Journal of Information and Data Management, 2025, 16:1, doi: 10.5753/jidm.2025.3657

EXPLORATORY ANALYSIS OF MICRODATA FROM THE NATIONAL HIGH SCHOOL EXAM - ENEM: PERFORMANCE AND SPECIFICITIES OF THE PARTICIPANTS

João Augusto Fernandes Barbosa [**University of São Paulo** | ***joaoaugusto_fb@usp.br***]
Robson Leonardo Ferreira Cordeiro [**Carnegie Mellon University** | ***robsonc@andrew.cmu.edu***]

Institute of Mathematics and Computer Science (ICMC), University of São Paulo (USP), Brazil.

Received: 4 September 2023 • **Published:** 2 November 2025

Abstract The National High School Exam (ENEM) is a significant test in Brazil that measures high school teaching quality and performance. It has also been used for evaluating undergraduate course candidates since 2004. ENEM has had a transformative impact on the education market, with schools now prioritizing exam preparation. However, there is a lack of comprehensive studies on the performance and characteristics of ENEM participants, particularly those with disabilities such as attention deficit or autism spectrum disorder. This article examines the challenges faced by these subgroups of participants considering the period between 2015 and 2019, and using analytical tools as clustering, heatmaps, and hypothesis testing to understand the main data patterns. The findings aim to support the development of more tailored and flexible study programs to meet the needs of participants. Our study reveals that individuals with certain disabilities, like Attention Deficit and Dyslexia, tend to achieve higher scores, while those with Mental disabilities and Deafness perform below the national average. Additionally, the results suggests that the grade disparity between students with and without disabilities may be influenced by socioeconomic factors.

Keywords: ENEM, DISABILITY, EXPLORATORY DATA ANALYSIS, EDUCATION, PERFORMANCE

1 Introduction

The ENEM - National High School Exam - is a standardized Brazilian national exam aiming to verify the quality of high schools while serving as the main gate to the Universities in the country. It consists of two days of testing, totaling 180 questions. The contents evaluated are Languages, Codes and their Technologies, Mathematics, Natural Sciences, Human Sciences, and an essay [Brazileducation, 2023].

Due to its academic significance, ENEM is always a subject of debate. Nevertheless, studies on the exam itself and its participants are rare. Still, such studies are essential, since through ENEM it is possible to find evidence of the main flaws and successes in Brazilian elementary and secondary education, which may help improve basic education.

However, in practice, such research does not happen. Recently, the FAPESP – São Paulo Research Support Foundation – published an article [Julião, 2021] that details exactly the fact that, despite having indicators on various segments of Brazilian education, rarely they are used. According to that paper, the head of FAPESP, Marco Antonio Zago, said: "Our basic education demands reforms in many aspects. But a central issue is that its planning and execution need to be much more grounded in science, in data, in evidence, than they are today. It is necessary to apply the results of the accumulated scientific evidence and, at the same time, work to have more information". This problem is very challenging to resolve. However, we seek to provide insights that may be helpful in addressing the issues related to this matter.

We present in our article a study that aims to serve as a basis for educational policies on particular issues, through the exploratory analysis of microdata from the ENEM exams conducted from 2015 to 2019, as well as the interpretation of the results obtained. Our work seeks to analyze the main difficulties of groups of participants with some type of disorder/disability, such as autism, attention deficit disorder, and dyslexia,by clustering them according to their performance, and using other analytical tools like heatmaps and hypothesis testing to understand the main data patterns. Therefore, we verify how individuals with each of these conditions perform in each subject. For example, it often assumed that people with autism spectrum disorder have certain above-average abilities. Discussions that revive the possibility of people with undeniable talent being autistic are not rare [Faria, 2013]. Can such supposed above-average skills be identified in the National High School Exam? Does the performance of these candidates stand out from the others? These and other related questions are studied in our work.

The second part of our article addresses groups with disabilities that presented very different results than expected. We studied their performance in the essay of the exam which is often considered as its most important part. It is in the essay that the capacity of interpretation and argumentation of those enrolled is put into question. Through the skills and indicators provided, we intended to study and understand the main mistakes made by the participants during the writing of their essays. The results obtained are analyzed taking into account socioeconomic contexts of each group of participants.

Intervention measures in teaching are out of the scope of our work because they must adapt to the reality of each educational institution and require prior planning. We aim to be grounded on the results of the analysis of ENEM's microdata to report the main trends, positive or negative, and, whenever possible, suggest actions to be considered by education professionals taking into account the local particularities.

Our results express that a large portion of individuals with disabilities has the same pattern score as those with no disability. The largest differences regard the Attention Deficit and Dyslexia groups, which have better-than-average grades. On the other hand, participants with Deafness and Mental disability have a score considerably inferior to expected. Also, a socioeconomic analysis suggests that those disabled groups with a high average score are from higher social classes.

We hope the information obtained may be of use to academic society as a whole, whether to support future research or to base institutional measures and debates that seek a better educational system. Brazil has experienced both economic and educational difficulties with huge cuts in funding for schools at all levels of education. Because of this, the investments made must be appropriately allocated, with well-planned and well-executed expenses. We believed that our study will be useful to guide such decision-making.

2 Goals

The major goal of this work is to serve as an overview of education in unexplored contexts. We aim to answer the following questions regarding participants with disability:

Q1 **Major difficulties per disability:** What are the major difficulties of students with each type of disability? Are their grades similar to the average?

Q2 **Performance per disability:** How each disability group behaves based on its performance when compared to the other ones? And how are the students divided considering High, Average, or Low performance?

Q3 **Essay difficulties:** What are the main difficulties for students in writing a good essay? Are these difficulties most related to technical writing questions (lexical and syntactic) or world knowledge (semantic)? Do the difficulties change for particular types of disabilities?

Q4 **External influence:** Is there any external factor that could relate to the differences in grades by disability?

Q5 **Statistical significance:** Are the results obtained for the previous questions statistically significant?

3 Related Work

Previous studies have analyzed ENEM's microdata made available by the National Institute of Educational Studies and Research Anísio Teixeira (INEP). The novelty of our work is to comprehend how students with disabilities integrate into the Brazilian educational landscape, by actively identifying, validating and describing patterns, which is precisely the main goal of data mining. We provide better insights into the current state of the school environment and how it has been accommodating or challenging for students with disabilities.

Lima *et al*. [2020] analyzed the performance of ENEM participants, separating them by socioeconomic characteristics, such as their location and school of origin. The analysis methodology used the K-means clustering algorithm, and in total, six editions of the exam were analyzed. Among other findings, it is remarkable the evolution of the average score obtained in ENEM by participants from the northeast region of Brazil. In addition, there was also a reduction in the total number of participants with some type of disability from public schools, while in private schools, the number of such participants increased. This fact suggests a possible migration of students with disabilities from public to private schools.

