

## RESEARCH REPORT

# A Text Analytical Study of STEM Inquiries in Grad Slam Competition

Jianjun Wang<sup>1</sup> , Anna Jacobsen , Debra Jackson 

California State University, USA

### Abstract

*Grad Slam competitions offer a unique opportunity for students to showcase their STEM research. As a popular platform to promote STEM education in the United States, the competition limits presentations to three minutes, demanding rigorous training to improve graduate students' presentation skills. This research incorporates text analytics to extract the substance of Grad Slam projects across 2019-2022 and assess the effectiveness of Grad Slam training. Videos of Grad Slam presentations are transcribed to enable the use of Natural Language Processing to transform the unstructured text into normalized data suitable for processing by machine learning algorithms. R scripts are developed to disentangle the overall features of the Grad Slam outcomes. Survey data are analyzed to report student feedback about workshop preparation for project presentation. The text mining not only shows strong connections between Grad Slam presentation contents and STEM subject inquiries but also reflects a trend of strengthening research culture at a Hispanic Serving Institution. The survey feedback reconfirms the benefit of workshop training. The Grad Slam competition and its related workshops emerge as a unique platform, aside from STEM coursework, to help graduate students excel in their specialty fields, as well as establish confidence in communicating their research to a general audience.*

**Keywords:** *Grad Slam, Three Minute Thesis, Text Analytics, Communication Skills.*

### Introduction

A Grad Slam is an academic competition that challenges graduate students to concisely present their research in an engaging manner to a general audience. It has gained increasing

---

<sup>1</sup> Corresponding: California State University, 9001 Stockdale Highway, Bakersfield, USA, [jwang@csu.edu](mailto:jwang@csu.edu)

To cite this article:

Wang, J., Jacobsen, A., & Jackson, D. (2025). A text analytical study of STEM inquiries in grad slam competition. *Journal of Research in STEM Education*, 11(1), 19-54. <https://doi.org/10.51355/j-stem.2025.185>

popularity in higher education because effective communication skills are critical for academic and professional success (Bishop, 2014). The event also contributes to the enhancement of science, technology, engineering, and mathematics (STEM) education by allowing participants and audiences to obtain insights across various subject-based inquiries. Therefore, the competitive aspect of Grad Slam not only inspires higher levels of academic performance but also creates a collaborative community to build learning capacities within and between departments. The competition further pushes students to establish a deep understanding of the challenges they may face in research dissemination. Held at a Hispanic Serving Institution (HSI), this event can help promote role models and shatter stereotypes in STEM education. Eventually, it paves the way for increased Hispanic representation to enhance a more inclusive and diverse scientific community.

Prizes and/or recognitions for Grad Slam presentations strengthen the institutional visibility among key stakeholders, including faculty, peers, and the public. As a result, the positive image can attract high-caliber students and faculty, advancing the institution's overall academic excellence. In addition, as a platform to accommodate diverse scholars, it enables the institutions to incorporate their cultural heritage in enriching the meaningfulness of STEM inquiries. Based on the overall impact at both student and institution levels, this research is designed to fit the interest of a dual target audience, i.e., the future STEM researchers who participated in Grad Slam competitions and the university leaders who cared about the development of campus research culture.

Research undertakings are inseparable from the population's ethnic composition. As the Latino population grows in higher education, assisting HSIs to become powerhouses of STEM research, especially at the postbaccalaureate level, supports the nation's global competitiveness and technological advancement. Beyond the subject competency, Good, Masewicz, and Vogel (2010) identified "communication gaps as the major barrier that impeded [Latino] student achievement" (p. 327). In this context, Grad Slam competitions at an HSI can foster subject belonging for traditionally underserved students who are hampered by language barriers in STEM education (Rodriguez et al., 2022).

To date, however, studies have yet to be conducted to assess the value of Grad Slam preparation and extract the essential information from student presentations. Meanwhile, text analytics has emerged as a new and powerful research tool in data science that encompasses a suite of computational techniques aimed at extracting meaningful information from unstructured textual data. Central to these techniques are Natural Language Processing (NLP) and machine learning, which collaboratively facilitate the aggregation, analysis, and interpretation of qualitative data. On the one hand, NLP is positioned in a subfield of artificial intelligence that focuses on the interaction between computers and human language (Stryker & Holdsworth, 2024). On the other hand, machine learning involves training algorithms on data to automatically categorize text, identify patterns, extract relevant information, and make trend predictions (Lexalytics, no date). In combination, text analytics becomes a useful tool that leverages NLP to

interpret human language and machine learning to analyze and derive insights for evaluating educational programs.

Due to its capacity to process vast amounts of unstructured text, the advancement of text mining has demonstrated tremendous potential for systematic examination of qualitative feedback in educational research. For instance, Dake and Gyimah (2023) applied sentiment analysis to educational data, providing insights into students' emotional states during presentations. Ferreira-Mello et al. (2019) also provided an overview of educational text mining applications, highlighting their role in uncovering trends in educational data. Huang et al. (2020) further discussed the use of text analytics to examine large-scale educational data for student engagement. These past examples provided additional background on text analytics applications in educational research.

Built on the extensive actionable capacity of text analytics, this article begins with a brief literature review to specifically clarify the context of Grad Slam competitions in STEM education. In a subsequent method section, an innovative approach to text analytics is described for investigating the features of student inquiries demonstrated in Grad Slam presentations. In addition, student survey data are employed to delineate the feedback about the learning opportunities. The implication of research findings is clarified in a discussion section to reflect on key questions derived from the current research literature.

## Literature Review

Academic contests for graduate students have been around for quite some time, but the specific term "Grad Slam" and its organized competitions may have varying origins in different countries. In the United States, the University of California (UC) system is often credited with popularizing Grad Slam (see Aguilera, 2019). For instance, it was reported that "UC Santa Barbara was the first UC campus to implement the Grad Slam format in 2013" (Cohen, 2015, p. 6). The idea was to have graduate students from different disciplines present their research to a panel of judges and a general audience. This platform is aimed at training students to communicate the significance of their research in a clear and understandable way to non-specialists. The competitions are sometimes branded as "Grad Slams" in California, referencing "poetry slams" during which poets perform spoken word to a live audience, as well as the term "grand slam" which designates a championship athletic contest, except for using the term "grad" as an abbreviation for "graduate student" in place of "grand."

