2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #33705

How Can We Identify Teams at Risk of Marginalizing Minoritized Students, at Scale?

Dr. Darryl A. Dickerson, Florida International University
Stephanie Masta, Purdue University, West Lafayette
Dr. Matthew W. Ohland, Purdue University, West Lafayette

Matthew W. Ohland is Associate Head and the Dale and Suzi Gallagher of Professor of Engineering Education at Purdue University. He has degrees from Swarthmore College, Rensselaer Polytechnic Institute, and the University of Florida. His research on the longitudinal study of engineering students, team assignment, peer evaluation, and active and collaborative teaching methods has been supported by the National Science Foundation and the Sloan Foundation and his team received for the best paper published in the Journal of Engineering Education in 2008, 2011, and 2019 and from the IEEE Transactions on Education in 2011 and 2015. Dr. Ohland is an ABET Program Evaluator for ASEE. He was the 2002–2006 President of Tau Beta Pi and is a Fellow of the ASEE, IEEE, and AAAS.

Dr. Alice L. Pawley, Purdue University, West Lafayette

Alice Pawley (she, her, hers) is an Associate Professor in the School of Engineering Education and an affiliate faculty member in the Women's, Gender, and Sexuality Studies Program, Environmental and Ecological Engineering, and the Purdue Climate Change Research Center at Purdue University. She was co-PI of Purdue's ADVANCE program from 2008-2014, focusing on the underrepresentation of women in STEM faculty positions. She runs the Feminist Research in Engineering Education Group, whose diverse projects and group members are described at pawleyresearch.org. She was a National Academy of Engineering CASEE Fellow in 2007, received a CAREER award in 2010 and a PECASE award in 2012 for her project researching the stories of undergraduate engineering women and men of color and white women, and received the Denice Denton Emerging Leader award from the Anita Borg Institute in 2013. She has been author or co-author on papers receiving ASEE-ERM's best paper award, the AAEE Best Paper Award, the Benjamin Dasher award, and co-authored the paper nominated by the ASEE Committee on Diversity, Equity, and Inclusion for ASEE Best PIC Paper for 2018. More recently, she received her school's Award for Excellence in Mentoring, the Award for Leadership, and a 2019 award from the College of Engineering as an Outstanding Faculty Mentor of Engineering Graduate Students. In 2020 she won the Sterling Olmsted Award from the Liberal Education/Engineering and Society Division of ASEE. She is president of Purdue's chapter of the American Association of University Professors (2020-22). She helped found, fund, and grow the PEER Collaborative, a peer mentoring group of early career and recently tenured faculty and research staff primarily evaluated based on their engineering education research productivity. She can be contacted by email at apawley@purdue.edu.

How can we identify teams at risk of marginalizing minoritized students, at scale?

Abstract

Teamwork is critical to engineering professional work. While some aspects of teaming with engineering students are well understood and implemented into instructional tools, tools for handling student teams dealing with implicit and explicit racism, sexism, and homophobia are infrequent. Instructors of large undergraduate courses need tools to help make team-level marginalization visible at the classroom level to interrupt discriminatory or marginalizing behavior amongst teammates, and to model allyship so teammates learn how to interrupt others' marginalizing behavior when instructors are not around. This paper describes the broader project, and describes some early results, focused on an algorithm that can help identify teams engaging in marginalizing behaviors against minoritized students, whether minoritized by race, gender, nationality, LGBTQ identity, or other categorization schemes. We describe how the algorithm is proving useful to identify student teams to focus on for analysis to answer some of our research questions focused on how engineering undergraduate teams marginalize minoritized members, and illustrate one such analysis. We describe our continuing work on the broader project.

Introduction

Teamwork is critical to engineering professional work, and engineering students can be taught to do it well. While some aspects of teaming with engineering students are well understood and implemented into instructional tools, tools for managing implicit and explicit racism, sexism, and homophobia in teams are still poorly resourced. Instructors of large undergraduate courses need tools to help make team-level marginalization visible at the classroom level to interrupt discriminatory or marginalizing behavior amongst teammates, and to model allyship so teammates learn how to interrupt others' marginalizing behavior when instructors are not around. Robust, valid, and accessible tools that help instructors see where to intervene will help create a more inclusive learning environment for minoritized students, a critical need for broadening participation in engineering. Additionally, such tools will help engineering programs to meet the new ABET criterion 4, where we must ensure that students have "an ability to recognize ethical and professional responsibilities in engineering situations and make informed judgements." Helping students learn to manage team dynamics of majority members to facilitate everyone's ability to contribute to the level of their talent is an ethical and professional responsibility to the field.