Terra [2019] focuses on predicting the scores of ENEM participants based on the answers obtained by the socioeconomic questionnaire completed in the exam registration. Through the work, the authors learned that the most important characteristics for the prediction of scores were the family's monthly income, what type of school the participant came from, the quality of his/her home infrastructure, e.g., the number of computers, access to the Internet, etc., and the level of education of the participant's mother.

Another work following a similar approach was performed by Stearns *et al*. [2017]. The authors used different methods of machine learning to find the most relevant features. The work was based both on clustering using the ADA algorithm and decision trees using Gradient boosting. The results were slightly different from those of Terra's work. In this scenario, the rank of features by their relevance shows first the location (Longitude, Latitude) and the student's age. Only then, indicators such as type of school and wage income are shown.

The goals of de Oliveira *et al*. [2020] are partially similar to ours. This related work uses a machine learning algorithm – the decision tree C4.5 – to predict the scores of participants with disabilities in the 2018 edition of ENEM. Their results match our own results; they suggest that participants with Attention Deficit performed better than average, while those with Mental Disability and Deafness tended to perform bellow average. Distinctly from our work, de Oliveira *et al*. [2020] do not analyze the major difficulties per disability, nor the essay difficulties and the external influences. Their work is also restricted to a single edition of ENEM.

Provided that we analyze the performance of the participants in each of the ENEM tests, it is interesting to highlight the work of Simon and Cazella [2017], which seeks to correlate the performance in the Natural Sciences test with the socioeconomic characteristics of the candidates. Their results suggest that the higher the socioeconomic level of the candidate, the greater are his/her chances of performing well.

Additionally, Thiago de Souza [2021] studied the profile of the participants of the ENEM's 2019 edition. His work relates the participants' performance to their age, location, school, race, parents' level of education, wage, and access to technology. The results indicate better scores in the Midwest region of Brazil, a higher score for those participants from private schools, and a large impact of the parents' education on the participants' performances. The higher the parents' educational level, the larger the participant's grades. In addition, it was noticed that those participants with access to technology had 28.58% larger scores on average.

Taking a closer proximity to the participants, Leria *et al*. [2023] article becomes intriguing as it conducts field research with visually impaired participants in the ENEM. The study revealed that the challenges of this entrance exam extend beyond the test itself. A significant portion of the infrastructure appeared to be limited for these participants, who reported confusing image descriptions, excessively long texts, and significant technological limitations. It was noted that the difficulty of composing an essay verbally, dictating it to a transcriber without the use of a computer with screen reader software, greatly complicated the time management. In this context, it can be observed that the guidelines of the ENEM itself pose limitations for participants with disabilities, extending beyond the content studied in high school.

In Viggiano and Mattos [2013], the main focus is the analysis of the results obtained according to each participant's location. The best scores were those from the South and Southeast regions of Brazil, while the Center-West region maintained an average performance and the North and Northeast regions presented a lower-than-average performance. A similar geographical analysis is carried out in our own work to verify possible similarities or divergences and determine if such regional characteristics were perpetuated in more recent editions of the exam, in addition to investigating whether they are also valid for subgroups of participants.

It is important to realize that socioeconomic and psychobehavioral characteristics have increasingly been linked to candidates' performance in exams and contests. In this case, studies like van Ewijk and Sleegers [2010], Polyzou and Karypis [2019], and O'Neill *et al.* [2012] provide several interesting analyses on the profile of candidates around the world. In the international scenario, Zhang *et al.* [2007] draw an interesting parallel with our own study but focuses on students with disabilities in North Carolina. It was found that most candidates performed below average, but with some exceptions excelling in subjects such as Algebra. The study also suggests as future work deeper analyses of candidates in high-stakes exams and how students with dyscalculia perform in Algebra tests. Importantly, both analyses suggested in this related work are performed in our work. A broader analysis of USA education focused in SAT, the College Board's Scholastic Aptitude Test, by Ragosta and Wendler [1992] revealed that students in disability groups took on average twice as long to complete the exam, with blindness requiring considerably more time. Also, issues such as accommodation and diagnosis were a challenge for the students. Contrary to our work, their main focus is related to the infrastructure of the exam — checking if the time was sufficient and the accommodations were good. However, they did not analyze the students' performance as deeply as we do.

To the best of our knowledge, except for the work by de Oliveira *et al.* [2020], no study delves as deeply into the performance of ENEM candidates with disabilities as we do. Even so, Oliveira's article has relevant limitations that we detailed before, both regarding sample size as they analyzed a *single edition* of ENEM, and also regarding the small number of disabilities considered. Because of that, our article holds significant novelty and offers valuable new insights to the community. Our study stands out among the related work due to its unique goals, results, and the broader understanding it provides regarding Brazilian education and the participant's conditions. We consider 11 different disabilities – a number much higher than the average of the studies in this field – and analyze 5 editions of the exam, with a significant amount of data that is much larger than in previous works.

4 Background

Clustering is the process of dividing a large dataset into smaller, more meaningful subsets (or clusters) of similar instances. It is usually seen as an unsupervised task. A clustering algorithm works independently to discover non-obvious patterns. Through that, the data is assigned to the most appropriate clusters containing only similar instances.

4.1 K-means

K-means [Ahmed *et al.*, 2020] is probably the most used clustering algorithm. Initially, it is necessary to inform the number of clusters K. With this information, it is created a hypercube with N dimensions, one per attribute of the dataset, where all instances are distributed according to their attribute values. Figure 1a shows a toy 2D dataset distributed as such.

Then, an initial choice of K centroids is made at random. All the points – which refer to dataset instances – are associated to the closest centroid. Figure 1b shows the random choice of centroids for the example data. The centroid color is red, and the clusters are shown by the different colors.

In the sequence, the position of each centroid is updated to be the mean of its points, and the points are reassociated to the closest centroid, as shown in Figure 1c. Because of this difference, it is possible to highlight some points that "changed their color". It means that the algorithm understood that they belong to a different cluster than the one initially chosen. The process of moving centroids and reassociating points is repeated until there is no change in the color/closest centroid of any point. Mathematically, we have:

1. Randomly choose k initial centers $\mathcal{C} = \{c_1, c_2, \cdots, c_k\}$.
2. For each $i \in \{1, 2, \dots, k\}$, let the cluster C_i be the set of points in dataset $\mathcal{X}$ that are closer to c_i than they are to c_j for any $j \in \{1, 2, \dots, k\}$ such that $j \neq i$.
3. For each $i \in \{1, 2, \dots, k\}$, update c_i to be the center of mass of all points in $C_i : c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$.
4. Repeat Steps 2 and 3 until there is no change in $C_1, C_2, \dots C_k$.

