Racing Against the Clock: A Meta-Analysis to Investigate the Effect of Topical Area on Time to Publication in STEM Education Outlets

Dr. Brooke Coley, Founding Executive Director, RARE JUSTICE, Arizona State University

About the ARC Network

Funded by the National Science Foundation ADVANCE Program, Awards HRD-2121468 and HRD-1740860, the ADVANCE Resource and Coordination (ARC) Network seeks to achieve gender equity for faculty in higher education science, technology, engineering, and mathematics (STEM) disciplines. As the STEM equity brain trust, the ARC Network recognizes the achievements made so far while producing new perspectives, methods and interventions with an intersectional, intentional and inclusive lens. The leading champion in North America to propel the inclusion of women in the field of engineering, the Women in Engineering ProActive Network (WEPAN), serves as the backbone organization of the ARC Network.

About the Virtual Visiting Scholars

The Virtual Visiting Scholars (VVS) program provides a unique opportunity for select scholars across disciplines to pursue research meta-analysis, synthesis, and big data curation on topics crucial to STEM faculty equity. VVS analyze existing research and data, synthesizing different, sometimes competing, perspectives, frameworks, metrics, and outcomes to offer new insights and applications to the broader community.

About the Author

Brooke C. Coley, Ph.D. is Founding Executive Director of the Center for Research Advancing Racial Equity, Justice, and Sociotechnical Innovation Centered in Engineering (RARE JUSTICE)—an unprecedented testbed for innovating and modeling antiracist and equitable engineering futures—and Assistant Professor of Engineering, both at Arizona State University. Across several national projects funded primarily by the National Science Foundation, Dr. Coley's research lies at the intersection of racial equity, mental health and qualitative research methods encompassing critical theory, participatory action research, and arts-based research methods. Her work is anchored in an intentional amplification of the voices of minoritized populations in STEM with the goal of informing disruption of the pervasive systemic inequities found in racialized organizations such as institutions of higher learning. Leveraging the outcomes of this work, Dr. Coley will continue to create exemplars of equity in action across realms of the academic enterprise—lived experience and restorative justice, scholarship generation and metrics, and rewards systems and structures.

Executive Summary

A specific focus on equity related work in STEM as pursued through an investigation of the Journal of Engineering Education, as the flagship journal of the field of engineering education, evolved to be the focal point of this study. This work conducted a meta-analysis of peer-reviewed manuscripts spanning this journals body of work from 2019-2024. The work presented herein entails the story of n = 208 manuscripts representing the full continuum of published work across the engineering education research taxonomy. The original significance of this work was in the greater realm of faculty equity, and in considering metrics that contribute to faculty success, specifically.

The role of metrics is quite established in the tenure and promotion process with research, teaching and service serving as the primary realms for which individuals are evaluated. As it relates to research, publications play a significant role in the attainment of tenure and promotion to associate faculty ranks. Yet, to date, the role that publication metrics play in the tenure and promotion process remain understudied with the ways faculty are differentially impacted by publication metrics receiving even less attention (Patel et al., 2023; Kolesnikov et al., 2018; Furst-Holloway & Miner. 2019). This work investigates a specific nuance of publication in seeking to determine how topical area along the continuum of equity work influences timeline to publication.

Research Goal/Question(s) and Significance of the Problem

This critical investigation centers the experiences of Black women STEM equity scholars. Specifically, in a moment where the pursuit of diversity, equity, and inclusion work is being met with great resistance and/or penalty, the myriad manifestations through which minoritized scholars endure costs for this work remain under explored. This research investigates what might be inherent disadvantages in the publication process as related to the influence of topical area on the timeline to publication. Said differently, this work will elucidate whether topics associated with diversity, equity and inclusion are met with longer publication timelines. Bound in the last ten years of STEM education literature, from 2019 – 2024, this mixed-method meta-analysis addresses the research question, “How does pursuit of work, along the continuum of equity, impact time to publication in STEM education research outlets?” The racialized theory of organizations will be used as a framework to guide the interpretation of the results. Bound by the last 10 years, this work will take place across four phases: collect, code, analyze and report.