This paper shares some early results from our broader NSF-funded project, titled "Identifying Marginalization and Allying Tendencies to Transform Engineering Relationships," or I-MATTER. The project's research questions are:

1. What does marginalization look like within engineering classrooms where teamwork is a primary feature?
2. How is marginalization legible (or not) to instructors at the classroom level?
3. What are the different ways that instructors respond to incidents of peer-to-peer marginalization?

4. How might the lessons of this work be implemented to systematically alert instructors when marginalization is likely occurring?

This paper focused on answering the first question using historical and contemporary peer evaluation data collected through CATME, a web-hosted instrument that measures behaviors necessary for effective team functioning (www.catme.org). In the broader project, we draw on theoretical frameworks framing gender and race as social constructions with real consequences, that operationalize microaggressions and coded language, and are rooted in commitments from critical race theory. We are collecting data at a Midwestern, historically and predominantly white institution with a large, yet innovative, first-year engineering program.

In this paper, we share our CATME algorithm that can help identify teams engaging in marginalizing behaviors against minoritized students, whether minoritized by race, gender, nationality, LGBTQ identity, or other categorization schemes. This paper will describe the algorithm, and demonstrate how it is proving useful in identifying student teams to focus on for analysis to answer our first and second research questions.

Background

This project is situated at the intersection of research on team dynamics (both in engineering and outside of it), on broadening participation in engineering, and on theory on marginalization and microaggressions.

Teamwork is a critical part of modern engineering education. Teamwork skills are necessary for the engineering workplace where teaming is pervasive, and are important for developing other work skills, including leadership, project management, communication, and conflict resolution.

In many educational contexts, we know that teammates marginalize their minoritized peers on the team; common instructional practices that organize teamwork may exacerbate minoritized teammates' marginalization. Marginalization can begin with team formation, where self-selection practices result in less gender- and racially-diverse teams [1]. Team formation by self-selection is still pervasive despite overwhelming evidence of how problematic it is [2]-[6]. In cases where team formation is controlled by instructors, instructors should carefully consider race, gender, and other salient characteristics that may serve as markers of difference within teams in the execution of team-based learning. The consequences of ignoring race and gender dynamics include limited learning opportunities for all students, experiences of isolation of marginalized students, and incidents of racism and sexism that can lead to students withdrawing from the team, leaving the course, or even leaving their major [7].

Minoritized teammates can be marginalized through microaggressions. Microaggressions are brief and commonplace verbal, behavioral, and environmental degradations with detrimental cumulative effects on marginalized people [8]. Three primary types of microaggressions that occur in classrooms are *microinsults*, *microinvalidations*, and *microassaults*. *Microinsults* are verbal and behavioral expressions that demean a person's identity (e.g., telling Black students they sound so articulate). *Microinvalidations* occur when the lived experiences of marginalized peoples are negated, invalidated, or diminished (e.g., claiming to be "colorblind"). *Microassaults* are more overt, constituting verbal and non-verbal attacks and avoidant behaviors (e.g not wanting to sit next to a student of color).

While researchers have investigated microaggressions that women and racialized minorities experience, there remain critical gaps in understanding the effects of marginalization on engineering student team dynamics associated with gender, race, and sexual orientation. Both social role theory and social identity theory have been used to explore psychological underpinnings of persistent sexism in workspaces [9-12]. These theories provide complementary views on gender bias in small groups generally, and more specifically in male-dominated environments [7], [13]-[14]. There is extensive research in organizational psychology on marginalization of racial minorities as a persistent characteristic of the workplace [15]-[16], and on different forms of marginalization of lesbian, gay, bisexual, transgender, and queer (LGBTQ) people in the workplace [17]. In STEM collegiate educational environments, researchers have documented the effects of marginalization on racial minorities, resulting in their higher psychological distress and decreased math self-concept [18] compared to their white peers. The scant body of research on sexual minorities in engineering has revealed LGBTQ students' social isolation, professional devaluation, and decreased physical and mental well-being compared to their heterosexual and gender-binary peers [19]-[20].