4.2 K++

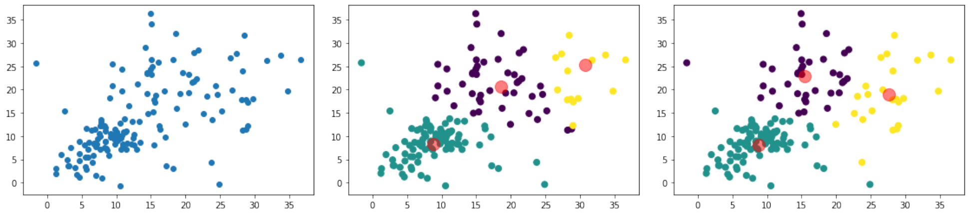
It is known that a different choice of the initial centroids can generate different results for K-means. This means results with a lower/higher speed performance and with a worse/better data division. It is in this scenario that the importance of the k++ algorithm fits [Arthur and Vassilvitskii, 2007]. It focuses on performing a better choice of the initial centroids. The process consists of the following steps:

1. Take a center c_1 chosen uniformly at random from $\mathcal{X}$.
2. Take a new center c_i, choosing $x \in \mathcal{X}$ with probability $\frac{D(x)^2}{\sum_{x \in \mathcal{X}} D(x)^2}$, where $D(x)$ is the shortest distance from a data instance x to the closest center previously chosen.
3. Repeat Step 2 until we have taken k centers altogether.
4. Proceed as with the standard k-means algorithm.

4.3 Elbow Method

Not only the initial centroids are important in the clustering, but, also, the number of clusters that the data is divided into. Actually, in data with many instances or dimensions, it is often hard to find the ideal number of clusters. This is because the division is a two-way decision. Too few clusters implies very different instances in the same cluster. Too many clusters means similar instances associated to different clusters.

The Elbow Method is an algorithm that works with K-means to determine the best K possible. The process consists in performing the clustering with different values of K. Then, for every K, it is calculated the squared sum of the distances from each point to its centroid. Finally, a simple plot



(a) Initial data **(b)** Centroids and colored clusters **(c)** Data after complete association

Figure 1. K-means process

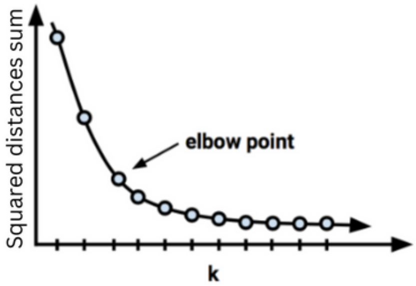


Figure 2. Elbow method, with the plot of squared sum of distances versus K. Illustration adapted from Dados [2021]

expresses the sum, as shown in Figure 2. It plots the squared sum of distances versus K. The optimal number of K is identified (manually or with an off-the-shelf elbow-detection algorithm) as the clearest "elbow" in the plot's curve. In the illustration of Figure 2, the "elbow", and also the best number of clusters for the K-means, is highlighted with an arrow.

5 Investigation

This section presents our main contributions, including methodology and results. We first present our dataset and then move on to the goals previously mentioned in Section 2.

5.1 Data Presentation

The raw data is composed of 5 files disposed by INEP [2022], each one containing 136 columns (attributes) and a variable number of rows (instances) according to the number of participants in each exam's edition. In general, each year's dataset has a volume between 3GB to 5GB. Importantly, Dyscalculia only started to be presented in 2015; it limited the analysis period of our article to be between the 2015 and the 2019 editions of ENEM. Microdata regarding editions after 2019 were not yet made available in INEP's repository.

In this dataset, in addition to disability indicators, there exist information about each participant such as hometown, age, race, gender, and others. Furthermore, it includes the participant's answers to the exam, data about his/her high school, and complete answers to the socioeconomic questionnaire. This questionnaire changed over the years, having 27 questions in some editions and reaching 50 in others. Its main objective is understanding the economic situation of the participant – trying to find out what his/her parents' relationship status is or if the participant has access to the Internet.

The data is presented with attribute names based on the dictionary disposed by INEP. For brevity, we use the shortened attribute names shown in Table 1. Also, Table 2 relates the ENEM's subjects with their acronyms. It is worth noting that we rely solely on the candidate's response to the exam questionnaire. This means that a candidate may have a disability and choose not to disclose it (either by personal choice or perhaps without even being aware of the disability). Another possibility is that someone may fabricate a disability to gain advantage such as extra time to perform the exam.

DISABILITY	ACRONYM
NO DISABILITY	ND
LOW VISION	LV
BLINDNESS	BL
DEAFNESS	DF
HEARING DISABILITY	HD
DEAFBLINDESS	DB
PHYSICAL DISABILITY	PD
MENTAL DISABILITY	MD
ATTENTION DEFICIT	AD
DYSLEXIA	DYx
DYSCALCULIA	DYc
AUTISM	AUT

Table 1. List of disabilities with the corresponding acronyms

5.2 Data Selection and Preprocessing

Knowing the dataset schema, as well as each identifier's meaning, the data was pre-processed. The first step was to delete every student that didn't take all the tests – these missing data could negatively influence the results obtained. As ENEM does not have the information about each student's presence in a test, it was inferred. It is known to be nearly impossible to get a zero score in any one of the exam's multiple-choice tests [de Andrade, 2014]. Thus, we deleted every student with a zero score on at least one of the multiple-choice tests. Then, for each year, a new attribute with all the scores normalized with min-max between 0 and 1 was created.

Finally, a pre-processed file was created for each edition of ENEM containing the essential attributes such as participant scores (original and normalized), state, high school type, disabilities, and ID.In general, the selected attributes were those related to the participants' scores (Humanities, Natural Sciences, Languages, Mathematics, and Essay). It was also important to store disability indicators to determine if the student falls into any special group. Finally, we selected some socioeconomic responses from the questionnaire, such as the participant's locality, type of school, the number of people in their household, and the gross income of the household.

5.3 Resulting Dataset

The resulting dataset has 5 files, one per ENEM's edition, with only the attributes we needed. It is publicly available for download in a *Kaggle* repository created for our work, at Barbosa [2023]. Figure 3 shows the percentage of each disability

SUBJECT	ACRONYM
HUMAN SCIENCES	HUMAN. SC.
LANGUAGE, CODES	LANG. COD.
NATURAL SCIENCES	NAT. SC.
MATHEMATICS	MATH.
ESSAY	ESSAY

Table 2. List of ENEM's subject with the corresponding acronyms

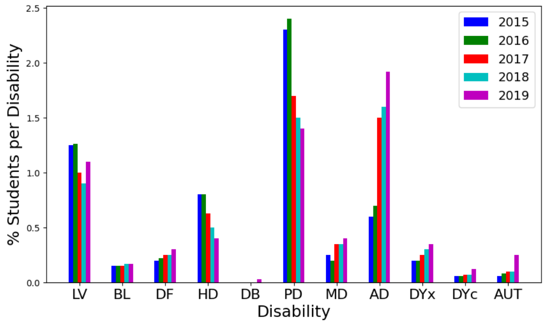


Figure 3. Percentage of participants with each type of disability per year

regarding all participants, i.e., the percentages after cleaning and pre-processing. One of the most notable aspects is the increase of participants with Attention Deficit over the years, which practically tripled since 2015. Is also noticeable the negligible presence of participants with Deafblindness. The limited number of participants with this particular disability biases any major study about their behavior and patterns.