Theoretical Framework

Ray’s Theory of Racialized Organizations is being utilized as the theoretical framework underpinning this work (Ray, 2019). In this theory, four tenets are identified to describe common aspects demonstrating how institutions of higher learning show up as being racialized. Those tenets are: 1) whiteness as a credential; 2) the unequal distribution of resources; 3) racialized decoupling; and 4) diminished agency. The choice to anchor the work in this framing is an intentional one to best capture the nuance of how inequities are baked into our ways of doing across academia, and specifically, as it connects to metrics that position advancement and success for faculty.

Whiteness as a credential relates to the context of the journal, its editorial board, its reviewers, its primary audience, and the content of the body of work under its umbrella of dissemination. Another relevant manifestation of whiteness as a credential is the widespread utility of white ideologies and epistemologies in STEM education research (i.e., Tinto’s model of departure (Tinto, 1988)). Both serve to establish the norms across the various publication outlets while positioning work produced through different intellectual approaches to be viewed at a deficit. The unequal distribution of resources relates in the limited number of reviewers with expertise in the areas of diversity, equity, and inclusion which perpetuates overuse of the same experts to address these areas within a journal outlet. I apply this framing that situates the available pool of reviewers with expertise as resources to support my hypothesis of there being an association between reviewer availability,

reviewer expertise/familiarity (with the content being reviewed), and publication process speed. It is hopeful that some nuance of this will be implicit in the outcome data.

The tenet of racialized decoupling also establishes a critical aspect of the framing as it places a direct focus on how race is factored into the policy aspects of the publication process. Particularly, these include how race and its implications are being considered in the peer review process (e.g., the limited reviewer pools or reduced manuscripts from the DEI area compared to non-DEI manuscripts in print, etc.), if at all, and if not, reflecting on why they are not and who is being impacted (and how) by the lack of such consideration. Applying the framing of the last tenet, diminished agency, introduces the impact of the process on self-efficacy, morale, professional belonging, and potentially, perseverance and retention in the field. Leading a publication whose process spans years can be an intense experience for scholars and one impacting their perceptions of efficacy and belonging in their production of scholarship. Having this comprehensive framing enables the simultaneous considerations of nuance and context of racialized manifestations in addition to an empirical understanding of whether differences exist in the publication timeline across topical realms.

Methodological Approach, Methods Used and Data Extraction

Bound by the last 10 years, this work occurs across four phases: collect, code, analyze and report. Collect. In the first phase of the project, collect, the primary data sources supporting the analysis were determined. Beginning with a feeder list of DEI related journal articles spanning the last decade and then conducting searchers with various STEM education literature databases, a comprehensive list of STEM education journal outlets was compiled. Through this approach, the following journals were identified to be included in the analysis: Journal of Engineering Education (C), Journal of Women and Minorities in Science and Engineering (JWMSE), Studies in Engineering Education (SEE), Science Education, Computing in Science & Engineering, Engineering Studies (C), Journal for Research in Mathematics Education, International Journal of STEM Education (C), Journal of Engineering Education Transformations, International Journal of Engineering Education (IJEE), Journal of pre-college engineering education research (J-PEER). From the total 11 journals identified, only those indicated with a (C) were positioned to be included in the analysis, which will later be discussed in detail.

Data Collection and Analysis Process

For the purposes of establishing and refining the research design for this study, it was decided to first execute all phases of the project with one journal outlet. This would enable a smooth process for establishing all of the components of the process while enabling real-time troubleshooting as warranted through the adoption of natural language processing approaches. Additionally, this also supported the completion of the entire study design in avoiding the data management challenges that can often result with such a comprehensive dataset across sources. The Journal of Engineering Education was the outlet selected for this exercise.

This decision was made after careful consideration of the original journal list to be included in the analysis. After initiating the first steps of the collection process, it became clear that several of the articles failed to provide the information necessary to be included in the analysis. This in itself was the first finding of the study that will later be discussed as only three of the aforementioned journals provided the dates that manuscripts were originally submitted, revised, and accepted. This information is paramount in being able to determine the timeline to publication. Thus, only journals where this information could be extracted were included. The data presented in this report is representative of that from the Journal of Engineering Education. However, ongoing analysis will integrate data from Engineering Studies and the International Journal of STEM Education as available across the years 2019-2024.

