This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Science Education on April 202, available online: http://dx.doi.org/10.1080/09500693.2025.2488409.

Science and Engineering for What? A Large-Scale Analysis of Theme Selection in K-12 Science and Engineering Fair Projects

Science and Engineering Fairs (SEFs) play a pivotal role in K-12 education, enabling students to engage in science and engineering practices. Particularly, students are given the chance to experience authentic and open inquiry processes, by defining which themes, questions and approaches will guide their scientific endeavors. This study analyses over 5,000 projects presented at FEBRACE, a national science fair in Brazil, over a span of 20 years, employing topic modelling to identify and examine the predominant themes guiding students' investigations and project designs. The findings indicate a diverse range of topics explored, significantly varying in response to changes in time, region, and educational settings. We argue our findings illustrate the authentic and open aspects of students' inquiries and provide valuable evidence supporting the role of science and engineering fairs as spaces for culturally responsive science learning. Furthermore, we propose that the methodology employed in this study can serve as a valuable tool for similar analyses in other educational contexts, encouraging further exploration of theme selection and project design across various settings. Finally, we argue our findings argue for the importance of equipping both students and teachers with resources to engage a more diverse range of participants and support the varied interests that students bring to their scientific and engineering endeavors, both within and beyond the context of SEFs.

Keywords: K-12, inquiry-based learning, science and engineering fairs.

Introduction

Science and Engineering Fairs (SEFs) are distinguished events within K-12 education that provide students with opportunities to showcase their projects and receive valuable feedback. This process is well-documented for both its benefits and challenges in supporting STEM learning (e.g., Bencze & Bowen, 2009; Grinell et al., 2020; Koomen et al., 2018; Schmidt and Kelter; 2017). SEFs are not merely exhibitions but represent the culmination of an authentic inquiry process, enabling students to deeply engage in science and engineering practices (NGSS, 2013) as well as competencies for scientific literacy (PISA, 2018). Previous research has demonstrated that students actively participate in scientific practices, including experiment design, data collection and analysis, and model development, through their participation in school science fairs (e.g., Koomen et al., 2018). Such involvement often leads to a deeper understanding of scientific concepts (Grinnell et al., 2020) and an increased interest in STEM careers (Schmidt & Kelter, 2017).

Contemporary literature suggests that SEFs can serve as venues for authentic participation in "doing science," where participants determine the questions that will drive their investigations (Chen et al., 2011). This characteristic enables SEFs to be seen as suitable environments for facilitating an open inquiry process (Zion & Mendelovici, 2012), where students select a theme or research topic that incorporates their personal backgrounds and interests into the scientific learning process. The ability of students to define research questions within SEFs is identified as a promising area for further academic contributions (LaBlanca, 2008), underscoring the importance of additional studies to explore how this autonomy is actualized in practice.

Analyzing data from 5,000 projects presented at FEBRACE, a national science and engineering fair in Brazil over two decades, this study employed topic modelling, a machine learning technique, to examine the central themes guiding student investigations and project designs. This study builds on preliminary findings from an analysis conducted by the authors (Eloy et al., 2023) to answer the following research questions:

(1) What are the main topics that students participating in FEBRACE choose to explore in their projects?

(2) To what extent do the topics chosen by students at FEBRACE vary across different geographical, educational, and temporal factors?

By addressing these questions, we aim to reveal the ways in which student interests shape their projects and how they adapt to their specific contexts. Moreover, the findings of this study have the potential to offer valuable insights for educators and curriculum designers, suggesting which resources might be most effective in supporting students engaged in open inquiry processes. This paper is organized as follows: "Related Work" reviews the role of SEFs in STEM Education; "Methods" details our methodological approach, with an emphasis on topic modelling; "Findings" presents the project themes, temporal trends, and regional variations; "Discussion" explores the implications of our findings for STEM education practice and policy; "Conclusion" summarizes our contributions and suggests directions for future research.

Related work

Our work builds on three distinct bodies of research. First, we draw from prior research that has outlined a diverse range of learning outcomes for students who participated in science fairs at various levels. In addition, we specifically consider science fairs as valuable learning environments where students engage in authentic and open inquiry. This entails pursuing open-ended investigations that are driven by their own interest-driven problems. Lastly, from a methodological perspective, we employ topic modelling. This promising technique is utilized for analyzing large collections of documents to identify emerging topics, serving as indicators of students' interests in their science and engineering projects.