5.4 Data Analysis

Once all the data was cleaned and pre-processed, the next steps aim to answer the questions presented in Section 2.

5.4.1 Q1: Major difficulties per disability

To understand the major difficulties per disability, we started by calculating the average score of the non-disabled group for the years between 2015 and 2019. Then, for each disability group, we calculated their corresponding average score and the percentage variation compared with the group of non-disabled participants. Then, we created the heatmaps shown in Figure 4 to express the percentage variation of the average score. In the heatmap, the closer to green, the higher the average score obtained, and the closer to red, the lower the average score obtained. An up-arrow means that the related group average score is greater than the group with No Disability (ND), while the down-arrow means that it was lower.

As shown, each disability group has its own pattern that remains nearly the same over the years. Apparently, some disabilities do not bias the grades – as for Low Vision or Physical Disability. On the other hand, participants suffering from some disabilities performed better than the non-disabled ones – the most notable case refers to Attention Deficit. Lastly, other disability groups performed worse than the non-disabled one; the most apparent cases are Deafness, Deafblindness, and Mental Disability.

Once again, we must consider the size of each disability group. For example, as shown in Figure 3, the number of Deaf-blind participants is negligible. Hence, any conclusion about tiny groups may not be statistically meaningful.

Answer: Analyzing Figure 4, we observe significant differences in performance, particularly in the Essay. Some groups score almost 30% above the non-disabled average, while others lag behind by 40%. Notably, both Human Science and Language Codes consistently produce poor results. This suggests that the main difficulty in the exam may not be linked to Math but rather to challenges within the Humanities part.

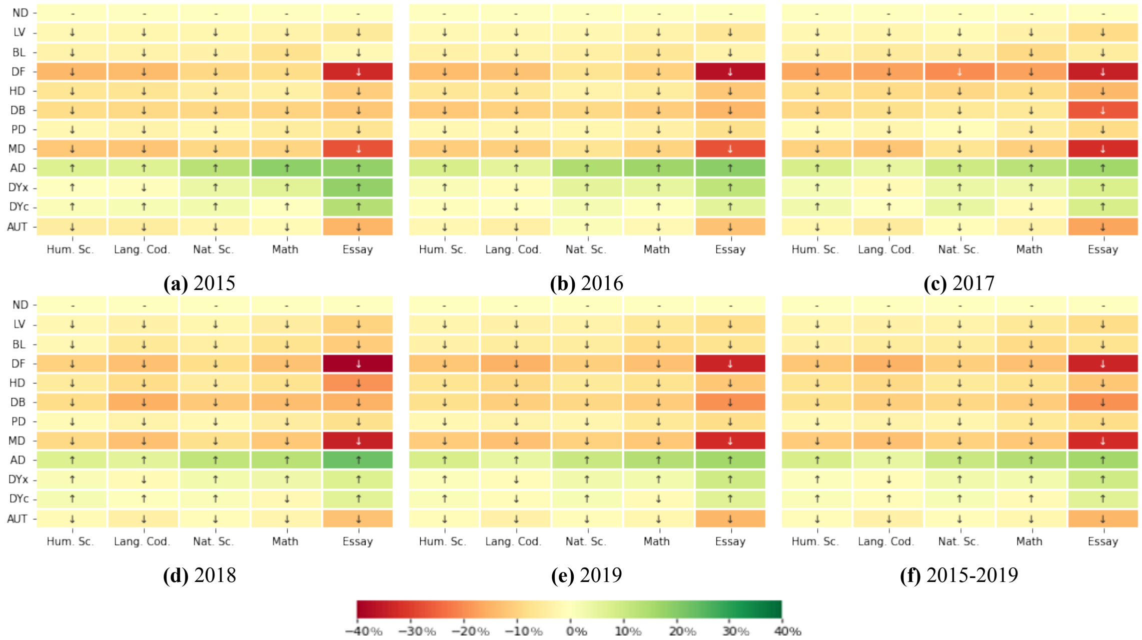
5.4.2 Q2: Performance per disability

While in Question Q1 the major goal was a generalist analysis, the purpose of Question Q2 is to study the performance distribution of one disability type at a time. To this end, we first divided all the participants into clusters using K-means. Aiming for a greater understanding of the clustering itself, we developed our own implementation of K-means. It is available at Barbosa [2021]. The initial choice of centroids was performed using the K++ logic and the distance calculations were made with the help of libraries that can process a large volume of data in parallel. The normalized scores of the participants were given to the algorithm together with the value of K. The Elbow method was used per year to identify the ideal value of K. In general, almost all the years followed a similar pattern, as shown in Figure 5, determining that the best option is to use K=3 clusters. The largest discordance occured in the year 2017, which was the only one where the ideal number of clusters wasn't 3, but, actually, 4.

Once the ideal number of clusters was found, we considered every execution of K-means that used 3 clusters. Therefore, for each year, every participant received a label from its cluster. The average grade of each cluster was calculated and the cluster with the highest average was defined as ”High performance”, the intermediate as ”Average performance” and the lowest as ”Low performance”. Hence, it was possible to analyze each disability's distribution. The current focus is to correlate each disability with its performance cluster defined by the use of K-means. To this end, the analysis consisted in computing a sum for each year and disability, where low-performance participants were accounted as 1, average performance as 2, and high-performance as 3. Then, the average value was calculated from the sum and the Heatmap in Figure 6a was created based on it. In this case, the closer to the maximum average value 3, the greener the color. The minimum value 1 is more red, and the value 2 has lighter colors. An up-arrow means that the related group is made up of more high-performing students, while the down-arrow means that it is made up of more students with low performance.

As shown, the participants generally had better performance in 2017 compared with the other years; it may indicate that Enem's 2017 edition was easier than the other editions, which also agrees with the different result of the Elbow method for this particular year. To facilitate the analysis, in Figure 6b, all values were normalized. Basically, this tends to equalize the difficulty of the exams. Specifically, we divided the average score of each disability by the average score of the ND group for their corresponding year. In this context, a greener color and, so on, an up arrow, indicates that the presence of high-performing students is more significant, percentage-wise, compared to the ND group, with the three greener disabilities being AD, DYx and DYc. A redder color and a down-arrow indicates the opposite with the three redder disabilities being DF, MD and DB.