An Excel workbook was created for the journal where each individual manuscript produced was listed explicitly across each of the respective volumes. Given the number of volumes and manuscripts per volume

varies with each journal, the number of manuscripts included in the analysis was variable for each journal. The Journal of Engineering Education publishes four volumes per year with an average of 10 research manuscripts per volume, generating a total of 40 manuscripts per year. Given the publication range of 2019-2024, this analysis includes n = 208 research manuscripts for the Journal of Engineering Education across the six-year period. For each research manuscript title, the following information was gathered and entered as headers across the columns of the sheet: manuscript title, first author, institution of the lead author, date submitted, date revised, date published, abstract, keywords.

This number is lower than the originally anticipated sample due to the change in format of the publication that occurred in 2019. While data is available from 2014-2024 to cover a decade of the journal, it was not until 2019 that the journal moved to a format where the dates the manuscript was originally submitted, revised, and accepted were included on the publication. Thus, I started with a host of data that, for this analysis approach, was unusable. In the continued analysis, the other manuscripts will be lumped in with the journal outlet adding in as a categorial variable that is not present in the initial analysis.

The volume of information that must be synthesized in the metadata set is large and benefits from a more automated way of extracting the data. To achieve this, R was used to develop and process a data extraction program (Abasi et al. 2020; Devlin, 2018; Hussain et al. 2019; Liu et al., 2020). The R program was designed to systematically gather publication dates associated with academic papers from two primary sources: PDF files and websites. For the analysis of the study data, all manuscripts were first saved as PDF files. Thus, no websites were used to extract dates. Gathering the publication dates from the PDF Files required text mining and corpus creation. The program initiates by conducting text mining on PDF files, each representing a publication from the journal being analyzed, here, the Journal of Engineering Education. The model creates a text corpus for each PDF to facilitate data extraction. This necessitates having access to a set of pdfs, which in the case of mining through all of the manuscripts of a given volume, and across a year, is not ideal. While this approach works well for a set of articles already saved and stored, it is a limitation to have to have all the pdfs for a given volume, or year, stored in one place and able to be recalled through the program. This also demonstrates how individuals' ability to conduct and/or replicate such studies is dependent upon access to journal volumes via pdf and/or online.

Nonetheless, a rule-based algorithm was developed to extract the important dates related to the publication process from each manuscript contained in the volume (Abasi et al., 2020). These include the submitted date, revised date, and accepted date. Given the variability in formatting and location of these dates across different journals, a customized algorithm is necessary to accurately identify and extract this information for each journal; however, starting with one journal to derive and deliver a functioning code leaves the changes required for customization for other journals to be minimal. The below function is an excerpt from the code extracting the received date from the Journal of Engineering Education:

Figure 1. Date extraction code excerpt from R.

```r
# Define a function to extract text between 'Received: ' and the next 'R'
extract <- function(string) {
  # Define the keyword to look for
  keyword <- "Received: "
  # Find the starting position of the keyword in the string
  pos_keyword <- regexpr(keyword, string)

  # Check if the keyword is found
  if (pos_keyword[1] != -1) {
    # Calculate the start position for the extraction
    start_pos <- pos_keyword[1] + nchar(keyword)
    # Get the substring from the start position to the end of the string
    rest_of_string <- substr(string, start_pos, nchar(string))

    # Find the position of the next 'R' in the substring
    pos_R <- regexpr("R", rest_of_string)

    # Check if 'R' is found; if not, return NA
    if (pos_R[1] == -1) {
      return(NA)
    }

    # Extract and return the text between 'Received: ' and the next 'R'
    return(substr(string, start_pos, start_pos + pos_R[1] - 2))
  } else {
    # Return NA if the keyword 'Received: ' is not found in the string
    return(NA)
  }
}
```

A similar process is repeated to also obtain the revise date as well as the published date of each manuscript. Once extracted, the dates from the PDFs were aggregated into a merged dataset for subsequent analysis. The dataset below shows an aggregation of the received date from a range of manuscripts, here labeled as Paper.ID, from the Journal of Engineering Education. This is how data was outputted from the model.

Table 1. Example date aggregation across Paper IDs 189 – 200.