In Australia, similar competitions can be tracked further back to 2008 when the University of Queensland launched a Three-Minute Thesis competition, or 3MT, as an innovative academic contest in graduate education. It challenged graduate students to present their research in 180 seconds or less to a general audience using one static slide. This competition has grown in popularity, with over 900 universities across 85 countries hosting the event each year (Three

Minute Thesis, no date). In particular, 3MT competitions are celebrated for their exhibition of graduate students' academic, presentation, and research communication skills (Three Minute Thesis, no date). As White (2021) acknowledged, the competition "stands out as a particularly important professional development activity" (p. iv). While there is a body of research examining the rhetorical strategies used by presenters during 3MT competitions (Carter-Thomas & Rowley-Jolivet, 2020; Hyland & Zou, 2022; Hu & Liu 2018; Kathpalia 2024; Ma & Jiang 2025), little is documented about participant experiences of participation (Muslimin & Zaki, 2024) or the impact of training and preparation activities on the content and quality of their presentations.

The Grad Slam or 3MT competitions are frequently recorded and posted on university and social media websites. For many participants, this can be intimidating as fear of public speaking is common both inside and outside of higher education settings (see Grieve et al, 2015; Lall & Biswas 2020). Students are expected to distill compelling STEM information into clear and common-sense explanations, which might trigger language barriers for some presenters. As Bell (2023) observed, "for graduate students who spend hundreds of hours on their research, Grad Slam is a challenge because they have to communicate the highlights in just 3 minutes!" (p. 1).

Nonetheless, the availability of Grad Slam presentations for viewing by a wider audience provides benefits. It allows viewers to pause and rewind, which can be particularly beneficial for understanding the projects of a group of Hispanic students who have English as their second language. Video recording also fits the needs of adult learners, including graduate students, who often balance their studies with work and family commitments. Knowing their presentation will be recorded and potentially viewed by a large audience can motivate them to refine their communication skills and deliver a clear presentation. Further, in STEM education inquiries, incorporating video recording is considered an effective approach. For instance, Gerald Bracey (2003), a columnist of Phi Delta Kappan, openly praised the TIMSS 1999 Video Study, "in which videotapes of teachers actually teaching eighth-grade mathematics in seven countries were studied" (p. 253).

While the UC system is a notable leader of Grad Slam support, the California State University (CSU) system serves many more Hispanic students in California. Unlike the high selectivity across UC campuses, CSU is known for its accessibility and focus on providing affordable and quality education to a broad range of minority student groups. Video recording of STEM activities offers a visual illustration of role models for future students, particularly those of Latino origin as the largest minority in the state. The method is effective because the digital materials can be presented to large numbers of students "without any additional effort on the role model's part and are at least as effective as live interactions" (Gladstone & Cimpian, 2021, p. 15). It also empowers Hispanic graduate students by bolstering their confidence in STEM subject inquiries.

Under Title V, Part B (Title Vb) provision of the federal government, the U.S. Department of Education established a Promoting Postbaccalaureate Opportunities For Hispanic Americans

(PPOHA) Program to address a two-fold purpose of expanding (1) postbaccalaureate opportunities for, and improve the academic attainment of, Hispanic students; and (2) the postbaccalaureate academic offerings as well as enhance the program quality in the institutions of higher education that are educating the majority of Hispanic college students and helping large numbers of Hispanic and low-income students complete postsecondary degrees. This study receives funding from PPOHA to support Grad Slam preparation at California State University, Bakersfield (CSUB), an HSI with a service area as large as the state of New Jersey. As the university president announced, “Grants and programs throughout the university’s four schools seek to improve Latinx representation, particularly in STEM professions” (Zelezny, 2022, p. 3). In supporting the PPOHA grant alignment with the commitment of CSUB to improving STEM education, dual questions are adduced to guide this investigation:

1. What are the top features of Grad Slam presentations prior to and during the period of program funding?
2. What is the student feedback about the learning experiences from workshop preparations for the Grad Slam?

While survey data are collected from Grad Slam workshops to address Question 2, Question 1 demands rich information extraction through text mining, which includes (1) the creation of an online portal to transcribe the video content into text files, (2) the application of NLP to transform the unstructured text into normalized data suitable for processing by machine learning algorithms, (3) development of R scripts to extract the overall features of the Grad Slam outcomes.

From the methodology perspective, the impact of text mining in educational research has been expanded in recent years, but not for the Grad Slam information extraction. For instance, a study by Yang, Kinshuk, and An (2023) utilized text mining to analyze educational studies, providing insights into how scholars employ text mining in their research. By examining the transcripts of presentations, researchers can identify prevalent themes, jargon usage, and the complexity of language. This analysis aids in understanding how effectively students convey complex ideas to a general audience.

It has also been generally agreed that “Compared to information obtained through surveys and interviews, the information provided by video analysis tends to be more objective” (Wang, Wang, Raymond, & Wang, 2023, p. 6). According to Ranjan and Mishra (2022), the adoption of text analytics holds immense value for video information extraction. In comparison to traditional methods in qualitative investigations (Best & Kahn, 2010), R can process large datasets, such as the ones from Grad Slam videos, much faster than manual coding. More importantly, it can ensure consistent result replications and has robust packages for data visualization. Built on the power of computational linguistics and machine learning, text mining techniques have largely revolutionized the way researchers aggregate a rich source of information from unstructured data to summarize student presentations of STEM inquiries

(Baker & Yacef, 2009). Details of the data selection, information cleaning, and R-based analytics are elaborated in the method section.

## Methods

### *Data Selection*

To strengthen the impact of this investigation, it is important to note a learning gap in which Latino students often face unique challenges in higher education, particularly in a public event like Grad Slam that requires science communication skills to present complex research succinctly to diverse audiences. Since HSIs play a pivotal role in educating Hispanic students, choosing an institution with a strong Hispanic serving mission can shed light on the unique challenges and strengths of the Grad Slam competition. According to Excelencia in Education (2024), the Hispanic full-time equivalent student ratio has been retained at 60.8% across CSUB's graduate programs, while the concurrent average ratio in the nation was 36.7% in Academic Year 2023-2024. Built on the data indicator for strong Latino student services, Grad Slam videos are selected from CSUB to address Question 1 during the period of Title Vb funding.

As a result, the video archives are gathered online and transcribed verbatim using Otter.ai, an AI-based transcription platform, to ensure accuracy and consistency. The texts were manually reviewed to correct errors in presenter identification, punctuation, and terminology. The transcription files were exported in txt format for compatibility with R scripts and subsequent preprocessing steps of data cleaning.

