 1: Scale of analysis, geographic region of Brazil, type of school, mean, population, and standard derivation of ENEM 2022 data

Scale	Region	Type of Schools	Mean	N	Std. Deviation
Big Scale	BR	Public	520.15	714,552	101.04
		Private	616.79	195,073	118.39
		Total	540.87	909,625	112.24
2th Level	N	Public	489.73	73,271	90.27
		Private	583.91	10,013	117.20
		Total	501.05	83,284	98.79
	SE	Public	541.67	236,427	102.84
		Private	624.42	94,701	116.43
		Total	565.34	331,128	113.25
3th Level	PA	Public	493.35	19,311	91.83
		Private	582.50	5,269	118.14
		Total	512.46	24,580	104.66
	SP	Public	549.63	86,674	100.51
		Private	624.60	47,142	115.14
		Total	576.04	133,816	111.79
Small Scale	Belém	Public	493.78	7,312	92.72
		Private	587.55	3,163	120.10
		Total	522.09	10,475	110.49
	Campinas	Public	553.90	2,933	105.52
		Private	631.05	1,766	115.03
		Total	582.90	4,699	115.40

The relationship between public and private schools and the results presented in Table 1 is complex and multifaceted. Private schools often benefit from greater financial resources, which can be reflected in more modern infrastructure, higher salaries for teachers, and greater access to quality educational materials. In contrast, public schools often face challenges related to insufficient funding, overcrowded classes, and a lack of adequate educational resources (Klein et al., 2007). These disparities can influence ENEM results, with private school students generally achieving higher scores due to these advantages (Travitzki, 2013). However, it is important to acknowledge that other factors, such as students' socioeconomic status and family support, also play a significant role in ENEM results (Marcom & Kleinke, 2021; Travitzki, 2013).

In Brazil, there is a strong correlation between socioeconomic status and the type of school in basic education. Students from public schools are generally from more disadvantaged socioeconomic backgrounds. Studies by Klein et al. (2007), Marcom (2015), Oliveira (2014), and Travitzki (2013) have shown that socioeconomic factors have a strong impact on candidates' performance on the ENEM. These studies indicate that students from less privileged social classes have lower success rates privileged ENEM items. The differences in the average scores shown in Table 1 reflect this divide between groups. Moreover, this difference in performance occurs at all levels of the scalar analysis, reinforcing the impact of socioeconomic conditions on student outcomes as documented in the literature.

The data reveal that mean score on the Mathematics test is higher for candidates from private schools compared to those from public schools. However, it can also be inferred that public schools play a significant role in reducing gender differences in Mathematics performance. It is important to consider not only the impact of the school environment on the candidates performance but also the role of cultural capital, which is lower among students in public schools. Therefore, it was anticipated that Cohen's distance for the Mathematics test would be smaller in public schools. However, in order for this statement to be better analyzed, it is necessary to apply the Student's t-test to the data. The application revealed a value of 79.42 ($p>0.01$), which confirms a statistically significant difference in favour of private schools.

Table 2: Cohen's d for Comparison of School Types by Scale and Geographic Region of Brazil

Scale	**Region**	**d**	**Region**	**d**
Big	BR	0.86	BR	0.86
2nd	N	0.95	SE	0.73
3rd	PA	0.85	SP	0.67
Small	Belém	0.85	Campinas	0.67

We observe that, in nearly all tests, the values demonstrate a large statistical distance between the groups. It is important to consider the Human Development Index HDI in these regions is relatively high when compared to the national scale. This suggests that the development index improves as we move from a smaller to a larger scale. According to Bourdieu (2002), the social dimension significantly affects students' performance; thus, socioeconomic factors – such as differences in the Human Development Index (HDI) – may result in performance disparities in exams outcomes, as highlighted by Travitzki (2013). As a result, even public schools tend to have a better performance than most other schools across the country. This is reflected in the decrease of Cohen's d as we transition from small to large scales.

In this way, we observe that candidates from private schools consistently outperform their public school counterparts across all areas of the ENEM exam, especially in Mathematics and Natural Sciences. According to Bourdieu (2002, p. 73), cultural capital is an indispensable concept to explain the performance inequality among children from different social classes, as well as the conservative nature of the school institution, which tends to reproduce the cultural patterns of the social universe experienced by students.

Analyzing the data of each scale, we can say that candidates from private schools have greater cultural capital than those from public schools. This capital stems not only from the type of education they receive but from the cultural patterns that these students inherit from their families. Thus, students from private schools whose families have higher socioeconomic status hold greater cultural capital, which contributes to their performance in assessments such as the ENEM and other selection tests.