	Paper.ID	Received_Date
189	189	8 March 2018
190	190	22 October 2018
191	191	8 January 2018
192	192	14 July 2017
193	193	17 May 2018
194	194	12 December 2018
195	195	14 July 2018
196	196	2 October 2018
197	197	28 September 2018
198	198	22 February 2018
199	199	19 July 2018
200	200	14 January 2019

Before continuing with the analysis, it is necessary to revisit the inability to use several of the original journal venues as planned. The primary issue as it related to the set of 11 journals was the variability in ways publication timeline related information was listed, if provided. As in the case with the Journal of Engineering Education, the publication information was changed to be contained within the manuscript starting in 2019. The journals whose publication timelines provided all of the pertinent information to be included in the analysis were designated above with a (C) in figure 1 for "coded." However, there were several journals that did not provide this information, and this introduces a concern for policy moving forward. Making the dates publications were first submitted, revised and accepted provide critical information to potential authors considering submitting to the journal in being able to determine what that process looks like for related work that has been published in the particular journal. When such information is not available, this simple, and what would seemingly feel to be public information, is not available and positions the journal to have limited transparency in its publication timeframes. This transparency will be summarized in the discussion.

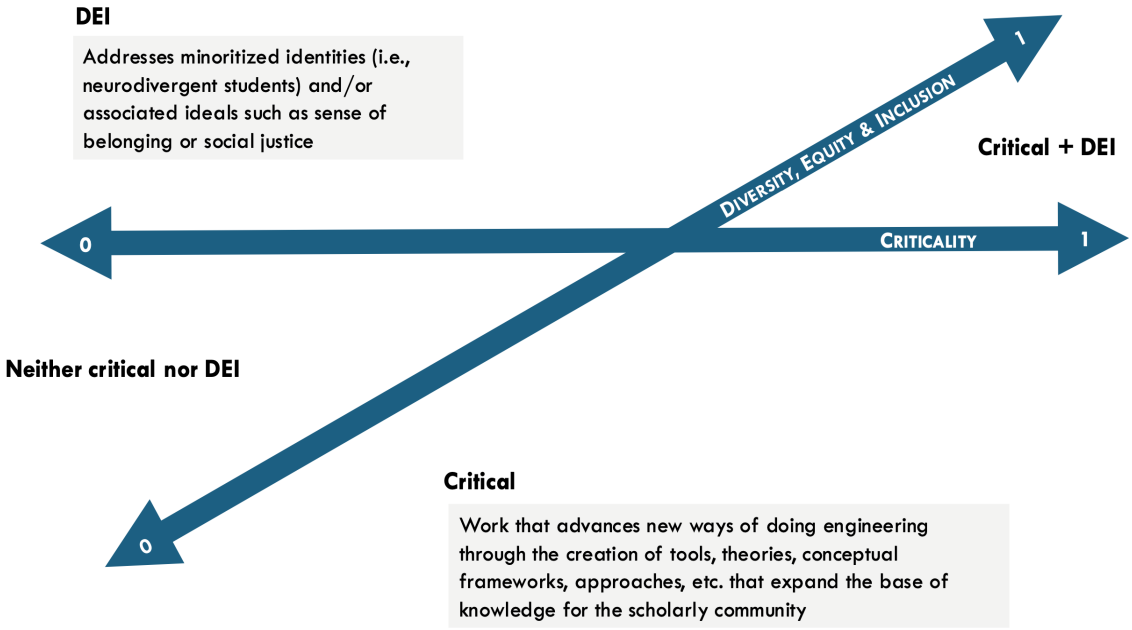
In R, a dataset was created containing all of the necessary dates for each of the manuscripts included in the analysis. This dataset contained volume 108 through 113 from 2019 to 2024 of the Journal of Engineering Education. There remains one outstanding issue #4 to be published to complete the data of 2024. However, the dataset that was created was saved as a .csv file and exported to Excel. In Excel, calculations were performed to determine the time to publication for each of the manuscripts. Having the publication timelines for every manuscript in the time window, the next step of the analysis involved investigating the actual effect of the topical area on time to publication. This would first necessitate the establishment of an association with topical area along the continuum of equity.

Upon completion of the date extraction program, content analysis was employed to scrutinize the titles of academic papers to identify those focusing on diversity, equity, and inclusion (DEI) related issues. For the greater meta-analysis, a sophisticated text classification algorithm to efficiently automate the identification process was leveraged. Specifically, utilizing advanced natural language processing (NLP) techniques, the program will implement a text classification algorithm with Bidirectional Encoder Representations from Transformers (BERT)-class models. This algorithm is designed to analyze the textual content of each paper's title and abstract, determining the presence of themes relevant to DEI studies. This is one of the areas where the

subset of papers being coded manually will serve to ensure the accuracy and reliability of the algorithm. Each paper will be assigned a binary code: '0' indicating the paper does not focus on DEI issues, and '1' signifying it does. This manual coding process will generate a high-quality training dataset for the algorithm and is currently in process.