Learning outcomes from science fairs

Previous scholars have reported different benefits from students engaging with science fairs at different levels. Koomen et al. (2018) found evidence of students engaging with NGSS science and engineering practices while developing projects for a school science fair, as well as the important role of teachers as mentors in that setting. Paul et al. (2016) also reported positive outcomes related to learning of experimentation through science fairs, based on interviews with 57 students who engaged with science fairs at a national level. More specifically, they pointed out that using well established methodologies and self-reflection by communicating with others are important aspects of science fairs that lead to science learning. Similarly, Schmidt and Kelter (2017) led focus groups discussions with students who participated in a city-level science fair, to investigate if and how science fair participation leads to increased interest and enjoyment in science. The authors identified that most students increased their understanding of science inquiry and demonstrated a positive attitude toward STEM courses and career, associated with their experience with the fair. Miller et al. (2018) used data from a large-scale sample of university students to investigate the role of science fairs and robotics competitions in their career interest in STEM. The results suggested student participation in those events relates to a 5% greater likelihood of STEM career interest at the end of high school; in addition, science fair participation was found to be a predictor of interest in a career in science/mathematics, but not in engineering or computer science.

In addition, prior research has reported both positive aspects and concerns with students engaging with science fairs. For instance, Bencze and Bowen (2009) analyzed data from four successive editions of a national-level science fair in Canada and identified students demonstrating learning outcomes in both doing and communicating science. At the same time, they raised issues of access, image and recruitment associated with the fair. Likewise, Grinnell et al. (2020) analyzed survey data from over 300 high-school students participating in science fairs across different states in the United States, to identify strengths and weaknesses of this approach for the learning and engagement with science. They identified increased interest in science as one of the most important potential positive outcomes of science fairs, but also reported students' focus on learning or winning awards were influenced by the nature of the science fair, i.e. if it was organized with a competitive or noncompetitive format. These and other studies on science fairs at the K-12 level inform our work by supporting the relevance and multiple benefits of this approach in general STEM Education, keeping in mind concerns and limitations of any findings, especially that SEFs at a national level, which is our case, do not fully represent the diverse range of contexts and students' interests.

Authentic and open inquiry in science fairs

Our work underscores the significant advantages of authentic and open inquiry processes in science learning. Authentic scientific inquiry has been defined as student engagement in scientific and engineering practices that mirror those of professional scientific communities, fostering both commitment and ownership of their projects. This type of inquiry has been documented in various studies; for instance, Koomen et al. (2018) observed students designing experiments, collecting and analyzing data, and developing models as part of their school science fair projects. Ramnarain (2020) further noted that a key benefit of science fairs is the autonomy students have in choosing their topics and designing their investigations. Additionally, Dionne et al. (2012) reported that a primary motivator for high-school students participating in national science fairs was the opportunity to explore topics of personal interest.

Scientific inquiry is often categorized by its level of openness, ranging from structured to open inquiry (Bell et al., 2005). While confirmation and structured inquiry are typically associated with a teacher-centered approach, open inquiry is seen to reflect the work performed by scientists, requiring ownership and higher-order thinking skills from students (Zion & Mendelovici, 2012). Previous scholars have highlighted science fair activities to demonstrate open inquiry (e.g. Bell et al., 2005), where students independently determine themes, problems, questions, and approaches through a continuous decision-making process. For instance, Chen et al. (2011) investigated the role of argumentation on supporting students to reflect on the validity and consistency of their open-ended scientific inquiries for a science fair. Similarly, LaBanca (2008) identified open inquiry as beneficial to juxtapose creative and logical perspectives in the problem finding process for projects to be developed for state and national science fairs.

A crucial component of inquiry-based learning at science fairs is the definition of research problems. Open inquiry approaches particularly emphasize the formulation of research questions, a practice central to scientific methodology (NRC, 2013). This allows students to integrate their personal interests into their research, which enhances engagement and learning. Adler et al. (2018) described the importance of appropriate support from teachers so that students can engage in open inquiry processes. Additionally, Lee & Cho (2017) found that the degree of structure of a problem situation influences problem-finding performance, with fifth-grade students finding more original and elaborate problems in ill-structured problem situations, and being positively influenced by students' motivation and scientific knowledge around a specific topic.

However, as LaBlanca (2008) pointed out, "there appears to be almost non-existent published research of open inquiry, in terms of science fairs, and problem finding" (pg.

2). We argue that, by identifying the main topics on students' projects for a science fair and analyzing the influence of different contextual factors on them, our research contributes to the understanding of how students identify problems that are relevant to their specific contexts to guide their scientific investigations and engineering designs.

Topic modeling in educational research

The growth in the availability of large volumes of data led to advancements in data analysis techniques based on unsupervised machine learning and natural language processing. These aim at reducing dimensionality, to discover patterns and relationships within the data, and provide insights that may be useful for future decision-making. Among these techniques, topic modeling stands out; it is a category of statistical models that consists in extracting hidden topical patterns within a collection of documents (Egger & Yu, 2022). In simple terms, topic modeling helps identify topics by analyzing the co-occurrence of similar words within a set of texts. For example, in a collection of texts, the relationship between the words "school", "student", and "teacher" could be found in the construction of a topic "education". In social sciences, topic modeling is often used to analyze textual data, helping researchers identify main themes, categorize texts, track changes over time, and extract insights from large sets of documents (Boyd-Graber, Hu & Mimno, 2017).