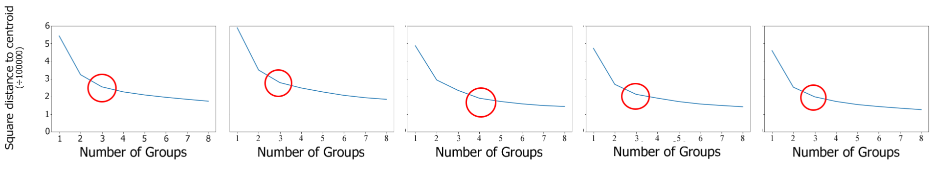
While this analysis makes it possible to relate the performance distribution between the disability groups, it does not express how each group is divided into the clusters of High, Average and Low performance. Then, we also report the percentage of each performance cluster by disability. If the disability has a small impact on the candidate's grade, it is expected that the percentage related to students with that condition be almost the same in the Low-performance cluster, Average and High. Figure 7 shows the aforementioned percentages for the most recent year, i.e., 2019. Similar results were



(a) 2015 **(b)** 2016 **(c)** 2017

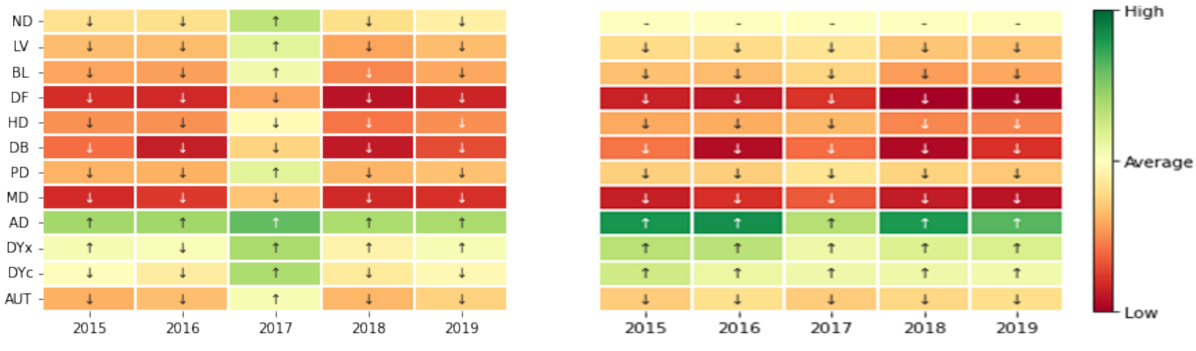
(d) 2018 **(e)** 2019 **(f)** 2015-2019

Figure 4. Average variation of grades of each disability group compared with the non-disabled group



(a) 2015 **(b)** 2016 **(c)** 2017 **(d)** 2018 **(e)** 2019

Figure 5. Average sum of squared distances to the closest centroid. An elbow in each plot reveals the ideal number of clusters; see the red circles



(a) Non-normalized **(b)** Normalized

Figure 6. Correlation between disability and performance

seen for the other years, which are not shown for brevity.

Answer: Based on Figure 6, we infer that the majority of disabilities is related to a bad performance in ENEM compared with the non disabled group. These results are complementary to those obtained for Question Q1 with no new insights being brought from them. On the other hand, while analyzing Figure 7, some disabilities have a more equal cluster distribution, such as Blindness, Dyscalculia, and Autism. Once again, the Attention Deficit group is the outlier, having its high-performance group larger than the sum of the other two.

5.4.3 Q3: Essay difficulties

At this point, we know that some disability groups express a different behavior than that of the non-disabled group. Based on information shown previously, especially in Figure 6, the disabilities having a greater impact on the grade are Attention Deficit (AD), Deafblindness (DB), Deafness (DF), Dyscalculia (DYc), Dyslexia (DYx), and Mental Disability (MD). However, when the groups' sizes shown in Figure 3 are considered, it can be inferred that DB and DYc have negligible sizes that may despair further studies. Consequently, we focus on the remaining four disability groups in the following.

For the essay, it is important to present how the scoring criteria works [MEC, 2022]. Each exam has a topic the participant needs to express and defend an opinion on. The text must follow the pattern of an argumentative essay, containing a maximum of 30 lines and respect some rules – for example, follow the Universal Declaration of Human Rights by the United Nations. The non-compliance with these points can zero the candidate's score. Also, the score is divided into the 5 competences shown in Table 3, each worth 200 points.

Then, using data of all years, we calculated the average overall essay score regarding all competences for each disability group, and for those with no disability. Figure 8a has the average scores, making it possible to compare the essay

Competence	Explanation
C1	Domain of the formal writing of the Portuguese language as accentuation, spelling, use of the hyphen, use of uppercase and lowercase letters and syllabic separation.
C2	Theme comprehension, evaluating the integrated reading and writing skills of the candidate
C3	Select, relate, organize and interpret information, facts, opinions and arguments in defense of a point of view
C4	Knowledge of the linguistic mechanisms necessary for the construction of arguments as prepositions, conjunctions, adverbs, and adverbial phrases
C5	Present an intervention proposal for the problem addressed that respects human rights, in detail and well related to the text

Syntactic
Semantic

Table 3. Explanation about each competence considered in the essay

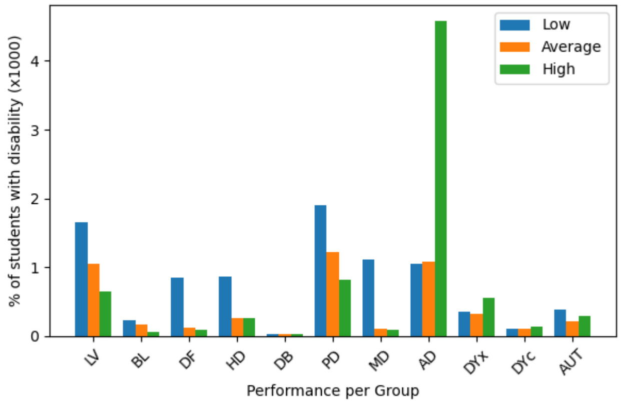
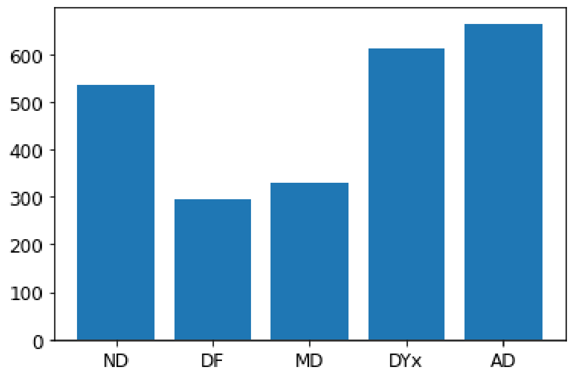
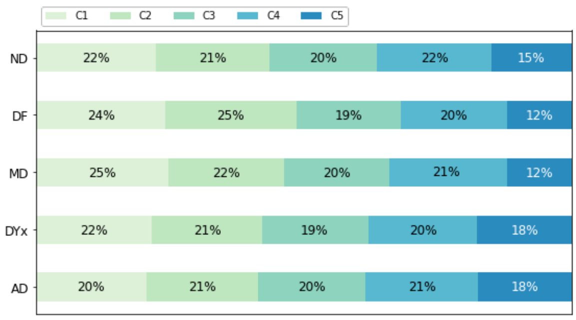


Figure 7. Percentage per performance cluster by disability for the year 2019

performance. To analyze the impact of each competence, we also computed the percentage distribution per competence within the score of each group, which is shown in Figure 8b.