It was important to me to capture a range of work along the continuum of equity for categorizing titles. This is a critical step to the process and involves a significant reliance on positioning of the researcher. After several iterations of visioning a taxonomy for the continuum of equity, I landed on a two-plane categorization. Figure 2 depicts the two-plane categorization of equity along the axes of criticality and address of diversity, equity, and inclusion. Criticality was defined as manuscripts that advanced new knowledge in the provision of theoretical/conceptual frameworks, methods, epistemologies and/or approaches.

Figure 2. Continuum of equity along considerations of criticality and association with diversity, equity and inclusion.



While all scholars would argue that their work advances new knowledge, to be indicated as critical meant the manuscript contributed a new knowledge and tools that could readily be adopted by the greater scholarly community. Specifically, manuscripts expanding our epistemic ways of doing were classified as critical. The other axes of the taxonomy identified manuscripts that were associated with diversity, equity, and inclusion. This was determined by the presence of words affiliated with a specific minoritized identity group (e.g., gender, women, Blacks, Latinx, persons with disabilities, first-generation, socioeconomic status, transfer students, veterans, etc.) as well as ideals and constructs related under the umbrella of equity such as social justice, public welfare, and belonging. This resulted in each manuscript being categorized by a double binary measure where manuscripts were found to be either critical (1-0), DEI (0-1), critical x DEI (1-1), or not critical or DEI (0-0). To the best of my knowledge, this is the first attempt to bucket manuscript titles topically along a spectrum simultaneously considering equity and DEI.

Results

The average time to publication for the journal across this six-year time span was 487 ± 210 days. This standard deviation, equating roughly to .6 years, demonstrates the level of variability in the data. Table 2 provides the categorical outcomes for each of the four yielded topical classification buckets.

Table 2. Average publication times for taxonomy categories.

Taxonomy Category	Binary Coding	N = observations (manuscripts) per category	Average ± Stdev
Critical	1,0	35	533.2 ± 284
DEI	0,1	35	478 ± 173
Critical + DEI	1,1	19	556.4 ± 268
Neither Critical nor DEI	0,0	119	465.0 ± 181

Additionally, to test whether there was significance in the timelines associated with publication across these taxonomy categories, an ANOVA was chosen as a special form of regression analysis to treat this data. The results of the ANOVA are pictured in Table 3.

Table 3. ANOVA output results.

```
Residuals:
    Min      1Q  Median      3Q     Max
-395.20 -144.60  -29.35   96.79 1002.51

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   464.35      18.44  25.181   <2e-16 ***
Critical       71.14      33.27   2.138   0.0337 *
DEI            16.71      33.27   0.502   0.6160
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 208.7 on 205 degrees of freedom
Multiple R-squared:  0.02461,   Adjusted R-squared:  0.01509
F-statistic: 2.586 on 2 and 205 DF,  p-value: 0.07776
```

A one-way ANOVA was performed to compare the effect of taxonomy category on timeline to publication. The one-way ANOVA revealed that there was not a statistically significant difference in mean timeline to publication between at least two groups (F (2,205) = [2.586], p = 0.07). However, the difference in mean time to publication was statistically different between critical and not critical nor DEI papers (p=.0337). No significant difference was found between DEI papers and not critical nor DEI papers (p=.616). These findings suggest that while statistical significance was not obtained, there are trends that significance might be obtained with greater power. Specifically, the not critical nor DEI category had n = 119 observations compared to only n = 35 in the critical and DEI categories and n = 19 in the critical and DEI category.

Looking at the means and standard deviations across the three categories, it becomes necessary to know what occurred for those manuscripts at the intersection of critical and DEI. Thus, a test of the interaction between critical and DEI was run to observe the outcome. While no significance was obtained as shown in

Table 4, the manuscripts categorized as ‘critical + DEI’ had the greatest mean across the groups and also with high variability.