Topic modeling has been widely applied in the context of educational research. Ming & Ming (2013) employed classical methods to relate topics students discussed in online forums to their final grades, finding that students with low grades ignored important topics. Similarly, other works have analyzed online discussions in MOOC forums (Amjad et al., 2022) and in social media (Zankadi et al., 2022) to identify student interests and assist in planning and improving the teaching methodology. Closer to our work, Chen et al. (2016) and Coelho & McCollum (2021) applied topic modeling to student essays. The first analyzed essays with a defined theme and found a correlation between topic relevance and the grades obtained. The second explored open-theme essays, offering insights into how the use of these analyses can help design culturally adaptive learning experiences. Following a similar approach, we aim to employ topic modeling to identify emerging topics from a substantial set of projects presented at a national level science and engineering fair, to illustrate the diverse range of themes and their responsiveness to contextual factors, such as the impact of students' geographical locations and significant temporal events.

Methods

Our approach to answer the research questions is structured as follows. First, we describe the setting at the FEBRACE, a significant Brazilian event promoting scientific culture among K-12 students nationwide. Secondly, we detail the data sources, specifically focusing on the dataset of 5,296 projects presented at FEBRACE from 2003 to 2022. Finally, we outline our data analysis techniques, particularly the use of *BERTopic* for topic modeling and additional analysis through statistical tests to explore distribution differences across variables.

Setting: The FEBRACE

FEBRACE is a major outreach program of the University of São Paulo, initiated in 2003, aimed at fostering a culture of scientific inquiry and entrepreneurship in Science, Technology, and Innovation among K-12 students in Brazil. Specifically targeting students from 8th to 12th grades, this program supports local science fairs, offers professional development through self-paced online courses and in-person workshops for teacher-mentors, and annually hosts the largest national science fair in the country.

Each year, FEBRACE attracts participation from hundreds of small, predominantly student-led teams from both public and private schools across all Brazilian states. These teams, guided by a teacher advisor but mainly driven by the students themselves, submit science and engineering projects which showcase significant student agency in project selection and execution.

The selection process for FEBRACE is rigorous and multi-staged. Initially, teams submit a detailed paper/report, along with a five-minute video presentation of their project for the first round of online evaluation. From over 2,000 submissions each year, up to 500 groups are then invited to a second round of online evaluations. Successful teams from this round are invited to present their projects in person at the main event in São Paulo, ensuring that the work showcased is predominantly the students' own effort. The projects are documented in annual proceedings published on the FEBRACE website, which include the title, authors, institution, abstract, and keywords.

Outstanding projects from the national fair are selected for further awards and opportunities, including participation in the International Science and Engineering Fair (ISEF), with which FEBRACE is affiliated. Furthermore, FEBRACE has inspired a network of regional fairs throughout Brazil, many of which serve as feeder events that recommend or promote projects for FEBRACE consideration.

Topic selection process

The selection of a topic or problem situation that will guide the science and engineering endeavors submitted and presented at FEBRACE is a key step in the project development process and is a relevant aspect to our study. A systematic investigation into the factors that influence the choice of topics by student groups, however, extends beyond the scope of this work. In addition, it is important to acknowledge the diversity of theme selection processes within the FEBRACE context, given its national coverage.

To shed light on the primary factors that influence the choice of topics, we contacted five teachers by phone, each of whom had mentored at least two projects in the past four years. Based on their experiences, we developed a short survey with two questions: (1) *How do students typically identify research problems or project themes?*, and (2) *Describe a specific project to illustrate the previous answer*. The survey was sent to 210 teachers who had mentored at least one project in the last two years. We received 35 complete responses, which were transcribed and coded by two authors. The main factors identified are detailed in Table 1.

[Insert Table 1 and title here]

These factors illustrate how multiple influences affect students during the topic selection process. Although nearly one-third of the responses cited external influences, these also include strategies that help students broaden their perception of topics to engage with. The most frequently mentioned factor, observing local problem situations, highlights the connection between project topics and students' local settings, which is a key assumption in this study. It is important to note that these findings are not the primary results of the study and are not conclusive regarding the main factors influencing topic selection. They serve only as a preliminary exploration to better understand the topic selection process in the context of the FEBRACE.

Data sources

This study utilized a dataset comprising 5,296 projects that were accepted and showcased at the in-person FEBRACE from 2003 to 2022. The dataset includes comprehensive project descriptions (title, keywords, abstract) along with additional data such as the year of presentation, school setting, and the state/region of origin. The primary source for this dataset were the FEBRACE proceedings, supplemented with information provided by the organizing committee. The dataset is available upon request.