Another striking point in the horizontal dimension related to gender differences is performance in mathematics. It is observed that male candidates outperform female candidates in the mathematics test, regardless of the educational institution.

The data in Table 3 indicates, first, that the performance of women is consistently lower than that of men on the Maths test across all scales of analysis. This result can be considered an indicator of the gender gap in secondary education in Brazil. However, to further validate this finding, it is necessary to apply the Student's t-test to this data. The application revealed a value of 77.56 ($p>0.01$), corroborating the existence of a significant difference in favour of men in relation to the performance of women in the Maths test.

To better visualize this performance differences based on gender, we calculated Cohen's d using as the focal group.

Table 4 shows that, at all scalar levels, the distances between the groups always remain statistically significant. According to Cohen's scale, all values are considered moderate. Therefore, the results presented in Table 4 are in line with what has already been presented in the international literature on the subject (Marcom & Aragão, 2022; Aylon & Liveh, 2013; Hyde & Linn, 2013; Ryan & DeMark, 2002).

Table 3: Scale of Analysis, Geographic Region of Brazil, Gender, Mean, Population and Standard Deviation of ENEM 2022 Data

Scale	Region	Gender	Mean	N	Std. Deviation
Big Scale	BR	F	523.53	541,449	105.37
		M	566.38	368,176	117.06
		Total	540.87	909,625	112.24
2nd Level	N	F	486.76	49,672	91.91
		M	522.18	33,612	104.64
		Total	501.05	83,284	98.79
	SE	F	545.99	197,149	107.04
		M	593.81	133,979	116.10
		Total	565.34	331,128	113.25
3rd Level	PA	F	494.63	13,861	97.66
		M	535.52	10,719	108.83
		Total	512.46	24,580	104.66
	SP	F	556.42	74,569	106.34
		M	600.74	59,247	113.57
		Total	576.04	133,816	111.79
Small Scale	Belém	F	504.94	5,986	103.95
		M	544.97	4,489	114.75
		Total	522.09	10,475	110.49
	Campinas	F	561.84	2,744	106.49
		M	612.46	1,955	120.82
		Total	582.90	4,699	115.40

Table 4: Cohen's d for Comparison of Gender in Scale and Geographic Region of Brazil

Scale	Region	d	Region	d
Big	BR	0.38	BR	0.38
2nd	N	0.36	SE	0.42
3rd	PA	0.39	SP	0.40
Small	Belém	0.36	Campinas	0.44

We observed in the horizontal dimension that there are significant differences in the performance of Public and Private schools, as well as between the performance about gender. The most important was to identify that these differences were

maintained in all scales analyzed, this means that there is a pattern that repeats itself, this way we can propose a generalization of these results. In this way in the vertical dimension, we can infer that the candidates of the private school will have higher performance than the candidates of the public school, this result is explained not only by the differences that exist in these two types of education in the country but also to the impact that the cultural capital has in the performance of the students in evaluations of this type.

Another generalization that we can make is related to the performance of candidates with regard to gender. We infer that, in large-scale assessments, male candidates tend to perform better in Mathematics tests than female candidates. Similar results are presented by Marcom & Aragão (2022), Hyde & Linn (2013), and Aylon & Liveh (2013), whose studies involved students from around the world. These findings demonstrate that gender-based performance differences in mathematics occur regardless of the scale and country. However, the values presented in these international studies show that Cohen's d value is small, unlike the values found in our study, which are considered moderate. This shows that in Brazil, gender differences are more pronounced than in other places around the world.

Some authors try to explain the gender difference in mathematics test by associating it with the perception among female candidates that the subject is not relevant to their future careers (Else-Quest et al., 2010; Aylon & Liveh, 2013). This perception is already observed in the 7^{th}-grade girls, who begin to experience difficulties with mathematics and lose interest in this subject (Else-Quest et al., 2010). However, we argue that the underlying causes of this disinterest are rooted in the social order, such as low family encouragement for girls to pursue scientific careers and the differentiated treatment by teachers, who tend to encourage boys more than girls to be interested in Mathematics.

Final considerations

Based on the data, we observed that there are differences in the general performance of candidates and high school graduates, as we move from the small scale (Brazil) and to the large scale (Campinas). These differences are closely linked to variations in the Human Development Index (HDI) across these scales.