Table 4. Test of the intersection of critical + DEI.

```
Coefficients:
 (Intercept)      Critical           DEI  Critical:DEI
     465.025        68.175        13.746         9.475
```

This suggests that with greater power, or more observations in this group, the results might establish significance. Although significance was not obtained in this analysis, this remains a critical finding in that the manuscripts at the intersection of criticality and association with DEI took the longest to publish. Table 5 provides more context into the titles of the manuscript along with their time to publication that were determined to populate the ‘critical + DEI’ category.

Table 5. Manuscripts at the intersection of critical and DEI with time to publication.

Manuscript Title	Time to Publication
Examining interests and goals as predictors of gender differences in engineers' pursuit of managerial roles	278
Engineering ableism: The exclusion and devaluation of engineering students and professionals with physical disabilities and chronic and mental illness	169
"The lab isn't life": Black engineering graduate students reprioritize values at the intersection of two pandemics	812
The fallacy of "there are no candidates": Institutional pathways of Black/African American and Hispanic/Latino doctorate earners	267
Impact of COVID-19 on sense of belonging: Experiences of engineering students, faculty, and staff at Historically Black Colleges and Universities (HBCUs)	696
Investigating culturally contextualized making with the Navajo Nation	701
How Latiné engineering students resist White male engineering culture: A multi-institution analysis of academic engagement	927
Circle of success—An interpretative phenomenological analysis of how Black engineering students experience success	529
Engineering stress culture: Relationships among mental health, engineering identity, and sense of inclusion	365
Learning from the experiences of Navajo engineers: Looking toward the development of a culturally responsive engineering curriculum	1077
The correlation between undergraduate student diversity and the representation of women of color faculty in engineering	357
Comparing students' engineering and science aspirations from age 10 to 16: Investigating the role of gender, ethnicity, cultural capital, and attitudinal factors	597
Engineering political fluency: Identifying tensions in the political identity development of engineering majors	761
Learning from small numbers: Studying ruling relations that gender and race the structure of U.S. engineering education	157
Resilient engineering identity development critical to prolonged engagement of Black women in engineering	632
Positionality practices and dimensions of impact on equity research: A collaborative inquiry and call to the community	271
Divergent thinking and academic performance of students with attention deficit hyperactivity disorder characteristics in engineering	830
"They are here to support me": Community cultural wealth assets and precollege experiences of undergraduate Black men in engineering	673
Recognizing the funds of knowledge of first-generation college students in engineering: An instrument development	473

Discussion, Implications and Recommendations

This work was designed to examine the relationships between critical publication metrics and a focus on critical and/or DEI issues within academic research. Specifically, this meta-analysis was conducted to determine whether there was a correlation between topical area and time to publication in STEM education related journal outlets. This initial investigation focused on one journal outlet, the Journal of Engineering Education, spanning the years 2019 to 2024. The findings of this study showed a correlation to exist with manuscripts at the intersection of being critical and associated with DEI to have longer timelines to publication.

The implications of this work are far-reaching and of importance to various stakeholders in faculty success such as Deans, department chairs, and others involved in faculty-related decision making (i.e., the determination of metrics used to evaluate faculty success). Having empirical evidence of time to publication being related to topical area suggests an inequity in the time constant of peer review. Thus, a major implication of this work is the apparent disadvantage imposed for the pursuit of work at the intersection of criticality and association with DEI. For those pursuing work along the continuum of equity, the nature of their work stands to delay publications time on the order of magnitude in years. When considering the push to promote more successful tenure outcomes for STEM faculty, this work urges a need for intentionality in considering how inequities might impact the acquisition of essential metrics. Given that a significant portion of the tenure and promotion process is situated around the successful publication of a specified number of peer-reviewed manuscripts, with each institution, discipline, and department/school potentially having variation in their exact criteria, it cannot be ignored that based on topic one faculty member's work might move seamlessly through a process that might be significantly delayed for another faculty member simply based upon their topic. Based on the outcomes of this work, faculty A and faculty B of the same program could have a paper published in 149 and 1,538 days, respectively, as influenced by where their work fell along the continuum.

If institutions truly desire to make the tenure and promotion process more equitable for all faculty, they must be more intentional to account for such inequities in the establishment of critical metrics. One recommendation would be for institutions to conduct disciplinary inventories, leveraging annual data reported by faculty, to determine average rates of production within a given program or department. This would enable the variability within that disciplinary unit to be maintained and not cancelled in a larger pool of merged disciplinary norms. Then, knowing the outcomes of such averages and standard deviations, metrics more relevant to the disciplinary trends of that group could be established.