Figure 1 illustrates the distribution of projects over the years. Table 2 categorizes the projects by geographical regions across Brazil and by types of educational institutions, including public schools, private schools, foundations (i.e., private schools subsidized through external foundations, offering significantly reduced fees), and other entities (e.g., projects by students participating in after-school programs or associated with NGOs, not directly linked to a specific school). It is important to note that the data for school types in Table 2 spans from 2010 to 2022, as schools were not classified into these categories prior to 2010.

[Insert Figure 1 and caption here]

[Insert Table 2 and title here]

Data analysis

Topic modeling was used to identify major topics among students' projects. We employed *BERTopic* (Grootendorst, 2022), a deep learning-based model that treats the task as a clusterization problem, leveraging recent advanced capabilities of Language Models (Reimers et al., 2019). *BERTopic* has been demonstrated to be highly effective for social sciences (Egger & Yu, 2022), and more specifically in educational research (Zankadi et al., 2022). The model takes the set of documents – each document comprising the title, abstract, and keywords of the project report, concatenated – clusters them together into topics, and then generates a set of representative words for each topic, which can be interpreted by humans in a later step. Before applying the text to *BERTopic*, we employed classical text pre-processing techniques to improve topic interpretation, including *Stopword Removal* (eliminating non-significant parts of the vocabulary, such as articles, connectives, prepositions) and *Lemmatization* (normalizing words by converting nouns and adjectives to their masculine and singular form and transforming verbs into their infinitive form), following the methodology proposed by Ferraz et al. (2021).

The model's learning process was guided by two automatic metrics, as proposed by Dieng et al. (2020): (i) ***coherence***, which measures an average degree of semantic similarity between the words that represent each topic; and (ii) ***diversity***, which measures the percentage of unique words representing the topics, indicating the overall variety of all topics. *BERTopic* managed to assign 58% of the projects in 72 topics with at least 10 projects, which is a reasonable amount considering the performance of this type of model for automatic categorization (Alcoforado et al., 2022). We analyzed the representative words provided for each topic and manually proposed concise terms to describe each topic after reviewing data from at least 5 examples of projects associated with them, examples of these terms are presented on Table 3. The terms were then validated and reviewed by external professionals from the FEBRACE organizing committee. Additionally, these professionals were asked to categorize each topic into its most representative UN Sustainable Development Goals (SDGs) (United Nations, 2015). The SDGs are generally used to communicate clear themes toward a peaceful and prosperous world, but in the context of FEBRACE, they are specifically used in its promotional materials to inspire and guide students' perceptions, encouraging them to explore themes they might find interesting and motivating them to propose science and engineering projects that are relevant to these goals.

[Insert Table 3 and title here]

Using the final topics, we compared their distribution across three variables: year (grouped in 4-year intervals), region (corresponding to the five macro regions of Brazil) and school setting (public vs. private, the most representatives in our sample). The school setting sample was smaller, because this information has only been available from 2010 onwards. We employed the Chi-square statistical test of independence (McHugh, 2013) to determine if there was statistically significant difference between groups for each variable (results are discussed in the Findings section). Finally, we identified topics with the highest values of dispersion and described top five lists for each variable to further explore the findings from the statistical analysis.

Findings

We organize the findings of this study around the two proposed research questions. First, we present the topics explored by students participating in FEBRACE, highlighting the most frequent ones with examples from our dataset. We also align those topics with the United Nations' Global Sustainable Development Goals. Following this, we examine the association of these topics with three contextual variables: year, region, and school setting. We identify the topics most influenced by these variables and illustrate their impact by analyzing the top five topics for each case within these variables. Additionally, we discuss preliminary factors that may explain these variations, providing deeper insights into the dynamics of student project themes.

What are the main topics that students participating in FEBRACE choose to explore in their projects?

The application of the topic modeling approach to our dataset categorized 3,087 projects into 72 distinct topics, with an average of 42.9 projects per topic and a standard deviation of 36.9. We employed automatic metrics to assess the quality of these topics: the coherence value was 0.62, aligning with the average results reported in existing literature (Röder et al., 2015), and the diversity score was 0.72, suggesting a reasonable distinction among the topics (Dieng et al., 2020).

Figure 2 displays the distribution of projects across these topics. The five most prominent topics, highlighted in yellow, comprise 748 projects, accounting for 24.2% of the total dataset. Additionally, the top 16 topics collectively encompass more than half of the projects. Notably, while the top five topics predominantly pertain to STEM subjects, we have marked in orange certain topics that exemplify the breadth of themes identified in our analysis. A complete list and occurrence of topics is available in Appendix 1.

[Insert Figure 2 and caption here]

In addition, Table 4 showcases samples of research questions from projects in the top five topics, which were also used for labeling these topics during the data analysis process. These questions were extracted from the projects' abstracts in the final report and were translated into English with minimal adaptations.