(a) Average overall score regarding all competences



(b) Percentage of the average overall score per competence

Figure 8. Essay scores per disability group

Answer: Analyzing the Figure 8a it is possible to see that the average scores per disability vary considerably. While Dyslexia (DYx) and Attention Deficit (AD) have better scores than the group with No Disability (ND), Deafness (DF) and Mental Disability (MD) are under the general result. It is, actually, a non-obvious performance. At least, one could assume that the deaf group would have a score similar or even higher than that of the others, because reading is one of the most essentials ways to connect with society for those who cannot listen. Moreover, based on Figure 8b, it is notable that the division of the score by competence is similar in AD and DYx, having a better conclusion (competence C5) of the essay than the others. For DF and MD, while they have a greater understanding of the theme (C2), apparently, they have difficulties in designing an efficient intervention. As a case study, we also analyzed the 2017 essay posteriorly. The theme was "Challenges for the educational formation of the deaf in Brazil". Analyzing the performance of the deaf group on this particular theme sounded interesting, and closely related to our goals. However, the results of year 2017 do not change significantly compared to those of other years. Despite an increase in the DF group average score, it is almost negligible and, thus, we omit these results for brevity.

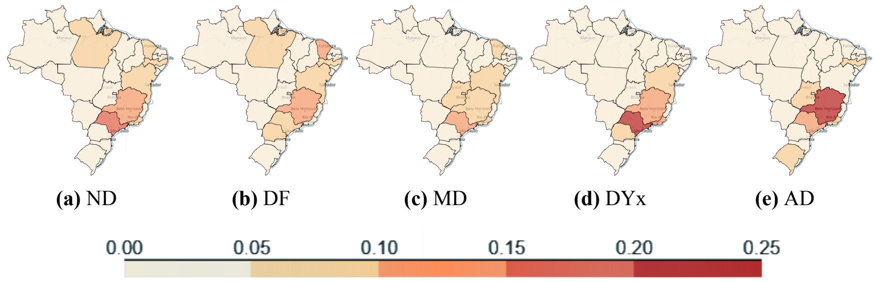
5.4.4 Q4: External influence

To determine the influence of external factors, we analyzed the participants' answers to the questionnaire they filled out before taking the exam. Based on it, we could correlate the grade with the location, economic situation, and school type. Note that the data of all the years were condensed together in this process. It means that this analysis does not rely on the year of the exam, but solely on the participants' information.

Location: Initially, the size of each disability group was calculated and the participants were grouped by state of residence. Then, a choropleth graph was created and shown in Figure 9. It presents the percentage of participants coming from each state of Brazil. Provided that the Figure 9a express the general distribution of the population, it can be used as a base to understand different patterns of distribution for each disability. It is interesting to point out again AD and DYx as both have a state responsible for more than 20% of its group size – respectively, Minas Gerais and São Paulo.

Economic condition: The first step was to calculate the average income of the students' families. While Question 5 of the ENEM's questionnaire regards the total money incoming in the participant's family, Question 6 expresses how many people live in his/her house. The incoming salary was presented as categorical groups, named from A to Q, each one having its own range of values which were also increased to reflect inflation over the years. As the same group is supposed to represent the same social class all the years, the incoming salaries were normalized based on the last year's values. This means that the income values of all the older years were adjusted to represent the same as those of 2019. Then, we could finally calculate the average income.

This result was later used to evaluate the economic difference between each disability group and the non-disabled one. Figure 10 reports that the groups with the largest aver-



(a) ND **(b)** DF **(c)** MD **(d)** DYx **(e)** AD

Figure 9. Percentage of participants with each type of disability grouped by the state of origin

age incomes are also the ones with the largest grades — in this case, DYx and AD. Both DF and MD present similar behavior, and an average income of nearly R$ 1.000, which is very close to that of the no disability group. Moreover, the average income was disposed according to the state, aiming for a broader correlation with the students' origins. Figure 11 presents the average income of each disability group by state.

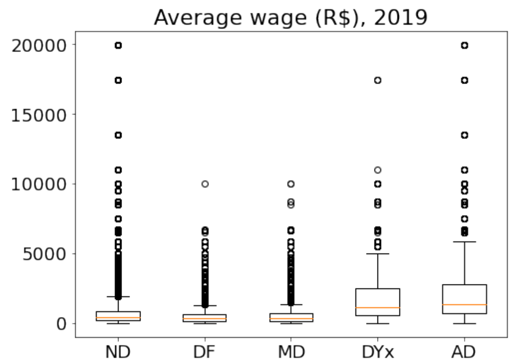


Figure 10. Boxplot of incoming salary grouped by disability

School Type: As shown in previous studies, such as the one of Lima *et al*. [2020], participants from private schools often have a better performance than those of public schools. With that in mind, we analyzed the percentage of each disability group in each type of school aimed to verify this hypothesis and how it can be related to the disability group performance.

There are four possible answers to the school type question: *1 - Prefer to not answer*, *2 - Public*, *3 - Private*, and *4 - Abroad*. Figure 12a reports the percentage of students per school type and disability type, where the disabilities and school types are given along the X and Y axes, respectively.

Note that the majority prefer not to answer and the studying-abroad group is negligible, regardless of disability. For a clearer distinction between private and public schools, we show in Figure 12b the results regarding only these two school types. It then becomes more obvious the difference in the distribution between the groups; the majority of participants with Dyslexia (DYx) or Attention Deficit (AD) come from private schools, and the opposite occurs for the other two disability groups, and also for the non-disabled one.