The other recommendation would be to mandate greater transparency from journal outlets, and specifically, provision of the details necessary to determine publication timelines for all manuscripts published in their journal. The main challenge identified in the work was the lack of transparency regarding publication details across journal outlets, and specifically, those disseminating STEM education research. Across the various journals anticipated to be included in the study, some reported all of the data required to determine publication timelines (date submitted, date revised, date accepted) while others provided only the publication date. This was indeed a late discovery as I was not privy to the fact that such information was not consistently and necessarily provided across publication venues. The lack of a consistent availability of this information signaled a greater disconnect and warranted the work as access to this important information has implications for authors and scholarly equity. Many scholars utilize information regarding time to publication as a means for determining where to submit their work. Being able to determine the timeframe for related work has proven to be insightful for scholars in informing that decision. However, for the eleven journals originally anticipated for inclusion in this work, only three transparently shared the information necessary for calculating individual manuscript publication timelines.

Limitations

There are two important limitations to consider. One limitation is the current approach only investigates papers that were actually published. Therefore, this work does not take into account manuscripts that were: 1) submitted, rejected and not further pursued; 2) submitted, revised and not finalized; and/or 3) initiated in the studied journal but ultimately published in a different journal outlet. Having this information would also be beneficial as in the current dataset all that we learn from are those manuscripts that prevailed the publication process in the given journal. Another limitation to the study is it does not contextualize information pertinent to the journal such as editorial team positioning. Having the ability to make connections between the journal and its manuscript outcomes beyond the variable of time would be insightful.

Future Work

This work turned out to be an iceberg project meaning it initiated the investigation only to uncover how much there is within this realm to be studied. Essentially, this study because a meta-analysis of the Journal of Engineering Education spanning years 2019-2024. However, the continuation of this work will expand to include three additional journals: International Journal of STEM Education, Engineering Studies, IEEE Transactions on Education. Additionally, now having greater power (n . 1000 manuscripts) the researcher-derived taxonomy will serve as training for the NLP models and used to automate the coding process that was done manually in the pilot. Specifically, the manually coded data will serve multiple roles in the development of the text classification algorithm, including:

- Training: The algorithm will learn from the human-coded dataset to identify patterns and keywords associated with DEI-related research.
- Validation: A portion of the manually coded dataset will be reserved for validating the algorithm's performance, ensuring it accurately classifies papers based on their placement along the continuum of equity.
- Testing: After training and validation, the algorithm will be further tested on unseen data to assess its generalizability and robustness in accurately identifying research along the continuum of equity.

The successful implementation of the text classification algorithm will enable the automated identification of papers, streamlining the content analysis process. The analysis will result in a curated dataset of academic papers categorized by their relevance along the continuum of equity. This dataset will be invaluable for researchers interested in DEI studies, policy-making, and educational program development. In the subsequent work, I will also analyze abstracts to determine if the content analysis better defines the placement of the manuscript along the continuum of equity. I will also be able to determine if greater context regarding the manuscripts (i.e., title information vs. information of the abstract) yields different outcomes in manuscripts assigned to the taxonomy categories and the associated times to publication of the categories.

References

Habersang, S., Küberling-Jost, J., Reihlen, M., & Seckler, C. (2019). A process perspective on organizational failure: a qualitative meta-analysis. Journal of Management Studies, 56(1), 19-56.

McGee, M., Patel, R., Smith, R., & Ajmera, S. (2023). The Impact of the COVID-19 Pandemic on Faculty Productivity and Gender Inequalities in STEM Disciplines. SMU Data Science Review, 7(3), 2.

Kolesnikov, S., Fukumoto, E., & Bozeman, B. (2018). Researchers' risk-smoothing publication strategies: Is productivity the enemy of impact?. Scientometrics, 116, 1995-2017.

Furst-Holloway, S., & Miner, K. (2019). ADVANCEing women faculty in STEM: empirical findings and practical recommendations from National Science Foundation ADVANCE institutions. Equality, Diversity and Inclusion: An International Journal, 38(2), 122-130.

Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. The journal of higher education, 59(4), 438-455.

Raja, M. P., & Abirami, A. M. (2023). Strategic approaches for DEI Implementation in Engineering Education. Journal of Engineering Education Transformations, 36 (Special Issue 2).