[Insert Table 4 and title here]

In-depth description of a sample topic

Delving deeper into projects within the same topic provides a richer understanding of the diversity within each topic (Figure 3). For example, a project from a public school in 2007 explored the use of solar energy for water purification, being classified as a Physics project. In contrast, a 2018 project from a private school designed an alert system for river flooding, approached as an Engineering project. Further illustrating this diversity, a 2021 project investigated the social impacts of unequal access to clean water in a large city, framed as a Social Sciences project.

[Insert Figure 3 and caption here]

These examples highlight that although projects may share a common topic, they can vary significantly in their specific questions, methodologies, and educational settings. This variation not only enriches the topic but also demonstrates the broad spectrum of inquiry that students engage in across different disciplines.

Mapping topics into Sustainable Development Goals

Finally, Table 5 illustrates the distribution of projects across the UNESCO Sustainable Development Goals (SDGs), following a validation exercise conducted with the FEBRACE Organizing Committee. The percentages indicate the distribution of projects across these goals. Fourteen out of seventeen SDGs had topics directly associated with them, while 3.5% of the projects did not align with any specific goal. Topics falling into this category included "Astronomy," "Acoustics," and "Aircrafts."

[Insert Table 5 and title here]

To what extent do the topics chosen by students at FEBRACE vary across different geographical, educational, and temporal factors?

Table 6 presents the results of the Chi-square Test of Independence applied to the distribution of topics using three variables: year (grouped into five ranges of four years each, from 2003-2006 to 2019-2022); region (corresponding to the five macro-regions of Brazil); and school setting (public versus private schools).

[Insert Table 6 and title here]

The degrees of freedom (DoF) for each variable were adjusted due to the test's constraint that requires at least five projects in the topics for each category. This adjustment led to a reduced number of topics (N_topics) for variables with five categories, such as 'year' and 'region'. Despite these adjustments, the p-values obtained ($p<0.05$) indicate a statistically significant relationship between the distribution of topics and each of the three variables. Notably, the p-values reveal clearer differences in the distribution of topics by school setting and year compared to region.

For each variable, we identified the topics that exhibited the highest dispersion across categories, as shown in Table 7. This identification was based on the standard deviation of each series, and we also noted the category where each topic was most prevalent. For example, we can infer from the table that “violence against women” was more common in private than public schools (#1 in “School setting”) and particularly frequent between 2019 and 2022 (#4 in “Year”).

[Insert Table 7 and title here]

Most frequent topics over time

In addition, we identified the five most frequent topics within each variable's categories to better illustrate the relationship between topics and variables. Table 8 displays the top five topics for each of the four-year intervals analyzed in this study. We employed a color scheme to highlight the 'intruders,' defined as topics that appear in only one or two intervals.

[Insert Table 8 and title here]

Some of these 'intruders' demonstrate how topics respond to the years in which projects were conceived. For instance, “dengue disease” ranks as the seventh most common topic among all projects, but its frequency significantly increased in the 2010s, coinciding with a rise in dengue cases in Brazil (Nunes et al., 2019). Similarly, “robotics”, the second most common between 2003 and 2006, might have lost space for the same technology being applied to other fields, such as in the example above of students developing an automated alert system for river flooding, or in projects in the topic “learning of robotics” (ranked #31 in the total frequency). More recently and notably, “Covid-19”' which is globally relevant, emerged as the fourth most common topic during the 2019-2022 period.

Most frequent topics across regions

Examining the most common topics for each region, as shown in Table 9 does not provide as much insight as the analysis by year intervals, with fewer 'intruders' among the top five topics. However, some preliminary observations suggest avenues for further research. For instance, “dengue disease” previously mentioned, is notably more relevant to students from the Northeast and Central West regions, which historically report the highest per capita incidence of the disease (de Castro Catão & Guimarães, 2011). Additionally, while "water resources” is a prevalent topic across all regions, its proportion varies significantly, with 10.2% of projects in the North, a region renowned for its extensive water resources associated with the Brazilian Amazon Forest.

[Insert Table 9 and title here]

Most frequent topics across school settings

Finally, three topics notably stood out in the top five results for the two school settings analyzed in this study (Table 10): “dengue disease”, which appeared in 4.1% of projects from public schools compared to 3.1% from private schools; “violence against women”, with 4.6% in private schools versus 1.4% in public schools; and “visual impairment”, observed in 4.6% of projects in private schools compared to 2.1% in public schools. While these differences alone do not warrant definitive conclusions, they suggest a need for further investigation into why certain topics are more concentrated in specific educational settings, which could inform future research directions.