Scholars have begun to explore the presence of gender-based marginalization within engineering student teams specifically [21]-[23], which manifests in a variety of ways, including assignment to gendered and non-technical roles, lower active participation, lower visible involvement, devaluation of educational credentials, and discounting of expertise [21]-[22], [24]-[26]. Additionally, current research provides evidence of the negative aspects of teamwork for racially minoritized students, with students reporting domineering teammates, limitations in learning opportunities, and exclusion from team roles [27]. Further qualitative inquiry revealed sexism and racism as themes within team interactions, and which provoked limited remedies from instructor intervention. The major takeaways for students were that bad team experiences were inevitable and working alone was preferable, counter to our expectations of students in their undergraduate education and in the workforce. Only recently has research on the engineering classroom experience of LGBTQ students emerged (e.g., [19]), and little of that research studies the team dynamics those students experience.

Literature on team dynamics in undergraduate education has focused on a variety of psychological constructs, but it is limited with respect to race and gender's effect in and on peer evaluations. Team composition can influence team effectiveness directly, such as when a team has members with strong skills and abilities to perform the team task. Team composition can also influence team performance indirectly by creating a negative environment [28]-[30].

In the context of this work, teams experiencing high levels of microaggression are likely to have higher conflict and lower cohesion. Conflict has components of task conflict, relationship conflict, and process conflict [31]. Marginalization, including microaggressions, seem most likely to affect relationship conflict, but to the extent that the contributions of some students are marginalized, task conflict and process conflict will also increase. Similarly, microaggression might also be measured through the construct of psychological safety [32].

CATME employs these psychological constructs as a component of team formation and teammate evaluation. The constructs are valuable in the ability to overcome self-selection biases, and to help prevent isolation by a salient demographic characteristic such as race or gender. However, research on the effect of race and gender on peer evaluations of teammates is particularly limited, and much of that research is inconclusive [23].

Current practice hinders the use of team process outcomes to detect bias and the occurrence of marginalization, including microaggressions, in two ways. First, these measures are commonly taken as consensus measures [33], in which responses are averaged over each team to measure an "average" level of conflict or psychological safety. This makes the experience of individual students invisible, because if one team member is marginalized, it is still possible for the average to be above the level of concern. Guidelines exist for determining whether such aggregation is appropriate [34], but when consensus breaks down, the data are typically discarded because the findings do not fit the theoretical assumptions under which those constructs were developed. By measuring these constructs as team processes, a second shortcoming arises. There is evidence that although team-level measures are useful when member agreement is considered, those processes also include a significant contribution from dyadic interaction [35]. A team member may report high conflict or low psychological safety, but their perception may be significantly influenced by one or more specific individuals. One goal of this work is to describe the microaggression landscape to better frame the process of detecting microaggressions using these team process outcomes.

Theory on coded language offers us one opportunity to assess marginalization in teams. Much of the research on coded language focuses on the way coded language reinforces racism and white supremacy. So even though race and racism are pervasive, there are a number of "code words" where we talk about race without naming race. It is far more normal to see words such as "urban," inner city," and "disadvantaged" than to see "white," "overadvantaged," or "privileged." Coded language reproduces racist images and perspectives at the same time as reproducing the illusion that race is something other people have. Teachers and students who use coded language in the classroom are acting in ways consistent with patterns present outside of schools. Using coded language is problematic for two reasons: 1) it hides the reproductive practices people engage in when related to race and inequity and 2) it prevents educators from engaging in conversations about the structural nature of racism. Given this, it's important to study the way teachers and students talk to each other and how they employ the use of coded language and how that affects non-white students in the classroom. Research on coded language often occurs in the context of K-12 education. However, understanding how teammates employ coded language with each other in the classroom setting and in teammate evaluations could provide insight a timely mechanism to detect marginalization and alert instructors to act. This research intends to analyze qualitative data from CATME for incidents of coded language as part of describing the microaggression landscape in engineering student teams.

Our research works towards filling the hole left by existing research and interventions. The extensive use of CATME in this university's First-Year Engineering classes creates an opportunity to observe and describe marginalization in engineering student teams. Those observations will be based on both quantitative and qualitative data collected through CATME's 5-factor model: contributing to the team's work; interacting with teammates; keeping the team on track; expecting quality; and having relevant knowledge, skills, and abilities.