### *Information Cleaning*

Meticulous data cleaning is indispensable for improving the reproducibility of research outcomes from text analytics (Upadhye, 2020). Techniques such as tokenization, stemming, lemmatization, and stopword removal are applied to strengthen the format uniformity. More specifically, each token (word) is defined as a sequence of alphanumeric characters. A heuristic approach is taken to trim prefixes and suffixes through word stemming to reduce the tokens to their root forms. Tokens are further simplified into the base structures in a lemmatization process to ensure that the resulting configuration is a valid word in English. Stop words (e.g., "and," "the," "is") are removed using predefined stop word lists from the stopwords package in R. Custom stop word lists are applied to exclude terms that are not analytically meaningful. Text cleaning also involves converting all characters to lowercase, removing numbers, punctuations, and special symbols, as well as replacing contractions with their expanded forms (e.g., "isn't" → "is not"). Correction of misspellings and wrong data entries has added another layer of protection for maintaining data integrity.

Altogether, the data cleaning process has addressed noise, enhanced consistency, corrected errors, and facilitated reproducible workflows to make the analyses both reliable and



replicable. Tokenization and text preprocessing standards adhered to best practices in NLP, enabling accurate, reproducible, and insightful analysis of presentation transcripts from academic competitions. By integrating precise transcription methods with robust text analytics techniques in R, this study ensured the comprehensive processing of textual data.

In summary, the data choice is designed to foster more inclusive Grad Slam support for STEM success among graduate students, particularly those of Hispanic origin, to promote diversity and advance STEM research communication. The ability to achieve consistent results using quality data and reliable methods is also a general cornerstone of scientific research, which demands the reduction of irrelevant information, inconsistencies, discrepancies, and non-standard abbreviations. Cleaning processes like text restructuring, tokenization, stemming, lemmatization, stop word removal, and spell-checking help mitigate these issues, leading to more accurate analyses. Upadhye (2020) emphasizes that text data cleaning is a crucial preprocessing step in NLP and text data analysis aimed at improving the quality, reliability, and usability of textual information.

#### *Method Choice*

The method choice is aligned with two questions for this investigation. While student feedback about workshop preparations for the Grad Slam competition is gathered from a participant survey (Question 2), the text analytical methods are relatively new. To address the objective of extracting the substance of Grad Slam projects across 2019-2022 in Question 1, this study followed Kellogg's (2024) classification of data analysis for studying (1) outcome differences, (2) variable relations, (3) trend descriptions on the time dimension.

As a result, a lexical dispersion plot (LDP) is employed to visualize the difference in content emphases among different data corpora. It displays the occurrences of target words across multiple documents to highlight where and how frequently terms appear (Svartzman et al., 2020). Hutchinson (2022) recommended LDP, which "each strike along the word offset axis signals that a specific word is mentioned within the corpus of data" (p. 56). Due to the focus on extracting essential information, the density of strikes in LDP is an important outcome. However, the specific time a word appeared depends on the sequence arrangement of the Grad Slam presentations, which has nothing to do with the content coverage. In other words, it does not matter whether the word appeared relatively early or late if it occurred. Hence, labels are omitted for the axis of the relative token index in LDP to avoid unnecessary distraction from the occurrence time. This practice follows the LDP literature to focus on the strike density. As Amin et al. (2022) asserted, LDP allows researchers to determine "how many times the word (or multiple words) occurs from the beginning to the end of the text" (p. 25).

The examination of Grad Slam outcome differences is further extended from the token tracking in LDP to a broad summary of multiple documents in a word cloud plot. Highlighting frequently mentioned words helps identify trends and gain an overall picture of the topic

coverage from 2019 to 2022. According to Mostafa, Feizollah, and Anuar (2023), “A word cloud plot is an appealing visual tool that can be used to summarize textual data” (p. 12434).

Another feature of text mining hinges on keyness plot construction. “Unlike frequency counts, the keyness of a word does not necessarily anticipate a high, but rather an unusual, frequency” (Đurović, 2023, p. 188). Koch et al. (2022) and Weinberg (2021) recommended the keyness method to contrast the text features, which permits this study to make a comparison between Grad Slam 2022 and the previous competitions. In addition, a plot of the most frequent words (MFW) is needed to visualize repeated themes through text data aggregation (e.g., Kostelev & Bagić Babac, 2022). Therefore, an MFW plot is constructed to identify common themes across Grad Slam presentations.

In terms of the relationship perspective, plotting token-indicator relations involves picturing the associations between specific words or phrases (tokens) and particular variables or outcomes (indicators), which is a powerful technique for uncovering meaningful patterns within textual data. “This analytic technique allows for evocativeness and interpretation as well as attunes the research to dynamic interrelationships in and across themes” (Scharp, 2021, p. 549). Therefore, a plot of Token-Indicator Relations (TIR) is created in this investigation. Due to its computing intensity, however, the Quantitative Analysis of Text Data (Quanteda) package in R limits the number of tokens for the network construction (Watanabe & Müller, 2023). In this study, STEM subject terms of environmental, nursing, health, and mental health sciences are extracted for the TIR plotting. The token selection is based on the graduate program offerings in geology, biology, nursing, and psychology at CSUB (2023).

In summary, the Quanteda package is employed to support the R script execution for key information extraction in Question 1. According to Benoit (2018), “Using C++ and multithreading extensively, quanteda is also considerably faster and more efficient than other R and Python packages in processing large textual data” (p. 774). Meanwhile, the data collection is protected by a research protocol of the Institutional Review Board (IRB). Starting in 2020, four workshops have been offered to prepare students for the Grad Slam competition. While Workshop A is designed to help students familiarize themselves with Grad Slam features, Workshops B, C, and D address specific themes of Crafting a Memorable Message, Designing a Compelling Visual Presentation, and Communicating with Confidence, respectively. Due to the pandemic impact in 2020 and 2021, inadequate survey data were gathered from the workshop attendees. Hence, feedback for this study is delimited to the workshop evaluation in 2022 (Question 2). All participants of Grad Slam were recruited from the graduate student group at CSUB. Per IRB stipulation, the survey did not collect individually identifiable data, and thus, it did not involve consent form administration or demographic information gathering. Statistical findings and qualitative comments are analyzed to summarize the survey results.



## Results

### *Findings from Text Mining*

After data cleaning, an LDP has been drawn from the text data to compare meaningful tokens across the four Grad-Slam events. Some of the tokenized terms, such as nurs[s] and scienc[e], are shown as truncated words in LDP to reduce the matrix sparsity. As Katre (2019) put it, “Lexical Dispersion Plots effectively depict how multiple topic keywords appear throughout the corpora” (p. 8579). In Figure 1, keywords are extracted from the presentation content to indicate STEM topic coverage among the Grad Slam sessions. Details of the result interpretation are elaborated across 2019-2022 in the discussion section.