[Insert Table 10 and title here]

Discussion

Our findings reveal that students at FEBRACE have been exploring a diverse array of topics in their science and engineering projects. Prominent among these are topics related to environmental studies and solutions, such as “water resources” and “sustainable agriculture,” reflecting responsiveness to Brazil’s rich natural resources and key economic sectors (e.g., The Economist, 2010). These topics typically align with traditional scientific inquiry within the Natural Sciences; however, the range of topics extends into the Social Sciences, including “teaching and learning,” “violence against women,” and “heritage languages.” Further research into these areas could demonstrate how K-12 students at science fairs might possess a more comprehensive understanding of science. Moreover, a detailed examination of projects within a single topic reveals how these inquiries span various disciplines, from biology to microelectronics and urban planning. This suggests that each topic warrants further exploration to understand the diverse reasons and contexts in which students frame their projects.

Aligning the topics with the Sustainable Development Goals (SDGs) has clarified the primary themes of student projects at FEBRACE, showcasing the potential of science fairs to engage middle and high school students in addressing real-world problems (Craven & Hogan, 2008). While three SDGs encompass nearly half of all projects, only three out of seventeen goals had no associated topics. This highlights the wide spectrum of issues accessible through SEFs. Considering the overlap among SDGs themselves, our analysis suggests that the SDGs effectively communicate the range of themes students can explore and serve as a catalyst for problem finding, prompting students to apply global perspectives to local challenges - for example, projects that enhance clean water and sanitation using indigenous plant species for water purification.

The results from our statistical tests reveal that the topics identified in this study vary depending on the time and location of the projects. For example, “Covid-19” has emerged as one of the predominant topics between 2019 and 2022, reflecting its significant impact on students' lives. Similarly, “dengue disease” varies regionally, affecting different parts of the country to varying degrees (Nunes et al., 2019). A more detailed analysis of the projects associated with these topics could further substantiate these findings, especially by examining how students navigate the problem-finding process that underpins their inquiries. Overall, our research demonstrates that SEFs provide a fertile ground for authentic and open inquiry (LaBlanca, 2008), as students navigate open-ended endeavors that are related to their local settings.

Implications for science and engineering fairs

The findings from our research highlight significant implications for science and engineering fairs (SEFs). First, the diversity of topics explored by students, as well as their responsiveness to specific contexts, helps to dispel the perception of students merely following “recipes” or replicating simple experiments (like the classic erupting volcanoes depicted in movies). Instead, our results indicate that science fairs serve as spaces for promoting authentic and open inquiries and fostering culturally responsive science education (Brown & Crippin, 2016). Communicating these alternative perspectives on SEFs is essential for attracting more diverse audiences to science fairs, addressing a challenge previously noted in the literature (e.g., Bencze & Bowen, 2009).

Our results also underscore the importance of providing students with adequate resources for their inquiries, considering the diversity of topics and objectives associated with their projects. We illustrate the Sustainable Development Goals (SDGs) not only as a framework for classifying topics but also as a tool for supporting the problem-finding process, broadening students' perspectives on different themes. It is equally important that this broader perspective is incorporated into research methodologies and tools for data collection and analysis, ensuring that projects not aligned with traditional views of science, particularly those associated with the Natural Sciences, are also supported.

Moreover, it is crucial to highlight the importance of supporting teachers in these processes. Previous research has emphasized the role of teachers as both mentors and content experts (Koomen et al., 2018). This is also true in our context, where projects are developed under the guidance of a supervising teacher, who is not responsible for defining the project's theme (see Table 1). More broadly, open inquiry processes can be mistakenly perceived as teacher-independent; instead, teachers play a vital role in scaffolding inquiry (e.g., Hmelo-Silver et al., 2007), particularly in the context of SEFs, which includes providing resources that enable students’ inquiries (Zion & Mendelovici, 2012). Therefore, we argue that the science fair community should recognize the importance of adequately equipping teachers with resources to support the diverse set of inquiries students might pursue.

Limitations

Several constraints limit the findings of this study, particularly regarding the data sources and students' experiences with SEFs. Firstly, we assume that the projects’ topics are defined by the students. Although preliminary evidence based on teacher input suggests this is the case for FEBRACE, more robust evidence through in-depth investigations, such as student interviews and case studies from various settings, would strengthen this assumption. Additionally, only selected projects presented at FEBRACE were included in the topic modeling analysis due to their public availability. Including data from all submissions might yield different results but would still face representativeness constraints, given the unequal access students have to SEFs. Moreover, the list of topics is specific to the context of this study and would likely vary in different contexts. Therefore, we do not intend to generalize these results to all SEFs. An important parallel contribution of this work is the proposed methodology, which can inspire the design of specific studies for other SEFs and open inquiry processes in contexts different from FEBRACE and Brazil. We hope this work will raise questions and encourage comparative studies in similar and varied settings.