Theoretical framework

To identify the relationship between student comments and the CATME algorithm, we draw from Critical Race Theory [36], which holds as core tenets the idea that racism is a normal part of American life, racism is structural and systemic, and that intersectionality is critical in understanding lived experiences. We also Sue's [8] framing of microaggressions as one form of daily harassment experienced by minoritized folks, and Cortina's selective incivility theory [15]-[16] which argues that subtle actions of interpersonal mistreatment directed towards particular social group are often masked.

Methods

For this part of our research, we downloaded numerical ratings and associated comments from three academic years of CATME peer-to-peer evaluations collected from a large first-year engineering course taught at a Midwestern public research university. The analysis performed in this paper focuses on one spring semester of the course. This required course is taught to over 2000 students per year, in sections of over 100 students that are then organized into teams with a target size of 4 students each. We designed, then applied, an algorithm based on CATME measures that identified teams that may demonstrate marginalizing behavior based on the team process measurement scales. CATME collects both quantitative measures and qualitative comments between teammates on a 4-person team, and the rater also must assess their own performance along with evaluating that of their teammates.

In this section, we describe the algorithm, and then the analysis we applied to cases the algorithm identified as high priority.

Algorithm

We designed a process to use CATME historical data to prioritize cases for qualitative review. The steps are:

1. Students were selected on the basis of membership in a group likely to be marginalized: women; Black, Native American, or Hispanic students; students reporting "Other or prefer not to answer" to a gender question; or international students.
2. We extracted quantitative team process measures on psychological safety, conflict, and satisfaction (defined in Step 5) from multiple academic terms of student data for this specific course. We did this on a large scale from the CATME database in deidentified form.
3. We matched these team process measures with the associated qualitative comments (which were not part of the deidentified CATME data). This required downloading ratings, team process measures, and comments for each peer evaluation (4 per semester) for each term separately as PDFs.
4. We associated the team process measures (which were only identified with an internal CATME PersonID) with semester/section/team/peer evaluation/name/email address in the PDFs. This was accomplished by accessing the source html for student tables in CATME, which include identifiable information (name, email address) alongside CATME's PersonID, which is used to track student data in deidentified form. Our IRB permits this re-association of deidentified data with instructor consent.
5. We then prioritized cases for review by aligning the magnitude and sense of the 3 team process measurement scales:
    a. *Satisfaction* is a positive construct relating to how satisfied a student is with their current group of teammates, and is measured on a scale from 1-5. It was shifted to a -2 to +2 scale by subtracting 3, then normalized to a -1 to +1 scale.
    b. *Conflict* is a negative construct relating to how much conflict is occurring in a team, and is measured on a scale from 1-5. It was shifted to a –2 to +2 scale by

subtracting 3, then normalized to a -1 to +1 scale, then multiplied by -1 so that higher numbers represented a better team dynamic.

   c. *Psychological safety* is a positive construct relating to whether a student feels accepted, respected, and confident within the team, and is measured on a scale from 1-7. Negatively coded items were already re-coded as positive in the data downloaded. The data were shifted to a -3 to +3 by subtracting 4, then normalized to a -1 to +1 scale.

6. We computed an average of the three resulting indices to create a scale from -1 to +1 including all three measures, such that -1 would represent the worst combination of team experiences (low satisfaction, high conflict, and low psychological safety).
7. We sorted the dataset by this new metric, and created a prioritized list of cases for further review.

Analysis

Based on the prioritization metric, we identified "focal students" for the analysis. For each focal student and their team, we extracted team comment data from the CATME system and created a "team experience transcript" within a single file to tell the story of the team during the semester. The team experience transcript consisted of the individual comments written by the each of the team members at the four team evaluation periods. At each evaluation period, team members provided a self-evaluation, an evaluation of other team members, and, optionally, confidential comments to the instructor. We read in the following order: the final evaluation periods (3 and 4) appearing first and the initial comment periods (1 and 2) appearing afterward. We used this approach to orient the reader to the portion of the team experience where conflict would likely appear. Within each evaluation period, we listed the team comments about one another first, followed by any comments to the instructor.