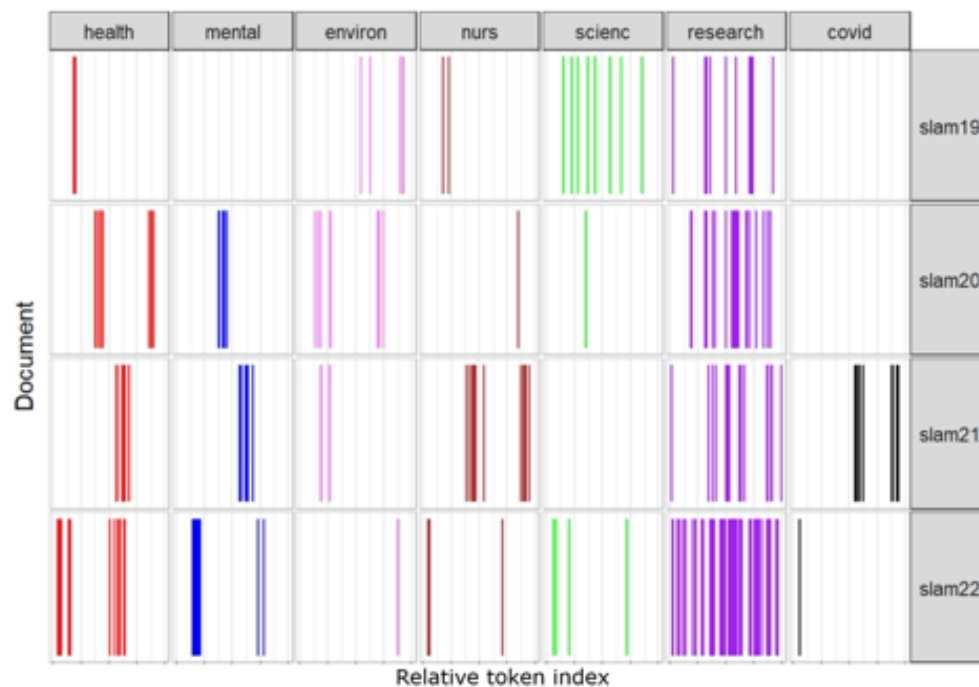


Figure 1. Dispersion of Content Coverage in Grad Slams

In Figure 2, a word cloud plot is constructed in such a way that “The size of each word and its closeness to the cloud center determine its significance” (Mostafa, Feizollah, & Anuar, 2023). The tokenized terms are closely related to biology, geology, and other STEM fields to echo the topic identification from a keyness plot in Figure 3. In comparison to Figure 2, Figure 3 leans toward the probabilistic inference because “Keyness analysis introduces the term ‘key word’, which is a word that occurs in a text more often than we would expect to occur by chance alone” (Rao & Taboada, 2021, p. 30).

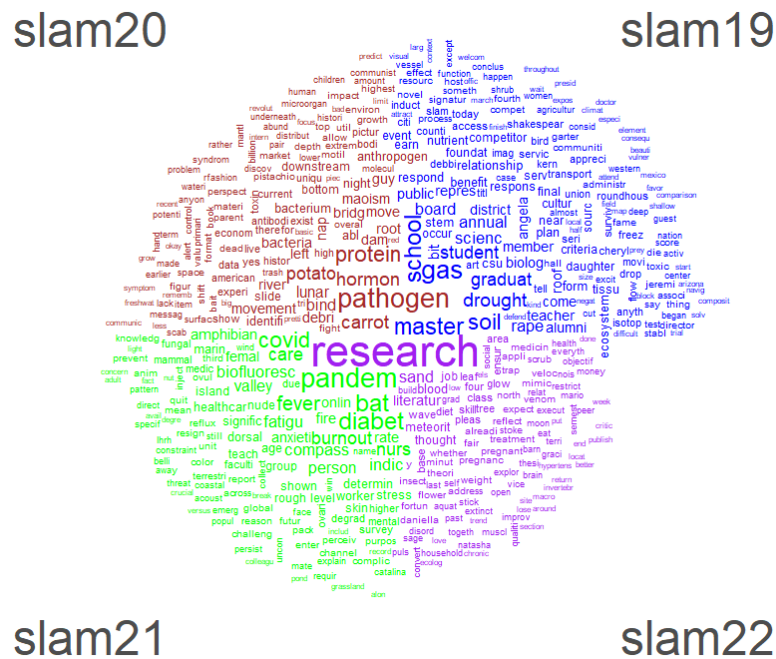


Figure 2. Word-Cloud Plot on Key Components of Grad Slams

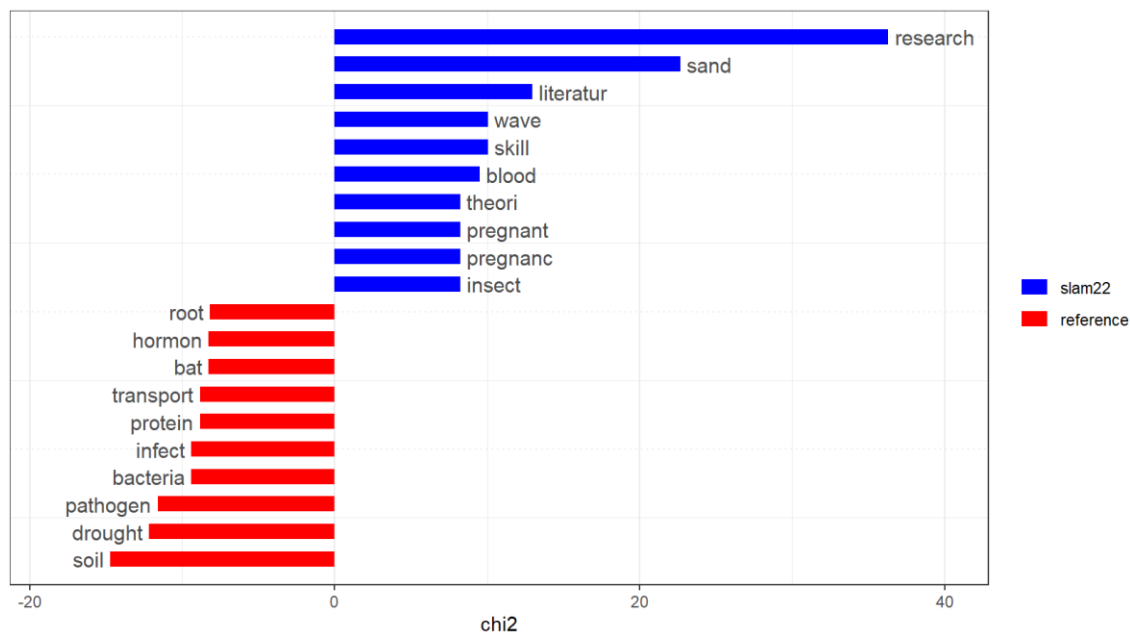


Figure 3. Contrast of Grad Slam 2022 on Year 3 Outcomes

With the Title Vb funding, student presentations in 2020, 2021, and 2022 consistently addressed topics in mental health, an area missed in Grad Slam 2019 before the pandemic. The increase in nursing research projects also coincided with the peak of COVID-19 impact in 2021.

