Analysis of ENEM Data in an Educational Context: An English Literature Review

Bernardo José Bueno[a]

[a] Federal University of Techonology - Paraná, Toledo, Brasil, bernardobueno@alunos.utfpr.edu.br.

Abstract. In recent years, data analysis has emerged as a critical tool for shaping public policies, offering a more comprehensive understanding of reality and uncovering patterns that may elude human perception. Within this context, Large-Scale Assessments (LSAs) stand out as invaluable sources of high-quality data, holding tremendous potential to inform and guide decision-makers. Among these LSAs, the National High School Exam (ENEM) administered in Brazil plays a pivotal role. With over 3 million candidates participating each year, the ENEM serves as a prerequisite for admission to the majority of the country's top public universities. This extensive examination encompasses 180 questions across diverse subject areas, including an essay component. Additionally, candidates are given the option to complete a socio-economic and cultural questionnaire. Importantly, the data of all participants are made publicly available every year while preserving their anonymity. This rich dataset empowers researchers to uncover performance patterns and meticulously scrutinize the influence of multifaceted factors, including socio-economic variables, cultural elements, teaching methodologies, school management practices, and geographical disparities on students' scores. This, in turn, allows for a profound understanding of critical national issues, ranging from the effectiveness of public policies to the far-reaching consequences of socio-economic inequalities. While a significant body of literature in Portuguese has been dedicated to exploring these themes, an observable dearth of articles published in English prevails within this realm, posing an obstacle to international access to essential information concerning one of the world's most extensive educational systems. This study endeavors to indentify the size of this scientific gap and shed light on the predominant themes addressed in English-language publications.

Keywords. Large-Scale Assessments, Education, Brazil, literature review, High School National Exam.

1

2 Introduction

The use of data analysis to shape public policies has witnessed significant growth in recent years, as demonstrated by [1]. Utilizing such data sources allows for a more comprehensive analysis of reality and the identification of patterns that may escape human perception.

Within this context, Large-Scale Assessments (LSAs) provide a substantial volume of high-quality data and an tremendous potential to provide valuable insights to decision makers [1].

The National High School Exam (ENEM) is one such LSA administered in Brazil, involving more than 3 million candidates each year it is required to admission in the majority of the country's top public universities[2].

The exam consists of 180 questions divided among Humanities and its Technologies, Natural Sciences and its Technologies, Languages, Codes and its Technologies, Mathematics and its Technologies, and an essay. In addition, candidates have the option to complete an optional socio-economic and cultural questionnaire. And finally, every year, the data of all participants, are made publicly while safeguarding their identities[2].

In this manner, we can employ this data to discern patterns of performance, scrutinizing the influence of various factors, including socioeconomic, cultural, teaching methodologies, school management, geographical, among others, on students' scores. This enables a deeper comprehension of a range of critical topics for the country, such as the efficacy of public policies and the consequences of inequality. Thus, an extensive body of literature in Portuguese delve into

these topics[3].

However, there is an observed dearth of articles published in English within this research domain. This gap hampers the international community's access to crucial information regarding one of the largest educational systems in the world, which could facilitate the formulation of theses on the education system, socioeconomic impacts on access to higher education and the enhancement of other systems.

This study aims to ascertain the extent of this academic gap and to identify the predominant themes addressed in English-language publications.

3 Methodology

The research involved the main words related to the theme in varied combinations, the keywords were "ENEM", "education", "analysis", "Data Science", "Data Mining", "Brazil", "LSA". The search engine used was ScienceDirect due to its renowned reliability and the quality of the associated databases.

4 Results and discussion

4.1 Research results

In the course of the research, hundreds of results were obtained using the selected terms, however. Subsequent examination of their titles led to the identification of 46 articles that appeared potentially relevant to the research scope. Further in-depth content analysis revealed that only 4 of these articles met the criteria for alignment with the research theme.

These researches were published in respected journals as illustrated in **Tab 1.**

Tab. 1 - Selected articles and local of publication.

Title	References	Journal
Through the looking glass: Can classroom observation and coaching improve teacher performance in Brazil?	[4]	Economics of Education Review
A data mining framework for reporting trends in the predictive contribution of factors related to	[5]	Expert Systems with applications
educational achievement		
Greenness around Brazilian schools may improve students' math performance but not science performance	[6]	Urban Forestry & Urban Greening
Centralized college admissions and student composition	[7]	Economics of Education Review

4.2 Selected Articles Themes

It is noteworthy that this number of results is considerably lower compared to the number of articles found in reviews written in the Portuguese language. However, the research theme does exhibit a certain similarity, with statistical analyses and data mining focused on student performance [3], as illustrated in **Tab. 2.**

Tab. 2 - Selected articles and adressed themes.