Once the transcripts were assembled, the analyst began data analysis by first reading through the entire team experience transcript without taking any notes or marking the document. The analyst then reviewed the team experience transcript again highlighting significant statements or phrases from the comments that provided an understanding of the team experience. The analyst sought to identify how each team member characterized work contributions of themselves and others, what team members valued from the team experience, commentary regarding personality, the appearance of coded language in any descriptions, and identification of team member marginalization. This process was a line-by-line analysis performed three times. During the initial review of team comments, the analyst identified programming skill as a salient observation discussed by students in their reviews of one another. Thus, the first review focused on identifying the perceived programming skill level of the team members. We categorized team members based on the descriptions of programming ability provided by in the comments. The second review focused on the perceived contributions of team members. We classified team members as "high", "medium", or "low" contributors based on the students' own perceptions and the perceptions of their team members. The final review concentrated on the relations between the focal student and the other team members. This approach developed from the initial review of the team experience transcripts. We generated notes for each team that summarized the analysis from each pass of the team experience transcript and demographic information on the team members. We assembled the analysis into a coherent team "journey" presentation, which contained the analyst's interpretation of the team experience. We discussed the team journey as an author team to consider alternative interpretations of the team experience.

Results

The algorithm we articulated above indeed identified teams where teammates engage in important and significant marginalization. Of the 1629 students enrolled in one spring iteration of the course, the algorithm above identified 46 focal students (so approximately 46 teams to focus on). For these focal students, the prioritization metric ranged from -0.35 to 0.40. To answer our research question, "How is marginalization legible (or not) to instructors at the classroom level," we explored the student comment text as a tool instructors may use to provide a potential window into experiences of marginalization after the initial flagging from the CATME metric.

We provide an illustration of the team experience journey analysis as an example of how the team comments may provide insight into team dynamics, reveal marginalization, and offer early indicators of potential issues. For this example, outlined in Table 1, the focal student is Carla, a Black woman who is a US citizen. Her teammates are Jamie (Hispanic woman, International); Jack (white man, US citizen), and Harry (man, International, no race reported). Carla's prioritization metric from the CATME data was 2.0.

Table 1: Demographic Information of Example Team Members

Pseudonym	Gender	Race	US Citizenship
Carla (focus)	Woman	Black	US
Jamie	Woman	Hispanic	International
Jack	Man	White	US
Harry	Man	Not Reported	International

At the first evaluation period, which occurred around 6 weeks into the semester, programming ability was a salient and valued characteristic observed by students of themselves and their teammates. Teammates identify Carla and Harry as the more experienced programmers. Jack comments that Harry's "skill in MATLAB is greater than mine …." Harry describes Carla as someone who "knows her work." Although differences in programming ability are apparent at this point in the semester, this team appears to function well with equal contributions among the team members.

At the second evaluation period, occurring at 10 weeks into the semester, there is a notable shift in the team dynamics observable through the comments about the team's high ability programmers. Jamie describes Harry as follows: "Harry is brilliant and super dependable." Carla comments that Harry "is very knowledgeable and contributes a lot of helpful information to the team." Jack says that Harry "is perfect. He does his portion of the work, sometimes above and beyond, and still manages to communicate well with the rest of us and maintain a positive attitude." The team's comments about Carla at this stage recognize her contributions generally. Harry notes that Carla "is always there to help everyone and she also makes sure we accomplish our short term goals" and Jamie describes Carla as "extremely supportive in every aspect." Jack's evaluation of Carla takes a different tone, however. While he recognizes the value of the work that Carla does, his comment reveals some negative perceptions: "[Carla] is a good group member who cares about the completion and quality of the work we do. Sometimes that means she can be *pushy* or overwhelmed, but I suppose that just comes with the stress of the workload" (emphasis added). At this point, the negative perception does not appear to be reciprocated as Carla comments that Jack "is very organized and does his fair share of the workload." Jack's use of the term "pushy" is the most explicit example of gender- and racially-coded language that appears in the comments. A more subtle aspect of the comments to note is the gendered language used in attributes ascribed to the two more skilled programmers. While Harry is viewed as "brilliant", "perfect", and "knowledgeable", Carla is described as "supportive" and helpful.

The third evaluation period (at approximately 13 weeks) marks a sharp shift in the team dynamics and contributions as noted by Carla. She comments that she and Harry have contributed more than their "fair share of work." Carla remarks that Jack "showed up and tried, but did not contribute much" and that Jamie "did not show up and has contributed nothing and done no work." This perception is confirmed in Jack's comments. Jack describes Harry as "an all-around wonderful teammate and person. He contributes a lot to the completion of each project and does so with a good attitude." Jack notes that Jamie "interacts well with everyone and contributes when necessary, but much like myself she is not quite as advanced with complex loops and a lot of the more challenging coding as Carla and Harry are." Jack notes for himself that "one shortcoming I had was finding time to meet my group members in person…. I intended to communicate with them and make sure this does not happen for future milestones." Jack's direct comments on Carla begins by recognizing Carla's skill and contribution: "Carla is a hard-working group member who contributes a lot to the completion of the assignment and cares deeply about the quality of the work we turn in." However, as he continues, he describes her negatively: "One thing she could work on would be interacting better with us under pressure. When stressed, she can be a bit *harsh* and *inflexible*" (emphasis added to highlight coded language in the comment). Here, we again see coded language appear in Jack's description of Carla.