Unlike the baseline count in 2019 (Figure 1), research is the tokenized term that shows the most coverage across the projects in Grad Slam 2022 (see Figure 3). Therefore, the grant funding from 2020 to 2022 is not only linked to the expansion of STEM topic coverage but also strengthened the research emphasis of the Grad Slam competition.

The text data are further aggregated across Grad Slam sessions to display a plot of the top impact words in Figure 4. After the text data cleaning, tokenized terms exhibit more project emphasis in the fields of biology, geology, and nursing. In part, this is because these departments have more graduate student enrollments at CSUB. These STEM program offerings are aligned with the need to enhance pollution control and health support that are essential for sustaining agricultural and petroleum industries in local communities. The emphasis on health and mental health also demonstrates how the Grad Slam competition has served as a dynamic platform that has evolved in response to global challenges over the years. The data reflect not only the adaptability of graduate researchers but also the importance of academic competitions in showcasing a timely and impactful investigation.

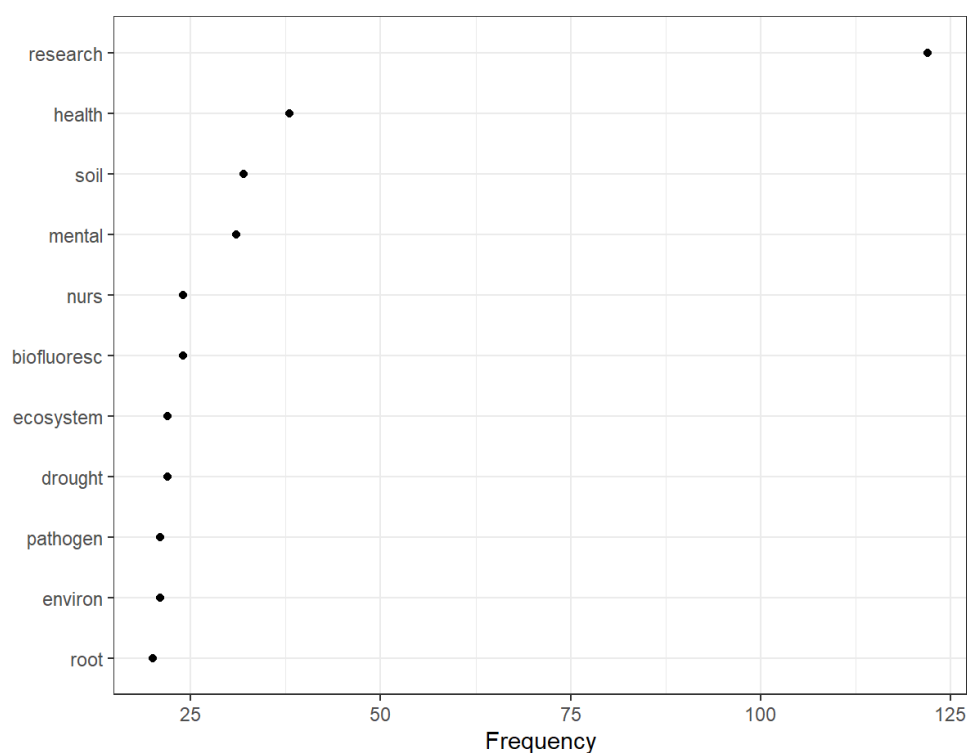


Figure 4. Most Frequent Words across Gram Slam Presentations

In alignment with the topic trend, a token-indicator relation plot further reveals the core role of research that articulates different field-based inquiries across the Grad Slam presentations. In Figure 5, the term research is located at the center to link multiple STEM fields of health, mental health, scienc[e], environ[ment], and nurs[ing]. Clearly, the analysis of Grad Slam presentations reveals a notable increase in the prominence of health and mental health-related themes in the years following the onset of the COVID-19 pandemic. This pattern reflects a shift in research focus and public interest, driven by the global health crisis.

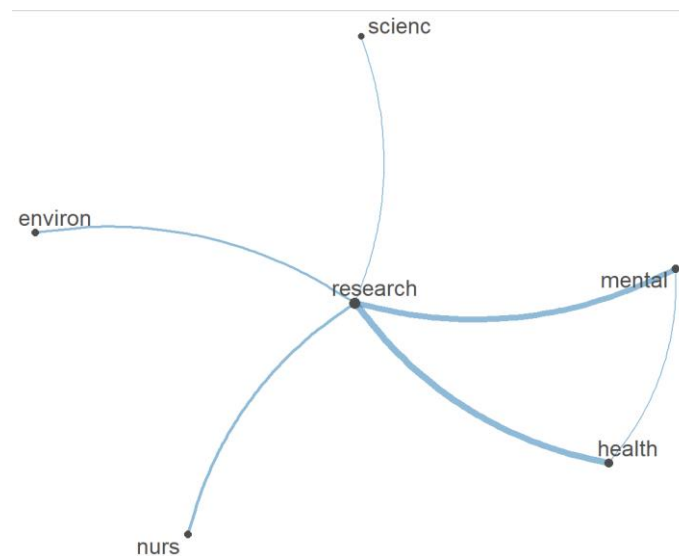


Figure 5. Token-Indicator Relations in Grad Slam Projects

### Survey on Grad Slam Preparation

In preparation for the Grad Slam presentation, four workshops were offered to describe effective strategies for handling the competition. Workshop A is intended to offer an introduction and help students understand the Grad Slam features. In contrast, Workshops B and C are more specific, focusing on useful techniques for Crafting a Memorable Message and Designing a Compelling Visual Presentation. Workshop D addresses Communicating with Confidence in general. As a result, 95% of the respondents believe that Workshop A is easy to understand, and all respondents rated Workshops B, C, and D in an extremely useful or very useful category. Meanwhile, 90 – 100% of respondents claimed the learning experiences met their expectations (Table 1).