Future Work

Future research can benefit from exploring the available data through different lenses. For instance, individual investigations of specific topics, combined with in-depth analyses of project samples, can provide meaningful insights into how students have conducted their projects and identify areas worth exploring. Similarly, detailed examinations of specific steps in project development, particularly problem-finding and solution design, can help identify key mechanisms that guide students toward topics. Additionally, incorporating additional variables associated with the projects, such as city size or Human Development Index (HDI), can expand our understanding of the factors that most influence topic distribution. Using the same topic modeling approach on datasets from other settings will also enable comparisons of common and unique themes across SEFs. Finally, similar to the work of Coelho and McCollum (2023), identifying these topics can inform design-based studies involving different actors in the SEF communities, leading to the creation of culturally informed resources that support students in their scientific and engineering endeavors.

Conclusion

In this study, we analyzed over 5000 projects from the FEBRACE national science fair to identify the predominant themes explored by students and how these themes vary across different contexts. Using topic modelling, we identified 72 distinct topics that could be linked to 14 out of 17 United Nations' Sustainable Development Goals, with significant variations influenced by time, region, and school setting, highlighting the relevance and responsiveness of their inquiries to global and local challenges. Our findings illustrate the authentic and open aspects of students' inquiries and provide valuable evidence supporting the role of science and engineering fairs as spaces for culturally responsive science learning. Moreover, we argue our findings argue for the importance of equipping both students and teachers with resources to engage a more diverse range of participants and support the varied interests that students bring to their scientific and engineering endeavors, both within and beyond the context of the fairs.

Acknowledgments

We thank the FEBRACE organizing committee for their support along the study. This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) in Brazil under Grant Projeto Ciência na Escola- 441081/2019-3.

Disclosure statement

The authors report there are no competing interests to declare.

References

Adler, I., Schwartz, L., Madjar, N., & Zion, M. (2018). Reading between the lines: The effect of contextual factors on student motivation throughout an open inquiry process. Science Education, 102(4), 820-855.

Alcoforado, A., Ferraz, T. P., Gerber, R., Bustos, E., Oliveira, A. S., Veloso, B. M., et al. (2022). ZeroBERTo: Leveraging zero-shot text classification by topic modeling. In *International Conference on Computational Processing of the Portuguese Language* (pp. 125-136). Springer.

Amjad, T., Shaheen, Z., & Daud, A. (2022). Advanced Learning Analytics: Aspect Based Course Feedback Analysis of MOOC Forums to Facilitate Instructors. IEEE Transactions on Computational Social Systems.

Bell, R. L., Smetana, L., & Binns, I. (2005). Simplifying inquiry instruction. The science teacher, 72(7), 30-33.

Bencze, J. L., & Bowen, G. M. (2009). A national science fair: Exhibiting support for the knowledge economy. International Journal of Science Education, 31(18), 2459-2483.

Boyd-Graber, Jordan, Yuening Hu, and David Mimno. "Applications of topic models." Foundations and Trends® in Information Retrieval 11.2-3 (2017): 143-296. http://dx.doi.org/10.1561/1500000030

Brown, J. C., & Crippen, K. J. (2016). Designing for culturally responsive science education through professional development. International Journal of Science Education, 38(3), 470-492.

Chen, J. J., Lin, H. S., Hsu, Y. S., & Lee, H. (2011). Data and claim: The refinement of science fair work through argumentation. International Journal of Science Education, Part B, 1(2), 147-164.

Chen, Y., Yu, B., Zhang, X., & Yu, Y. (2016, April). Topic modeling for evaluating students' reflective writing: a case study of pre-service teachers' journals. In Proceedings of the sixth international conference on learning analytics & knowledge (pp. 1-5).

Coelho, R. & McCollum, A. (2021). What Can Automated Analysis of Large-Scale Textual Data Teach Us about the Cultural Resources that Students Bring to Learning?. In de Vries, E., Hod, Y., & Ahn, J. (Eds.), Proceedings of the 15th International Conference of the Learning Sciences - ICLS 2021. (pp. 565-568). Bochum, Germany: International Society of the Learning Sciences.

Coelho, R., & McCollum, A. (2023). Teachers’ Engagement with Analysis Outputs of Large-Scale Identity Productions as a Move Towards Culturally-Informed Curriculum Development. In Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023, pp. 91-98. International Society of the Learning Sciences.

Craven, J., & Hogan, T. (2008). Rethinking the science fair. Phi Delta Kappan, 89(9), 679-680.

de Castro Catão, R., & Guimarães, R. B. (2011). Mapeamento da reemergência do dengue no Brasil-1981/82-2008. Hygeia-Revista Brasileira de Geografia Médica e da Saúde, 7(13), 173-185.

Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2020). Topic modeling in embedding spaces.

Transactions of the Association for Computational Linguistics, 8, 439-453.

Dionne, L., Reis, G., Trudel, L., Guillet, G., Kleine, L., & Hancianu, C. (2012). Students' Sources of Motivation for Participating in Science Fairs: an Exploratory Study within the Canada-Wide Science Fair 2008. International Journal of Science and Mathematics Education, 10, 669-693.