The final evaluation period (at 15-16 weeks) marks a continuation of the dynamics of the third period. Comments from all team members again confirm the high contribution level of Carla and Harry and relatively limited contributions of Jack and Jamie. Jack's pattern in describing Carla is consistent with the prior evaluation periods. He begins by acknowledging her contributions and then adds commentary related to his perception of her teamwork skills that employs coded language: "Carla really cares about the quality of the work we submit and contributes a lot to the completion of the work; however, her team *behavior skills could use some work* and *she is not a very flexible person.* It was kind of *her way every time* and *she was always grumpy*" (emphasis added). Jack's statements stand in contrast to how other teammates perceive Carla. Harry comments that "Carla also had major contributions to the project with the work documents and she also gave some input on the code." Jamie describes Carla as "an extraordinary teammate, she is always willing to go out of her way to help others and is a joy to spend time with." Additionally, Jack's comments about Carla are starkly different from his comments about his other teammates. Jack states that "Harry is perfect. He really understands the material, works hard to contribute to the group work, and does it with a good attitude" and "Jamie is a lot like me in that she found herself not knowing as much about MATLAB and thus, not being as useful." Carla's comments for this period reflect continued frustration with the uneven work distribution within the team. She states, "I have contributed more than my fair share of work to each and every milestone…. I feel the quality of the work I have been doing is high and that I have been an effective team member."

Beyond the added workload, Carla's experience may have been even more negatively impacted by her interactions with Jack. While we do not know how their in-person interactions played out, we can see documented evidence from the students' comments that indicate gender and racial bias. Jack's comments for Carla during the final three evaluation periods contain coded language with negative racial and gender stereotypes and that language becomes more extreme as time progresses. This team journey documentation demonstrates the potential utility of the defined prioritization metric. Using the metric as a flag, potential marginalization can become legible through analysis of the comments on team experience provided by students as a normal part of their course requirements.

Conclusion

While this research is in its early stages, we see from the data above that there are early signs in CATME data of marginalizing team dynamics which can be identifiable by a trained observer, and which could benefit from intervention. We believe not everyone may see those signs at first, particularly people who identify with a majority culture in US engineering education, including people who identify as white, as cis-gender male, and as the overlap of these two cultures. However, our work strives to develop tools to help particularly demographic majority members know when their teams need intervention around issues of marginalization, and of what sort. The algorithm we have articulated above helps identify where there could be problems, and if the instructor is tuned to key language that identifies those problems, they can intervene early.

We see from this early work in our project that helping students improve their skills in working on diverse teams takes intense work. Instructors need to see this work as part of the educational process, and as a way of working that is critical to engineering. Many instructors do not have these skills themselves, and so expecting them to be able to teach students to work in diverse teams when they themselves do so poorly will not work. These claims are borne out in our interviews with instructors, which we do not describe here but which is part of the rest of the project's work.

One of the limitations of the current work is that, while the algorithm we developed using CATME team ratings provides a useful numerical gauge of individual experience, the algorithm may yield "false positives" for marginalization. We are currently reviewing comments in those cases to better understand the distinctions in the team dynamics which may help us further refine the algorithm. Another limitation associated with this approach is that the numerical measures may miss some cases of marginalization. We selected this approach as a first step to analyze the very large data set. Our next steps for this CATME portion of our research include developing a codebook to look for language to search for in other teams' evaluations. We also will search team evaluations which the algorithm did not pick up, and then explore why the algorithm did not pick up on them, to determine whether there is something else we should be noticing regarding marginalization in those evaluations. Lastly, our research may help identify ways to better design the tool used to provide feedback (CATME, in this case) to prevent its use as a vehicle for marginalization.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1936778. The authors thank the reviewers for their constructive comments.