Table 1.

### *Student Feedback about Grad Slam Training*

Workshop	N	Results
A	21	- 95% reported the topic “easy to understand” - 90% claimed “meeting their expectations”
B	3	- 100% reported “extremely useful” or “very useful” - All believed that the event met their expectations
C	7	- 100% reported “extremely useful” or “very useful” - All believed that the event met their expectations
D	29	- 100% reported “extremely useful” or “very useful”

Students also provided the following short comments to confirm the effectiveness of Grad Slam preparation:

- Thank you so much for putting this workshop together today. It was very empowering, useful and practical.
- This is good to boost confidence. I wish we have more of this!
- Great experience. Learned so much about so many different topics.

In summary, the survey results and text analytic findings have resulted in credible evidence to substantiate the benefit of student preparation for the Grad Slam presentations during the Title Vb grant funding. The results of text mining repeatedly indicated the focus of Grad Slam presentations on STEM fields. Additional support from the grant funding has led to more research emphasis on the Grad Slam competitions. The learning experiences from workshop training are considered useful according to the complimentary feedback from a participant survey.

### **Discussion**

This research has a dual focus to (1) disentangle the top features of Grad Slam presentations and (2) evaluate student feedback about the learning experiences from the workshop preparation. These emphases correspond to two research questions that are derived from the needs assessment in the literature review section.

Regarding the first focus, a more detailed discussion is needed to interpret the trend of Grad Slam presentations over 2019-2022, especially concerning health and mental health topics, to highlight the impact of COVID-19 on the project themes. In particular, the heightened awareness of health issues has led graduate students to align their research involvements with the unprecedented challenges of society. Across the United States, the impact was felt by all students who faced campus shutdown and had to take classes online (Lee et al., 2021). The aftermath inevitably reduced social interactions, decreased learning engagement, and created feelings of isolation. Meanwhile, universities and funding agencies have prioritized health-related research in response to the pandemic, which enabled students to pursue projects in these areas for Grad Slam competitions. As indicated in the results section, the pandemic has catalyzed

a substantial shift in graduate research interests driven by personal experiences, societal needs, and institutional support.

To clarify the trend in Grad Slam coverage, Figure 6 is created from Figure 1 to highlight the "Health" and "Mental Health" topic counts in 2019-2022. Prior to COVID-19, the trend showed no student presentations on mental health and one mention of health in 2019. The pattern grew to eight presentations in mental health and seven projects in health in 2022. This outcome is corroborated by the word plot in Figure 2, which incorporates all words from the Grad Slam presentations into consideration. The tokens of larger size indicate research as the primary feature of the STEM inquiries in 2022, while Pandem[ic] gained more project coverage in 2021 after the eruption of COVID-19 in 2020. Besides the trend description, a summative examination of the keyness plot in Figure 3 reconfirmed the enhancement of inquiry emphasis by indicating "research" and "literatur[e]" among the top three most important tokens in 2022. It also highlighted "theori[es]" along with tokenized terms in biology [e.g., blood, insect, pregnant, preganc(y)], physics, and geology [e.g., wave, sand] to show enrichment of confirmatory STEM inquiries in student projects in the same year. In contrast, no such features were extracted from the video recording of Glad Slam presentations before 2022. When the most frequent words were aggregated across Gram Slam presentations, Figure 4 shows that research was mentioned more than three times than the second most frequent term of health, making it the token of paramount importance in this period.

These findings jointly supported the positioning of research at the center of the token network in Figure 5, which involves STEM inquiries in science, environments, nursing, health, and mental health across the Grad Slam projects. Hence, the top features prior to and during the period of program funding not only show timely alignments of the Grad Slam topics with the contemporary issues of the pandemic and the consequences of mental health stress but also consistently promotes the strengthening of research culture per encouragement of the federal program funding (see Wang, 2023).

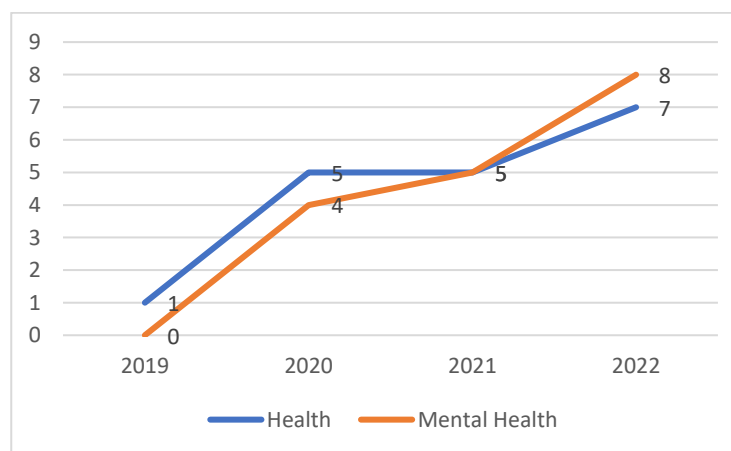


Figure 6. Trend of "Health" and "Mental Health" Topic Counts in Grad Slam Presentations



For Question 2, the quantitative results in Table 1 indicate that the learning experiences from the Grad Slam preparation were useful and meeting student expectations. The workshop training is also related to the competition setting. Sadlowski (2018) observed, “With only three minutes to present, summarizing and simplifying years of research can be a challenging task for Grad Slam participants” (p. 2). A student recapped the benefits from both Grad Slam participation and workshop preparation.

The Grad Slam competition was a rewarding experience and instrumental in providing me with the confidence to discuss my research. Engaging with my fellow Grad Slam participants and watching everyone discuss their research was very enriching. Collaborating on my research with my mentor has provided me with important skills and knowledge to enhance my career. I’m very fortunate to have received this grant and the opportunities it has provided.

By introducing text analytics in this study, the student learning experiences can be enriched in future investigations to facilitate Grad Slam preparation. More specifically, the past Grad Slam performance can be disentangled through an examination of the presentation transcripts to evaluate clarity, coherence, and the effective communication of key concepts. The NLP approaches also enable evaluators to identify the use of technical jargon, assess the logical flow of ideas, and determine the alignment of the content with the intended audience's comprehension level. The tool can provide immediate, personalized feedback for the student to better engage their audience and convey enthusiasm for their subject matter in alternative ways, thereby enhancing communication skills. Furthermore, the machine learning mechanism incorporated in text analytics can help train and extract textual features from presentations to predict performance outcomes for additional support and timely interventions.

