INVESTING MICRO-BEHAVIORS IN TEAM INTERACTIONS BETWEEN ENGINEERING STUDENTS

An Undergraduate Research Scholars Thesis

by

KIARA BERRY

Submitted to the LAUNCH: Undergraduate Research office at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Faculty Research Advisor: Dr. Theodora Chaspari

May 2023

Major: Computer Science

RESEARCH COMPLIANCE CERTIFICATION

Research activities involving the use of human subjects, vertebrate animals, and/or biohazards must be reviewed and approved by the appropriate Texas A&M University regulatory research committee (i.e., IRB, IACUC, IBC) before the activity can commence. This requirement applies to activities conducted at Texas A&M and to activities conducted at non-Texas A&M facilities or institutions. In both cases, students are responsible for working with the relevant Texas A&M research compliance program to ensure and document that all Texas A&M compliance obligations are met before the study begins.

I, Kiara Berry, certify that all research compliance requirements related to this Undergraduate Research Scholars thesis have been addressed with my Faculty Research Advisor prior to the collection of any data used in this final thesis submission.

This project required approval from the Texas A&M University Research Compliance & Biosafety office.

TAMU IRB #: 2021-0220D Approval Date: 12/08/2022 Expiration Date: 12/07/2023

TABLE OF CONTENTS

Page

ABSTRACT

Investigating Micro-behaviors in Team Interactions between Engineering Students

Kiara Berry
Department of CSCE
Texas A&M University

Faculty Research Advisor: Dr. Theodora Chaspari
Department of CSCE
Texas A&M University

This thesis examines the interpersonal behaviors between first and second-year undergraduate students pursuing a STEM field, in a group setting. Dynamics in team settings can contribute to the overall success of a team, mental health of the students, and even long-term success of the students' career. Interdisciplinary research has begun to study how technology can improve human interactions and communication for humans to reduce racism, sexism, and hate speech. Many of these technologies have been built by taking textual examples from social media and other such platforms. This thesis explores the linguistic markers of team interactions between Engineering students. The data comes from a newly collected corpus of team interactions conducted in an online format, where teams of 3 students (1 female, 2 male; or 1 male, 2 female) and 4 students (2 female, 2 male) work together to solve a set of programming problems. These teams met daily for 5 consecutive days solving 2 coding challenges taken from Leetcode medium to hard problems at each meeting, with 35 minutes to solve each problem. From the conducted linguistic analysis, we identified over 100 statistically significant correlations between linguistic features generated by the Linguistic Inquiry and Word Count (LIWC) toolbox and self-reported psycho-social constructs related to individual emotion and team interaction quality that the participants filled out at various points throughout the study. These relationships give us further insight to the categories of verbal

language and its associative underlying emotion. As part of the future work, this dataset will be coded for microaggressions and microaffirmations and will then be used to design an Artificial Intelligence model that can predict, from a real time interaction, whether a statement is a microaggression or microaffirmation. A technology such as this can be used to enhance interactions in group settings and improve individuals' communications to reduce microaggressions.

ACKNOWLEDGMENTS

Contributors

I would first and foremost like to sincerely thank my faculty advisor, Dr. Theodora Chaspari, and her PhD student, Projna Paromita, for their guidance and support throughout the course of this research.

Furthermore, I would like to thank my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience and supporting me throughout my studies.

I would like to thank my brother, Nolan, for always inspiring me and setting an example for me throughout my schooling and time here at Texas A&M. Thank you to my parents for their encouragement and their continuous support, not only throughout this research study but my entire academic life, and for being my greatest mentors.

Funding Sources

This work was also made possible in part by the Texas A&M Triads for Transformation (T3) Program. The internal grant we received was titled Tracking Microaggressions In Science, Technology, Engineering, And Mathematics; Texas A&M Triads for Transformation (T3); Theodora Chaspari (PI), Ruihong Huang (Co-PI), Srividya Ramasubramanian (Co-PI); 2021-2022.

1. INTRODUCTION

Within the STEM workforce there are many minoritized groups that are underrepresented. Women specifically remain underrepresented in the STEM disciplines of computing and engineering. A previous study examining the gender gap in STEM identified that women held a disproportionately lower share of STEM degrees, specifically in engineering, and only 26% of women that did earn a STEM degree pursued a job in the STEM field, whereas for men this was 40% [1]. Women in Engineering disciplines have higher attrition and turnover rates from STEM occupations than men after college graduation and this can be attributed to the multitude of challenges faced by women in the STEM field [2]. This increasing gender, racial, and ethnic diversity gap and its long-term outlook is closely tied to the representation in the STEM educational system, particularly at colleges and universities. College culture can influence students' dedication, experiences, and persistence in their degree, and it has been shown that these minoritized groups are more likely to switch out of STEM majors [3]. To decrease this gender and racial gap, it is important to try to understand how these groups are being treated in the educational system and in interpersonal settings, and whether the lack of inclusion potentially plays a factor in them leaving the STEM discipline.