References

[1] D. R. Bacon, K. A. Stewart, and W. S. Silver, “Lessons from the Best and Worst Student Team Experiences: How a Teacher can make the Difference,” *Journal of Management Education,* vol. 23, no. 5, pp. 467–488, Oct. 1999.

[2] N. A. Lloyd and J. Szymakowski, “Student experiences in first-year engineering classrooms – exploring issues of gender in an Australian programme,” *Australasian Journal of Engineering Education*, vol. 22, no. 1, pp. 23–31, Jan. 2017.

[3] S. Jaikaran-Doe, A. Henderson, E. Franklin, and P. Doe, “Strategies for promoting cultural diversity within student laboratory groups in an engineering degree course at an Australian university,” Australasian Association for Engineering Education Annual Conference 2018, Hamilton, New Zealand.

[4] M. V. Jamieson and J. M. Shaw, “Applying Metacognitive Strategies to Teaching Engineering Innovation, Design, and Leadership,” *Proceedings of the Canadian Engineering Education Association,* 2017.

[5] S. Beecham, T. Clear, J. Barr, M. Daniels, M. Oudshoorn, and J. Noll, “Preparing Tomorrow’s Software Engineers for Work in a Global Environment,” *IEEE Software*, vol. 34, no. 1, pp. 9–12, Jan. 2017.

[6] J. R. McCusker, A. Carpenter, and J. Sosnik, “Improving the Student Experience in First Year Engineering Design Courses,” presented at the First Year Engineering Experience Conference, 2017, p. 6.

[7] S. V. Rosser, “Group Work in Science, Engineering, and Mathematics: Consequences of Ignoring Gender and Race.” *College Teaching*, vol. 46, no. 3, pp. 82–88, 1998.

[8] D. W. Sue, C. M. Capodilupo, G. C. Torino, J. M. Bucceri, A. M. B. Holder, K. L. Nadal, and M. Esquilin, “Racial microaggressions in everyday life: Implications for clinical practice,” *American Psychologist*, vol. 62, no. 4, pp. 271–286, 2007.

[9] A. H. Eagly and S. J. Karau, “Gender and the emergence of leaders: A meta-analysis.,” *Journal of Personality and Social Psychology*, vol. 60, no. 5, pp. 685–710, Jan. 1991.

[10] A. H. Eagly, M. G. Makhijani, and B. G. Klonsky, “Gender and the evaluation of leaders: A meta-analysis.,” *Psychological Bulletin*, vol. 111, no. 1, pp. 3–22, 1992.

[11] H. Tajfel, *Human groups and social categories: Studies in social psychology*. 1981.

[12] M. A. Hogg and D. I. Terry, “Social Identity and Self-Categorization Processes in Organizational Contexts,” *Academy of Management Review* vol. 25, no. 1, pp. 121–140, Jan. 2000.

[13] C. Leaper and M. M. Ayres, “A Meta-Analytic Review of Gender Variations in Adults’ Language Use: Talkativeness, Affiliative Speech, and Assertive Speech,” *Personality and Social Psychology Review*, vol. 11, no. 4, pp. 328–363, Nov. 2007.

[14] S. R. Bird, “Welcome to the Men’s Club: Homosociality and the Maintenance of Hegemonic Masculinity,” *Gender & Society,* vol. 10, no. 2, pp. 120–132, 1996.

[15] L. M. Cortina, “Unseen Injustice: Incivility as Modern Discrimination in Organizations,” *Academy of Management Review,* vol. 33, no. 1, pp. 55–75, Jan. 2008.