In summary, the in-depth analysis of text data and student feedback indicated effective learning experiences from the Grad Slam workshop training that have been rarely studied before. To fill the void of research literature, an innovative approach has been taken to explore features of extracurricular learning with the cutting-edge method of R-based text analytics. By providing objective, scalable, and insightful analyses of presentation content and delivery, text analytics contributes to a more comprehensive assessment process for evaluating student competitions like Grad Slam.

## Conclusions

Findings from this investigation have a strong implication in STEM education, particularly for graduate students at HSIs. It is well-known that STEM education, traditionally viewed as a domain for experts, is increasingly recognized as vital for all. Thus, the Grad Slam competition emerges as a unique platform to help graduate students not just excel in their specialty fields but also establish confidence in communicating their research to a general

audience. Improvement of communication skills also enhances a sense of belonging to STEM inquiries, which is essential for career advancement, interdisciplinary collaboration, and leadership preparation (Morreale, Osborn, & Pearson, 2000). Hence, studies of Grad Slam competitions have tackled a unique platform, aside from STEM coursework, to help graduate students excel in their specialty fields, as well as establish confidence in communicating their research to a general audience.

Drawing from the methods and findings used in the Grad Slam analysis, implementing text analytics in other educational competitions can offer valuable insights to improve participants' preparation, presentation quality, and audience engagement. More specifically, LDP can be recommended for identifying patterns in the timing and recurrence of impactful words to ensure consistent use of key terms in strategic locations of debate competitions. The word cloud plot can help visualize the central themes of a presentation to guide participants' attention to profound topics that resonate with judges of STEM fairs. The keyness plot is effective in highlighting the importance of distinctive words with strong associations to specific judging criteria that are aligned with the competition's goals. The MFW plot supports workshop designs by replacing overused terms to enhance originality for audience attraction. The TIR plot is a useful tool for training participants to use tokens with strong positive correlations to performance metrics, such as clarity, relevance, engagement. Hence, the methods of text analytics illustrated in this study represent a powerful tool in elevating the quality of participation and evaluation in a wide range of educational competitions. These approaches directly help participants meet judging expectations while fostering innovation and clarity in their presentations.

While the value of this investigation has been adequately justified, three limitations are adduced for future improvement. First, this study relies on textual data from Grad Slam presentations, excluding non-verbal elements such as body language, tone, and visual aids, which are integral to effective communication. The partial information analysis might lead to an incomplete understanding of what contributes to successful presentations. Secondly, each presentation in Grad Slam has a limited scope due to the three-minute constraint. Consequently, findings may not fully capture the breadth of topics for a broad generalization across the STEM educational contexts. Thirdly, the analysis is built on automated techniques such as tokenization, keyness, and word frequency, which may miss subtleties like sarcasm, metaphors, or cultural references that contribute to the richness of the presentations. This limitation might imply that important qualitative aspects of the content could be overlooked for a comprehensive information extraction.

To address these gaps identified from this study, future investigations may incorporate new tools from R-based text analytics, such as sentiment analysis of audience reactions, to identify the elements of presentations that spark the most interest or confusion. The outcome could be culture-dependent, which supports more comparative studies across different cultural or institutional contexts to optimize the STEM project communication. Similar to the highlight of

COVID-19's impact in this study, additional attention needs to be devoted to examining how the focus of Grad Slam projects evolves over time, particularly in response to global challenges such as climate change, pandemics, artificial intelligence, and other technological advances. By tackling these limitations and pursuing the outlined future research avenues, the PPOHA grant funding is expected to further strengthen its positive influences in improving the effectiveness of STEM education events like Grad Slam and, thus, ultimately bridge the gap between academic competency in specialized subjects and scientific communication to the public.

## References

- Aguilera, M. (2019). *A tale of tails wins Grad SLAM*. <https://today.ucsd.edu/story/a-tale-of-tails-wins-grad-slam>
- Amin, M., Shah, B., Sharif, A., Ali, T., Kim, K. I., & Anwar, S. (2022). Android malware detection through generative adversarial networks. *Transactions on Emerging Telecommunications Technologies*, 33(2), e3675.
- Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3-17.
- Bell, J. (2023). *Julia Mouat is named winner of the UC Davis Grad Slam competition*. <https://theaggie.org/2023/05/05/julia-mouat-is-named-winner-of-the-uc-davis-grad-slam-competition/>
- Benoit, K. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774.
- Best, J. W., & Kahn, J. V. (2010). *Research in education* (10 th ed.). PHL Learning Private Limited.
- Bishop, K. (2014). The importance of being an effective oral communicator. *Academic Leader*, 30(8), 1-2.
- Bracey, G. W. (2003). Teachers around the world. *Phi Delta Kappan*, 85(3), 253.
- Carter-Thomas, S., & Rowley-Jolivet, E. (2020). Three Minute Thesis presentations: Recontextualisation strategies in doctoral research. *Journal of English for Academic Purposes*, 48, 100897.
- Cohen, J. (2015). *The three-minute challenge*. <https://news.ucsb.edu/2015/015330/three-minute-challenge>
- CSUB (2023). *University catalog*. <https://www.csub.edu/catalog>
- Dake, D. K., & Gyimah, E. (2023). Using sentiment analysis to evaluate qualitative students' responses. *Education and Information Technologies*, 28(4), 4629-4647.
- Đurović, Z. (2023). Frequency or Keyness?. *Lexikos*, 33(1), 184-206. <https://doi.org/10.5788/33-1-1807>
- Excelencia in Education (2024). *HSIs database for researchers*. <https://www.edexcelencia.org/HSIs-database-for-researchers>
- Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(6), e1332.
- Gladstone, J. R., & Cimpian, A. (2021). Which role models are effective for which students? A systematic review and four recommendations for maximizing the effectiveness of role models in STEM. *International journal of STEM education*, 8(1), 1-20.
- Good, M., Masewicz, S., & Vogel, L. (2010). Bridging achievement and cultural gaps between schools and families. *Journal of Latinos and Education*, 9(4), 321-339.
- Grieve, R., Woodley, J., Hunt, S. E. & McKay, A. (2021). Student fears of oral presentations and public speaking in higher education: A qualitative survey. *Journal of Further and Higher Education*, 45(9), 1281-1293.
- Hyland, K. & Zou H. (2022). Pithy Persuasion: Engagement in 3 Minute Thesis presentations. *Applied Linguistics*, 43(1), 21-44.
- Hu, G., & Liu, Y. (2018). Three Minute Thesis presentations as an academic genre: A cross-disciplinary study of genre moves. *Journal of English for Academic Purposes*, 35, 16-30.