The most striking conclusion of this study was the identification of differences in the performance of candidates associated with gender. Our findings showed that there is a moderate difference in the candidates' performance on the Mathematics test, with male candidates outperforming female candidates. This difference was evident across all scales analyzed and was independent of the type of educational institution, providing sufficient data to generalize that in large-scale evaluations such as ENEM, male candidates tend to perform better than female candidates in Mathematics. Thus, we conclude that scalar analysis is a methodology of extreme importance, especially in the context of student performance evaluation, since it enables a deeper analysis of specific phenomena and allows for less subjective generalizations about school performance. Thus, it is essential that the school community be aware of these findings because they have a significant impact on the teaching of Mathematics.

References

Alfonsin, B. M. (2021). *O direito à cidade como agenda de pesquisa: coletânea de doze anos de investigação científica do Grupo de pesquisa em Direito Urbanístico e direito à cidade da FMP*. Editora Dialética.

Ayalon, H., & Livneh, I. (2013). Educational standardization and gender differences in mathematics achievement: A comparative study. *Social Science Research*, **42**(2), pp. 432–445.

Bourdieu, P. (1998). *Escritos de Educação*. Maria Alice Nogueira e Afrânio Catani (org.). Petrópolis, Rio de Janeiro: Vozes, p. 67.

Bourdieu, P. (2002). *A escola conservadora: As desigualdades frente à escola e à cultura*. Petrópolis, RJ: Vozes.

Bourdieu, P. (2004). *Usos sociais da ciência*. Unesp.

Bourdieu, P. (2007). *A distinção*. São Paulo: Edusp.

Bourdieu, P. (2010). *A dominação masculina*. Rio de Janeiro: Bertrand Brasil.

Bourdieu, P., & Passeron, J. C. (2009). *A Reprodução: Elementos para uma Teoria do Sistema de Ensino*. Petrópolis: Ed. Vozes.

Brasil. (2012). LEI Nº 12.711, de 29 de Agosto de 2012. *Diário Oficial da União*.

Bruhn, M., de Souza Leão, L., Legovini, A., Marchetti, R., & Zia, B. (2016). The impact of high school financial education: Evidence from a large-scale evaluation in Brazil. *American Economic Journal: Applied Economics*, **8**(4), pp. 256–295.

Burgoa Etxaburu, M. B. (2014). *La transferencia de contenidos matemáticos a contextos científicos: el concepto de función*. Servicio Editorial UPV/EHU.

Cohen, J. (1998). *Statistical Power Analysis for Behavior Sciences*. Hillsdale.

Compiani, M. (2007). O lugar e as escalas e suas dimensões horizontal e vertical nos trabalhos práticos: implicações para o ensino de ciências e educação ambiental. *Ciência & Educação*, **13**(1), pp. 29–45.

Corrêa, R. L. (2003). Uma nota sobre o urbano e a escala. *Revista Território*, **7**, pp. 11–12.

Moura Castro, C. (2005). *Crônicas de uma educação vacilante*. Rocco.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*, **136**(1), pp. 103.

Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science*, **314**(5799), pp. 599–600.

INEP. (2009). *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. ENEM (Exame Nacional do Ensino Médio): Matriz de Referência para o ENEM 2009*. Brasília.

INEP. (2021). *Sobre o ENEM*. Recuperado de `http://portal.inep.gov.br/web/enem/sobre-o-enem` (Access in 25/05/2021).

Klein, R., Fontanive, N. S., & Elliot, L. G. (2007). O exame nacional do ensino médio-tecnologia e principais resultados em 2005. *REICE – Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, **5**(2), pp. 116–131.

Krasilchik, M. (2000). Reformas e realidade: o caso do ensino das ciências. *São Paulo em Perspectiva*, **14**, pp. 85–93.

Marcom, G. S., & Aragão, T. Z. B. (2022). Gender Inequality and Science Education: Comparison with Brazilian Students in PISA and ENEM. *Acta Scientiae*, **24**(4), pp. 57–80.

Marcom, G. S., & Kleinke, M. U. (2021). Indicadores Formativos para o Ensino de Física através do Exame Nacional do Ensino Médio (ENEM). *Caderno Brasileiro de Ensino de Física*, **38**(3), pp. 1388–1419.

Ryan, J. M., & DeMark, S. (2002). Variation in achievement scores related to gender, item format, and content area tested. In Tindal, G. & Haladyna, T. M. (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* (pp. 67–88).

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

Guilherme Stecca Marcom
University of Campinas,
Gleb Wataghin Institute of Physics,
Campinas, São Paulo, Brazil
e-mail: marcomgs@unicamp.br

Mayra Picoloto Papani
University of Campinas,
Multiunits Program in Science and
Mathematics Teaching (PECIM/UNICAMP),
Campinas, São Paulo, Brazil
e-mail: m184537@dac.unicamp.br