[16] L. M. Cortina, D. Kabat-Farr, E. A. Leskinen, M. Huerta, and V. J. Magley, “Selective Incivility as Modern Discrimination in Organizations,” *Journal of Management*, vol. 39, no. 6, pp. 1579–1605, Sep. 2013.
[17] L. Zurbrügg and K. N. Miner, “Gender, Sexual Orientation, and Workplace Incivility: Who Is Most Targeted and Who Is Most Harmed?,” *Frontiers in Psychology*., vol. 7, 2016.
[18] K. N. Miner, I. Diaz, and A. N. Rinn, “Incivility, Psychological Distress, and Math Self-Concept Among Women and Students of Color in Stem,” *Journal of Women and Minorities in Science and Engineering*, vol. 23, no. 3, pp. 211–230, 2017
[19] E. Cech and T. Waidzunas, “‘Engineers Who Happen To Be Gay’: Lesbian, Gay, And Bisexual Students’ Experiences In Engineering,” presented at the 2009 ASEE Annual Conference & Exposition, 2009, pp. 14.1384.1-14.1384.28.
[20] M. R. Woodford, C. L. Kolb, G. Durocher-Radeka, and G. Javier, “Lesbian, Gay, Bisexual, and Transgender Ally Training Programs on Campus: Current Variations and Future Directions,” *Journal of College Student Development*., vol. 55, no. 3, pp. 317–322, 2014.
[21] K. Beddoes and G. Panther, “Gender and teamwork: an analysis of professors’ perspectives and practices,” *European Journal of Engineering Education*, vol. 43, no. 3, pp. 330–343, May 2018.
[22] A. Joshi, “By Whom and When Is Women’s Expertise Recognized? The Interactive Effects of Gender and Education in Science and Engineering Teams,” *Administrative Science Quarterly,* vol. 59, no. 2, pp. 202–239, Jun. 2014.
[23] C. E. Williams, D. J. Woehr, A. C. Loignon, M. L. Loughry, and M. W. Ohland, “Men behaving badly or gender bias? An examination of gender effects in peer assessment ratings,” presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL, 2018.
[24] L. A. Meadows and D. Sekaquaptewa, “The Effect of Skewed Gender Composition on Student Participation in Undergraduate Engineering Project Teams,” in ASEE Annual Conference and Exposition, Vancouver, BC, Canada, 2011.
[25] L. A. Meadows and D. Sekaquaptewa, “The Influence of Gender Stereotypes on Role Adoption in Student Teams,” in ASEE Annual Conference and Exhibition, Atlanta, Georgia, 2013.
[26] D. A. Trytten, R. Pan, C. Foor, R. Shebab, and S. Walden, “Inclusion or Exclusion? The Impact of the Intersection of Team Culture and Student Identity and Pathway on Team Diversity,” in ASEE Annual Conference and Exhibition, Seattle, WA, 2015.
[27] J. Wolfe, E. Powell, S. Schlisserman, and A. Kirshon, “Teamwork in Engineering Undergraduate Classes: What Problems Do Students experience? : American Society for Engineering Education,” presented at the 2016 ASEE Annual Conference & Exposition, New Orleans, LA, 2016.
[28] D. J. Beal, R. R. Cohen, M. J. Burke, and C. L. McLendon, “Cohesion and performance in groups: a meta-analytic clarification of construct relations,” *Journal of Applied Psychology*, vol. 88, no. 6, pp. 989–1004, Dec. 2003.
[29] F. R. C. de Wit, L. L. Greer, and K. A. Jehn, “The paradox of intragroup conflict: a meta-analysis,” *Journal of Applied Psychology*, vol. 97, no. 2, pp. 360–390, Mar. 2012.
[30] S. M. Gully, K. A. Incalcaterra, A. Joshi, and J. M. Beauien, “A meta-analysis of team-efficacy, potency, and performance: interdependence and level of analysis as moderators of observed relationships,” *Journal of Applied Psychology*, vol. 87, no. 5, pp. 819–832, Oct. 2002.

[31] K. A. Jehn and E. A. Mannix, "The Dynamic Nature of Conflict: A Longitudinal Study of Intragroup Conflict and Group Performance," *Academic Management Journal*, vol. 44, no. 2, pp. 238–251, Apr. 2001.
[32] A. Edmondson, "Psychological Safety and Learning Behavior in Work Teams," *Administrative Science Quarterly*, Vol. 44, no. 2, p. 350, Jun. 1999.
[33] D. Chan, "Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models," *Journal of Applied Psychology*, vol. 83, no. 2, pp. 234–246, 1998.
[34] D. J. Woehr, A. C. Loignon, P. B. Schmidt, M. L. Loughry, and M. W. Ohland, "Justifying Aggregation with Consensus-Based Constructs: A Review and Examination of Cutoff Values for Common Aggregation Indices," *Organizational Research Methods*, vol. 18, no. 4, pp. 704–737, Oct. 2015.
[35] J. S. Thomas, A. C. Loignon, D. J. Woehr, M. L. Loughry, and M. W. Ohland, "The relative effects of warmth and competence for team viability," presented at the *Society for Industrial and Organizational Psycholog*y, Philadelphia, PA, 2015.
[36] Delgado, R., & Stefancic, J. (2017). *Critical race theory: An introduction*. New York University Press.