Theme	References	Year of publication
Influence of School Teachers' Training Methods	[4]	2018
Influence of socio-economic and education related factors	[5]	2023
Influence of school environmental conditions	[6]	2022
Effects from the adoption of the centralized admission system	[7]	2021

4.3 Description of the articles

In a complete analysis, each article provides the

following information.

[4] Conduct a randomized assessment in the efficay of a expert coaching program for teachers based on Teach Like A Champion, a resource by Doulgas Lemov focused on pratical strategies for enhancing teacher effectiveness and student engagement. The program lead to statistically significant academic improvement in schools where it was implemented. The improvement was visualized in Enem data from those schools students.

[5] The authors introduce a new framework for analyzing large-scale assessment (LSA) data using data mining and machine learning techniques. It aims to understand the driving factors behind ML models predicting LSA outcomes, offering insights for educators and policymakers. The proposed score-based metrics are designed to address limitations of existing scoring methods, providing a new approach to quantify the impact of variables, including highly correlated ones like socioeconomics.

The research validates these metrics using both synthetic and real data, applied to Brazilian LSA data (ENEM) from 2009 to 2019. Key findings include confirming the strong influence of known determinants of school performance, such as income, race, and parental education. Additionally, insights into school infrastructure and faculty-related factors, like students' computer access and faculty workload, present opportunities for further investigations into educational policy.

[6] The study explore the relationship between green spaces and school-level academic performance in Brazil, considering the vital role of green areas in people's well-being. Using mixed-effects regression models and adjusting for air pollution, socioeconomic status, and spatiotemporal factors, the Normalized Difference Vegetation Index (NDVI) is employed as the exposure variable. Considering ENEM results, it is revealed a positive association between green surroundings and math performance, though no significant impact on science. These findings have implications for policymakers and urban planners, highlighting the potential to enhance school environments with green spaces to foster public health and improve academic outcomes.

[7] The study investigates the effects of transitioning to centralized admission systems in the education market, a shift that's gaining traction. Focusing on the higher education sector in Brazil and utilizing comprehensive administrative data, the staggered implementation of a centralized clearinghouse across institutions is analysed to assess its impact on student demographics.

The results suggest that centralization reduces application hurdles and heightens competition, leading to a decrease in the proportion of female students and an increase in the average student age. Furthermore, institutions adopting centralized admission tend to attract students from different locations with higher test scores. These findings also hint at centralization exacerbating institutional stratification by quality, widening the divide between low and high-quality institutions. The Enem data is vital for the analysis and after the adoption of the Unified Selection System (SISU), it is noticeable an increase in competitiveness among participants.

5 Conclusions

The predominant use of the English language in international publications makes essential for researchers to publish in this language.

Several advantages are offered, including broader global reach, access to the international research community, greater citations and impact, standardization and clarity, editorial and peer review quality, career advancement, compatibility with international databases, and facilitation of innovation and knowledge transfer [8].

Unfortunately, despite a significant number of articles published in Portuguese on the subject under analysis, there is a notably limited quantity of works published in English.

This constitutes a missed opportunity regarding one of the world's largest and most diverse educational databases, which could be better harnessed to enhance education and reduce inequality in Brazil and numerous other countries.

It is essential that more seminal works from this field be translated and published.

6 References

[1] Ertl, B, Hartmann, F, Heine, JH. "Analyzing Large-Scale Studies: Benefits and Challenges". Frontiers in Psychology 2020; 11.

[2] Travitzki R. *ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar* [Doctoral Thesis]. São Paulo: , Faculdade de Educação; 2013 [cited 2023-09-30]. doi:10.11606/T.48.2013.tde-28062013-162014.

[3] Lima P da SN, Ambrósio APL, Ferreira DJ, Brancher JD. Análise de dados do Enade e Enem: uma revisão sistemática da literatura. *Avaliação (campinas).* 2019;24:89–107. https://doi.org/10.1590/S1414-40772019000100006.

[4] Bruns B., Costa L., Cunha N. "Through the looking glass: Can classroom observation and coaching improve teacher performance in Brazil?". *Economics of Education Review 2018*; 64:214-250.

[5] Silva Filho R. L. C., Brito K.,Adeodato P. J. L. "A data mining framework for reporting trends in the

predictive contribution of factors related to educational achievement". *Expert Systems with Applications* 2023; 221:119729.

[6] Requia W. J., Saenger C. C., Rejane Cicerelli R. E., Abreu L. M. de, Cruvinel V. R. N.. "Greenness around Brazilian schools may improve students' math performance but not science performance". *Urban Forestry & Urban Greening* 2022; 78:127768.

[7] Machado C., Szerman C. "Centralized college admissions and student composition". *Economics of Education Review 2021*; 85:102184.

[8] Hamel R. "The dominance of English in the international scientific periodical literature and the future of language use in science". AILA Review 2007; 20:53-71.