We also conducted an analysis exclusively on students from private schools, recognizing that in this environment students with certain disabilities are more likely to be diagnosed. For each disability, we calculated the average score for each of the five subjects in the exam taken by private school students. Subsequently, we summed these averages to obtain a single "overall average" value. Next, we performed the same process for candidates without disabilities from private schools and calculated the ratio between the scores of those with disabilities and those without. The results are as follows, with the first value being the acronym and the second value representing the division of scores: (LV, 0.89), (BL, 0.80), (DF, 0.56), (HD, 0.80), (DB, 0.70), (PD, 0.85), (MD, 0.53), (AD, 0.97), (DYx, 0.89), (DYc, 0.82), (AUT, 0.76). As shown, none of the average scores for candidates with disabilities surpass that of non-disabled candidates.

Finally, we created Figure 13 to study the correlation between students from public and private schools according to their state of origin. Note that the prefer-not-to-answer and the studying-abroad groups are not considered in these maps. **Answer:** Figure 9 indicates that participants with AD and DYx are more concentrated in the Southeast of Brazil than the remaining participants. Alongside of the fact that students from this region often perform better than others [Lima *et al.*, 2020], the location may, therefore, have biased the good results of the participants with these two disability types. Based on the economic conditions shown in Figure 10, it is also obvious the difference in average incomes. Again, both DYx and AD's households have a much higher income than the others. Finally, for the school type, as shown in Figure 12b, it is clear that participants with DF and MD follow a pattern similar to those without disability, which is distinct for the groups DYx and AD that come mainly from private schools. However, once we considered only the private schools, both AD and DYx participants performed below average. All these indicators corroborate the understanding that the superior performance of participants with AD and DYx may have been influenced by several issues, such as economic conditions, school type, or even failure to declare a disability when completing the questionnaire due to free will or lack of diagnosis, and not necessarily by their disability.

5.4.5 Q5: Statistical Significance

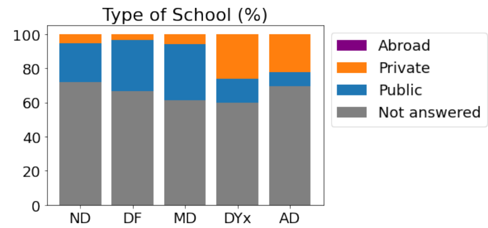
We also conducted an analysis to ascertain the statistical significance of exam score differences between candidates with and without disability. These results are summarized in the Table 4. Despite the potential for exploring various hypotheses, we focused exclusively on exam scores to better understand the specific impact of disabilities on academic performance. Using the well-known T-student test with $\alpha = 0.05$, we found compelling evidence to reject the null hypothesis that the grades are similar, thus indicating a significant difference in performance between candidates with and without disabilities. These results also support the understanding that each disability has a different impact on student learning.

6 Conclusion

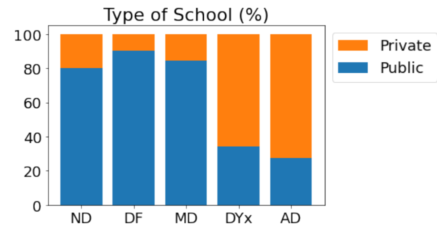
This article presented an analysis of students with disabilities who attended ENEM from 2015 to 2019. The results ob-

(a) ND **(b)** DF **(c)** MD **(d)** DYx **(e)** AD

Figure 11. Average income in Reais (R$) regarding each type of disability and the state of origin

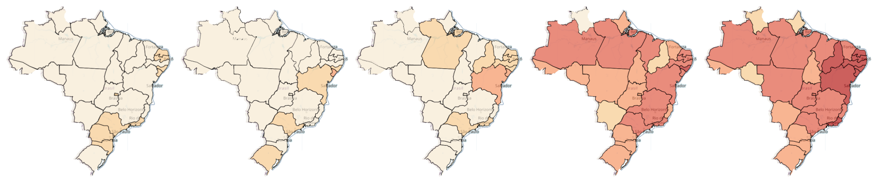


(a) All school types



(b) Only private and public

Figure 12. Type of school by disability



(a) ND **(b)** DF **(c)** MD **(d)** DYx **(e)** AD

Figure 13. Percentage of students from private schools by state and disability

Table 4. T-student test results comparing the grades of students with and without disabilities. As expected, the null hyphotesis that the grades are similar is rejected for most cases.

	Human Sc.		Lang. Cod.		Nat. Sc.		Math		Essay	
	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value
LV	30.33	7.77e-199	40.96	0.0	33.25	1.02e-237	44.60	0.0	44.87	0.0
BL	15.89	6.71e-55	24.17	5.59e-119	22.89	8.21e-108	37.09	4.23e-252	12.13	3.62e-33
DF	101.62	0.0	121.32	0.0	98.94	0.0	90.97	0.0	82.28	0.0
HD	64.62	0.0	76.51	0.0	54.92	0.0	49.92	0.0	47.41	0.0
DB	7.59	5.62e-11	8.80	2.56e-11	8.75	3.17e-13	8.04	7.55e-12	4.40	3.35e-5
PD	34.25	8.25e-254	55.79	0.0	43.39	0.0	68.10	0.0	64.12	0.0
MD	77.04	0.0	103.67	0.0	61.10	0.0	73.02	0.0	79.39	0.0
AD	-92.64	0.0	-73.86	0.0	-126.08	0.0	-110.60	0.0	-97.98	0.0
DYx	-6.49	8.83e-11	6.64	3.31e-11	-21.13	1.18e-95	-16.85	2.54e-62	-27.55	1.92e-157
DYc	-2.88	3.95e-3	-0.766	0.44	-5.92	4.11e-9	1.70	0.08	-8.78	4.49e-18
AUT	7.59	4.15e-14	16.38	1.55e-57	1.62	0.10	6.23	5.34e-10	20.69	2.85e-88

tained express a singular behavior. The majority of students in the disability groups had a score pattern similar to those in the ND group. However, Attention Deficit and Dyslexia groups had a higher-than-average performance. Once the socioeconomic conditions of these students were analyzed, we noted that their average income is considerably above the others. Also, they are more concentrated in the Southeast region of Brazil, where previous work [Lima *et al.*, 2020] observed a performance better than others in the ENEM. Finally, most students with those disabilities come from private schools. Distinctly, students with Deafness and Mental disabilities had much worse scores, despite of having a socioeconomic pattern similar to the non-disabled group. It may indicate a poor inclusion of individuals with these conditions.

These results suggest that the differences in scores are not only based on the disabilities. Dyslexia and Attention Deficit do not have obvious diagnoses because targeted patient follow-up is required. The difficulty of noting the condition and the expenses of the diagnosis can corroborate for these students to be of upper social classes – which may explain the better grades and school origin. Under this logic, students with Deafness or Mental Disability can be more easily diagnosed, which explains their less "elitist" conditions.