- Huang, C., Yang, C., Wang, S., Wu, W., Su, J., & Liang, C. (2020). Evolution of topics in education research: A systematic review using bibliometric analysis. *Educational Review*, 72(3), 281-297.
- Hutchinson, J. (2022). Australian and Canadian far-right extremism: a cross-national comparative analysis of social media mobilisation on Facebook (Doctoral dissertation, Macquarie University). [https://figshare.mq.edu.au/articles/thesis/Australian\\_and\\_Canadian\\_far-right\\_extremism\\_a\\_crossnational\\_comparative\\_analysis\\_of\\_social\\_media\\_mobilisation\\_on\\_Facebook/19428185/1/files/34520042.pdf](https://figshare.mq.edu.au/articles/thesis/Australian_and_Canadian_far-right_extremism_a_crossnational_comparative_analysis_of_social_media_mobilisation_on_Facebook/19428185/1/files/34520042.pdf)
- Kathpalia, S. S. (2024). Three-Minute Thesis presentations: simplification and engagement strategies. *ESP Today* 12(2), 196-220.
- Katre, P. D. (2019). NLP based text analytics and visualization of political speeches. *International Journal of Recent Technology and Engineering*, 8(3), 8574-8579.
- Kellogg, S. (2024). *Network inference & applications*. <https://laser-institute.github.io/network-analysis/module-4/sna-4-conceptual-overview.html#/a-quick-refresher>
- Koch, S., Wessels, M., Altpeter, B., Olvermann, M., & Johns, M. (2022). Keeping privacy labels honest. *Proceedings on Privacy Enhancing Technologies*, 4(2), 486-506.
- Kostelej, M., & Bagić Babac, M. (2022). Text analysis of the harry potter book series. *South Eastern European Journal of Communication*, 4(1), 17-30.
- Lall, H., Tapader, B. & Biswas, S. D. (2020). Speech anxiety: Improper communication & fear of public speaking. *International Journal of English Learning & Teaching Skills*, 1(4), 352-356.
- Lee, J., Solomon, M., Stead, T., Kwon, B., & Ganti, L. (2021). Impact of COVID-19 on the mental health of U.S. college students. *BMC Psychology*, 9(95).
- Lexalytics (no date). Machine Learning (ML) for Natural Language Processing (NLP). <https://www.lexalytics.com/blog/machine-learning-natural-language->
- Ma, Y., & Jiang, F. K. (2025). Guiding and engaging the audience: Visual metadiscourse in PowerPoint slides of Three Minute Thesis presentations. *English for Specific Purposes*, 77, 56-70.
- Morreale, S. P., Osborn, M. M., & Pearson, J. C. (2000). Why Communication is Important: A Rationale for the Centrality of the Study of Communication. *Journal of the Association for Communication Administration*, 29, 1-25.
- Mostafa, M. M., Feizollah, A., & Anuar, N. B. (2023). Fifteen years of YouTube scholarly research: knowledge structure, collaborative networks, and trending topics. *Multimedia Tools and Applications*, 82(8), 12423-12443.
- Muslimin, A. I. & Zaki, M. (2024). Navigating the visual odyssey: Perceptions, hurdles, and creative solutions in Three Minute Thesis presentations. *Issues in Educational Research*, 34(2), 588-607.
- Ranjan, B., & Mishra, B. (2022). Overview of big data and natural language processing. In *Handbook of Research for Big Data* (pp. 113-135). Palm Bay, FL: Apple Academic Press.
- Rao, P. & Taboada, M. (2021). Gender bias in the news: A scalable topic modelling and visualization framework. *Frontiers in Artificial Intelligence – Language and Computation* 4(82). doi: 10.3389/frai.2021.664737.
- Rodriguez, S. L., Perez, R. J., & Schulz, J. M. (2022). How STEM lab settings influence graduate school socialization and climate for students of color. *Journal of Diversity in Higher Education*, 15(1), 58-72.

- Sadlowski, E. (2018). *UC Davis Grad Slam competition features graduate research*. <https://www.proquest.com/docview/2115974539?parentSessionId=y9YT6NFV9Z0WMSuvHHfFGYfpUSC5Zf7bsmLGyuFU15s%3D&pq-origsite=primo&accountid=10345>
- Scharp, K. M. (2021). Thematic co-occurrence analysis: Advancing a theory and qualitative method to illuminate ambivalent experiences. *Journal of Communication*, 71(4), 545-571. doi: 10.1093/joc/jqab015.
- Stryker, C., & Holdsworth, J. (2024). *What is NLP?* [https://www.ibm.com/topics/natural-language-processing?utm\\_source=chatgpt.com](https://www.ibm.com/topics/natural-language-processing?utm_source=chatgpt.com)
- Svartzman, G. G., Ramirez-Marquez, J. E., & Barker, K. (2020). Social media analytics to connect system performability and quality of experience, with an application to Citibike. *Computers & Industrial Engineering*, 139, 106146.
- Benoit, K. (2018). quanteda. *Journal of Open Source Software*, 3(30), 774.
- Three Minute Thesis (no date). The University of Queensland. <https://threeminutethesis.uq.edu.au/higher-degrees-researchstart-your-3mt-journey-here>.
- Turner, L., & Restrepo, L. (2021). *Higher education performance and accountability*. <https://www.brookings.edu/wp-content/uploads/2021/06/Restrepo-Turner-higher-ed-accountability.pdf>
- Upadhye, A. (2020). A comprehensive survey of text data cleaning techniques: Challenges, methods, and best practices. *Journal of Scientific and Engineering Research*, 7(8), 205-210.
- Wang, J. (2023). *Strengthening STEM education for Latinx students. A peer-reviewed report for the federal PPOHA grant*.
- Wang, Y., Wang, J. T., Raymond, F. T., & Wang, J. T. (2023). Elementary pre-service teachers' horizon knowledge for teaching addition and subtraction: An analysis of video presentations. *EURASIA Journal of Mathematics, Science and Technology Education*, 19(6), em2276.
- Watanabe, K., & Müller, S. (2023). *Plot a network of feature co-occurrences*. [https://quanteda.io/reference/textplot\\_network.html](https://quanteda.io/reference/textplot_network.html)
- Weinberg, J. (2021). Who's listening to whom? The UK House of Lords and evidence-based policy-making on citizenship education. *Journal of Education Policy*, 36(4), 576-599.
- Zeletzny, L. (2022). *Promoting graduate education In Hispanics/Latinos/xs from Kindergarten*. <https://www.hispanicoutlook.com/articles/csu-bakersfield>