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. Frontiers in Sociology, 7.

Eloy, A., Ferraz, T. P., Alves, F. S., & de Deus Lopes, R. (2023). Science and engineering for what? A large-scale analysis of students' projects in science fairs. In Blikstein, P., Van Aalst, J., Kizito, R., & Brennan, K. (Eds.), Proceedings of the 17th International Conference of the Learning Sciences - ICLS 2023 (pp. 946-949). International Society of the Learning Sciences.

Ferraz, T. P., Alcoforado, A., Bustos, E., Oliveira, A. S., Gerber, R., Müller, N., D'Almeida, A. C., Veloso, B. M., & Costa, A. H. R. (2021). DEBACER: a method for slicing moderated debates. In Anais do XVIII Encontro Nacional de Inteligência Artificial e Computacional (pp. 667-678). SBC.

Grinnell, F., Dalley, S., & Reisch, J. (2020). High school science fair: Positive and negative outcomes. PloS one, 15(2), e0229237.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). Educational Psychologist, 42(2), 99–107.

Hume, A., & Coll, R. (2010). Authentic student inquiry: The mismatch between the intended curriculum and the student-experienced curriculum. Research in Science & Technological Education, 28(1), 43-62.

Koomen, M. H., Rodriguez, E., Hoffman, A., Petersen, C., & Oberhauser, K. (2018). Authentic science with citizen science and student driven science fair projects. Science Education, 102(3), 593-644.

LaBanca, F. (2008). Impact of problem finding on the quality of authentic open inquiry science research projects. Western Connecticut State University.

Lee, H., & Cho, Y. (2007). Factors affecting problem finding depending on degree of structure of problem situation. The journal of educational research, 101(2), 113-123.

Miller, K., Sonnert, G., & Sadler, P. (2018). The influence of students' participation in STEM competitions on their interest in STEM careers. International Journal of Science Education, Part B, 8(2), 95–114. https://doi.org/10.1080/21548455.2017.1397298.

Ming, N. C. & Ming, V. L. (2013). Visualizing Topics, Time, and Grades in Online Class Discussions. In Rummel, N., Kapur, M., Nathan, M., & Puntambekar, S. (Eds.), To See the World and a Grain of Sand: Learning across Levels of Space, Time, and Scale: CSCL 2013 Conference Proceedings Volume 2 — Short Papers, Panels, Posters, Demos & Community Events (pp. 105-108). Madison, WI: International Society of the Learning Sciences.

NGSS Lead States. (2013). Next Generation Science Standards: For states, by states. Washington, DC: The National Academies Press.

Nunes, P. C. G., Daumas, R. P., Sánchez-Arcila, J. C., Nogueira, R. M. R., Horta, M. A. P., & Dos Santos, F. B. (2019). 30 years of fatal dengue cases in Brazil: a review. BMC public health, 19, 1-11.

OECD (2018), "PISA for Development Science Framework", in PISA for Development Assessment and Analytical Framework: Reading, Mathematics and Science, OECD Publishing, Paris. https://doi.org/10.1787/9789264305274-6-en

Paul, J., Lederman, N. G., & Groß, J. (2016). Learning experimentation through science fairs. International Journal of Science Education, 38(15), 2367–2387. https://doi.org/10.1080/09500693.2016.1243272.

Ramnarain, U. D. (2020). Exploring the Autonomy of South African School Science Students When Doing Investigative Inquiries for a Science Fair. Eurasia Journal of Mathematics, Science and Technology Education, 16(12).

Reimers, N., Schiller, B., Beck, T., Daxenberger, J., Stab, C., & Gurevych, I. (2019, July). Classification and Clustering of Arguments with Contextualized Word Embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 567-578).

Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In Proceedings of the eighth ACM international conference on Web search and data mining (pp. 399-408).

Schmidt, K. M., & Kelter, P. (2017). Science fairs: a qualitative study of their impact on student science inquiry learning and attitudes toward STEM. Science Educator, 25(2), 126-132.

The Economist (2010). How to feed the world. The emerging conventional wisdom about world farming is gloomy. There is an alternative. The Economist. Retrieved August 20, 2024, from https://www.economist.com/leaders/2010/08/26/how-to-feed-the-world.

United Nations (2015) Transforming our world: the 2030 Agenda for Sustainable Development: A/RES/70/1. Retrieved from August 20, 2024, from: https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf.

Zankadi, H., Idrissi, A., Daoudi, N., & Hilal, I. (2022). Identifying learners' topical interests from social media content to enrich their course preferences in MOOCs using topic modeling and NLP techniques. Education and Information Technologies, 1-18.

Zion, M., & Mendelovici, R. (2012). Moving from structured to open inquiry: Challenges and limits. Science education international, 23(4), 383-399.