Once again, it is crucial to emphasize that we rely solely on the candidate's response to the exam questionnaire, so there may be students with unidentified disabilities who have been categorized under the "no disability" group in our study. This dynamic encompasses the challenges of not being able to thoroughly study every disability group and also considers the socio-economic relationship of individuals with the diagnosis of their disabilities. In this regard, there are specific situations that we cannot address due to a lack of information from the questionnaire. For instance, it would be valuable to study the performance of candidates with Attention Deficit Disorder who use medication for focus, such as Ritalin. This is only our interpretation of the results obtained during the work and it cannot be guaranteed by the INEP's data. Also, the small size of some disability groups may not ensure statistical significance of the results. Nevertheless, we believe our work can serve as a basis to the elaboration of public interventions, aiming to make access to education, and its quality, more equal. Initially, it is important to encourage the professionals – especially those from primary schools – to refer students with low grades or lapses in class to a psychologist to previously determine potential disabilities. With the school's support, an early diagnosis can facilitate a more targeted follow-up for these students.

In addition, more use of sign language would be beneficial. Actually, LIBRAS – Brazilian Hands Signal – is the second official language in Brazil. However, it is not taught in school and interpreters are rare. The facilitation of learning LIBRAS would be beneficial to society and could have a very positive impact on the performance of the deaf group.

As future work, further analyses could be performed regarding the context of ENEM. The exam will pass by reform and a new test model will be applied in 2024. The major difference is that the test will have discursive questions related to an area of knowledge chosen by the student [Politize, 2022]. Once the data from these future editions become available, it will be interesting to understand which specific area each disability group prefers. Furthermore, as the exam will count with more textual development – due to discursive questions – an analysis of the groups with bad/good performance in the essay can give insights into how the adaptation process of groups with disabilities will be in this new model.

Finally, it is worth noting that our work does not cover the most recent editions of the exam, as the data corresponding to the tests after 2019 were not yet available at the time. Hence, a future work could be done to validate if the performance of the students with disability continued following the same patterns observed in our article. Furthermore, a study can be carried out to understand how the Covid pandemic and distance learning impacted these groups' behavior in the exam.

References

Ahmed, M., Seraj, R., and Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8). DOI: 10.3390/electronics9081295.

Arthur, D. and Vassilvitskii, S. (2007). k-means++: The Advantages of Careful Seeding. https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf.

Barbosa, J. A. F. (2021). ENEM ANALYSIS. https://github.com/joaoaugusto05/ENEM-ANALYSIS.

Barbosa, J. A. F. (2023). ENEM DATASET 2015 - 2019. https://www.kaggle.com/datasets/joaoaugusto05/enem-dataset-2015-2019.

Brazileducation (2023). ENEM in Brazil. https://www.brazileducation.info/tests/higher-education-tests/enem-in-brazil.html.

Dados, M. (2021). Entenda o Algoritmo K-means e Saiba como Aplicar essa Técnica.

de Andrade, D. F. (2014). Você nunca vai tirar zero na prova do Enem, saiba por quê. https://educacao.uol.com.br/noticias/2014/10/31/voce-nunca-vai-tirar-zero-na-prova-do-enem-saiba-por-que.htm.

de Oliveira, C. G., Barwaldt, R., and Lucca, G. (2020). Análise do desempenho de pessoas com deficiência que prestaram o exame nacional do ensino médio - enem. *Revista de Educação, Ciência e Tecnologia*, 9(1). DOI: 10.35819/tear.v9.n1.a4038.

de Souza, T. O. (2021). Análise de dados: Um estudo do perfil dos participantes do ENEM 2019. https://repositorio.ufersa.edu.br/handle/prefix/6916.

Faria, R. S. (2013). https://twitter.com/RomarioOnze/status/376852643244937216?s=20.

INEP (2022). Microdados do ENEM. https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem.

Julião, A. (2021). Faltam pesquisas para embasar as políticas educacionais no Brasil, afirmam especialistas. https://agencia.fapesp.br/faltam-pesquisas-para-embasar-as-politicas-educacionais-no-brasil-afirmam-especialistas/37118/.

Leria, L. d. A., Alves, G. d. S., Benitez, P., and Silva, F. J. F. d. (2023). Perspectivas de pessoas com deficiência visual sobre a acessibilidade no exame nacional do ensino médio (Enem). *Revista Educação Especial*, 36(1):e57/1–31. DOI: 10.5902/1984686X84416.

Lima, A. M. S., Florez, A. Y. C., Lescano, A. I. A., de Oliveira Novaes, J. V., de Fatima Martins, N., Junior, C. T., de Sousa, E. P. M., Junior, J. F. R., and Cordeiro, R. L. F. (2020). Analysis of ENEM's attendants between 2012 and 2017 using a clustering approach.

MEC (2022). Critérios da redação. http://portal.mec.gov.br/ultimas-noticias/418-enem-946573306/81381-conheca-as-cinco-competencias-cobradas-na-redacao-do-enem.

O'Neill, L., Markward, M. J., and French, J. P. (2012). Predictors of graduation among college students with disabilities. *The Journal of Postsecondary Education and Disability*, 25:21–36.

Politize (2022). Novo Enem: quais são as novidades para 2024? https://www.politize.com.br/novo-enem/.

Polyzou, A. and Karypis, G. (2019). Feature extraction for next-term prediction of poor student performance. *IEEE Transactions on Learning Technologies*, 12(2):237–248. DOI: 10.1109/TLT.2019.2913358.

Ragosta, M. and Wendler, C. (1992). Eligibility issues and comparable time limits for disabled and nondisabled sat® examinees. *ETS Research Report Series*, 1992(1):i–34. DOI: https://doi.org/10.1002/j.2333-8504.1992.tb01466.x.

Simon, A. and Cazella, S. (2017). Mineração de Dados Educacionais nos Resultados do ENEM de 2015. *Congresso Brasileiro de Informática na Educação*, page 754–763.

Stearns, B., Rangel, F., Firmino, F., Rangel, F., and Oliveira, J. (2017). Prevendo desempenho dos candidatos do enem através de dados socioeconômicos. In *XXXVI Concurso de Trabalhos de Iniciação Científica da SBC*. SBC.

Terra, K. (2019). Seleção de características para predição de notas no ENEM 2018. https://github.com/kterra/enem-project-impa.

van Ewijk, R. and Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, 5(2):134–150. DOI: https://doi.org/10.1016/j.edurev.2010.02.001.

Viggiano, E. and Mattos, C. (2013). O desempenho de estudantes no Enem 2010 em diferentes regiões brasileiras. *Revista Bras. de Estudos Pedagógicos*, 94(237):417–438.

Zhang, D., Katsiyannis, A., and Kortering, L. J. (2007). Performance on exit exams by students with disabilities: A four-year analysis. *Career Development for Exceptional Individuals*, 30(1):48–57. DOI: 10.1177/08857288070300010601.