Bruning, M. J., Bystydzienski, J., & Eisenhart, M. (2015). Intersectionality as a framework for understanding diverse young women's commitment to engineering. Journal of Women and Minorities in Science and Engineering, 21(1).

Sochacka, N. W., Walther, J., Rich, J. R., & Brewer, M. A. (2021). A narrative analysis of stories told about engineering in the public discourse: Implications for equity and inclusion in engineering. Studies in Engineering Education, 2(2).

Henderson, J. A., Hines, E. M., Boyce, A., Golden, M., Singleton II, P., Davis, J. L., ... & Junqueira, W. (2022). Factors impacting engineering advanced degree pursuit and attainment among black males. Journal of Women and Minorities in Science and Engineering, 28(4).

Henderson, J. A., Junqueira, W., Benjamin, L. S. S., Hines, E. M., Alarcón, J. D., Davis, J. L., & Cavazos, S. (2023). Circle of success—An interpretative phenomenological analysis of how Black engineering students experience success. Journal of Engineering Education.

Holly Jr, J., & Lee, W. C. (2024). Black Men Committed to Making Engineering More Humane: A Collaborative Autoethnography of Two Engineering Education Scholars. Journal of Women and Minorities in Science and Engineering, 30(1).

Morton, T. R., & Parsons, E. C. (2018). # BlackGirlMagic: The identity conceptualization of Black women in undergraduate STEM education. Science Education, 102(6), 1363-1393.

Mack, K., Johnson, L. R., Woodson, K. M., Henkin, A., & Dee, J. R. (2010). Empowering women faculty in STEM fields: An examination of historically black colleges and universities. Journal of Women and Minorities in Science and Engineering, 16(4).

Allen, D., Dancy, M., Stearns, E., Mickelson, R., & Bottia, M. (2022). Racism, sexism and disconnection: Contrasting experiences of Black women in STEM before and after transfer from community college. International Journal of STEM Education, 9(1), 20.

Leyva, L. A. (2021). Black women's counter-stories of resilience and within-group tensions in the white, patriarchal space of mathematics education. Journal for Research in Mathematics Education, 52(2), 117-151.

Carrigan, C., Tanguay, S. K., Yen, J., Ivy, J. S., Margherio, C., Horner-Devine, M. C., ... & Grant, C. S. (2023). Negotiating boundaries: an intersectional collaboration to advance women academics in engineering. Engineering Studies, 15(1), 9-29.

Coley, B., & Thomas, K. (2023). "The lab isn't life": Black engineering graduate students reprioritize values at the intersection of two pandemics. Journal of Engineering Education, 112(2), 542– 564. https://doi.org/10.1002/jee.20518

Jauchen, J. G. (2023). Institutional Activism in Diversity, Equity, and Inclusion Faculty Service in STEM. Journal of Women and Minorities in Science and Engineering, 29(2).

Hutchins, H. M., Pereira-Leon, M. J., & Lee, M. M. (2022). Can a Workshop Make a Difference? Influencing Promotion-Seeking Behaviors of Women Associate Professors. ADVANCE Journal, 3(2).

Robinson, W. H., McGee, E. O., Bentley, L. C., Houston, S. L., & Botchway, P. K. (2016). Addressing negative racial and gendered experiences that discourage academic careers in engineering. Computing in Science & Engineering, 18(2), 29-39.

McGee, E. O. (2015). Robust and fragile mathematical identities: A framework for exploring racialized experiences and high achievement among black college students. Journal for Research in Mathematics Education, 46(5), 599-625.

McGee, E. O., Main, J. B., Miles, M. L., & Cox, M. F. (2021). An intersectional approach to investigating persistence among women of color tenure-track engineering faculty. Journal of Women and Minorities in Science and Engineering, 27(1).

Abasi, A. K., Khader, A. T., Al-Betar, M. A., Naim, S., Alyasseri, Z. A. A., & Makhadmeh, S. N. (2020). A novel hybrid multi-verse optimizer with K-means for text documents clustering. Neural Computing and Applications, 32, 17703-17729.

Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Hussain, S. F., & Haris, M. (2019). A k-means based co-clustering (kCC) algorithm for sparse, high dimensional data. Expert Systems with Applications, 118, 20-34.

Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., & Wang, P. (2020, April). K-bert: Enabling language representation with knowledge graph. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 03, pp. 2901-2908).