1.1 Background

Effective communication leads to better services, boosts employee morale, engagement, productivity and is key for better team collaboration–which is essential in the STEM workplace and is what cultivates a positive work environment. Microaggressions, which are not as obvious as direct insults, are daily verbal (or nonverbal) slights, either made intentionally or unintentionally, that discriminate against individuals of a marginalized group [4]. An example of this in the context to our study would be a man saying to a woman that they were only accepted into their degree solely because of their gender. The underlying gender stereotype behind this remark would be that women are not skilled in STEM and would need special consideration to be accepted. Microaggressions are mostly unconscious biases so it can be sometimes difficult to detect but can have a deep and pervasive impact. Microaggressions can be in the form of microassaults (outright prejudice-the easiest to identify), microinsults (communications that demean a person's identity but made usually unintentionally), and microinvalidations (communications that negate or exclude others' experiences, thoughts, or feelings) all which can distract from a person's focus, concentration, and success in an academic or professional setting [5]. It has been shown that daily experiences of microaggressions have stronger negative effects on the emotional and mental well-being than major discriminatory events and can even lead to chronic illness [6]. Microaffirmations, on the other hand, are subtle kindness cues that include nods, choices of words, and tones of voice that convey inclusion and active listening. An example of this in the context to our study would be students recognizing the achievements of their teammates, listening, and showing genuine interest to their opinions. Microaffirmations are important for minoritized groups' integration into the STEM community.

An individual's ability to empathize with others and effectively communicate is essential in preventing conflicts and fostering an uplifting workplace environment. It has been shown that team members with high emotional intelligence are effective at building trust and fostering collaboration [7]. An individual's ability to recognize when they are speaking or on the receiving end of a microaggression/microaffirmation can boost emotional intelligence and awareness, improve communication skills, and help foster a healthier work environment. An individual's choice of words in their interactions can be very impactful, therefore both positive (microaffirmative) and negative (microaggressive) language used in academic and professional settings in STEM can be highly influential to cultivating an inclusive (or non-inclusive) environment and ensuring success to minority individuals.

1.2 Our Work

Observing students in STEM disciplines interact in a team setting provides an opportunity to gain knowledge into inclusive or exclusive behaviors with respect to women in STEM and how these micro-behaviors affect the individual and overall group dynamic. Our goal with this research is to investigate spoken examples of microaggressions and microaffirmations taken from real time interactions. Micro-behaviors are subtle in nature and context-dependent, therefore it can be challenging for humans or machines (e.g., machine learning algorithms) to accurately identify the occurrences of these behaviors. As a first step toward accomplishing this, data from student interactions was collected in the context of small teams solving programming exercises. We conduct a linguistic analysis of the transcripts obtained by these interactions and capture the content/sentiment of the individuals' statements. The linguistic measures are cross-examined with self-reported psycho-social constructs related to individual emotion and team interaction quality. The collected data and corresponding analysis can help provide context for real life situations such as in the STEM workforce that resemble the interactions from university students pursuing a STEM degree and can be used to train and test micro-behavior detection models.

While our dataset is not an exhaustive collection of microaggressions and microaffirmations, it is a first look at identifying microaggressions from students in university pursuing a STEM degree. Previous research involved focus group discussions and interviews that showed evidence that microaggressions exist in STEM-where they would ask students to share their experiences in a retrospective manner for social sciences research. But microaggressions have not been observed in the moment in this manner. If we can identify microaggressions and microaffirmations in the real world this would not only help humans improve their emotional intelligence, but it could also help encourage these minoritized groups to continue with their degree and enter the STEM workforce-reducing the gender, racial, and ethnic gap that currently exists in STEM. Our dataset has the potential to be used across further research disciplines to further investigate the impact of microaggressions and microaffirmations on human emotion and mental and physical health. Companies and similar agencies in the workforce can make use of an Artificial Intelligence (AI) model that detects these micro-behaviors in real time. A technology such as this could improve the emotional intelligence of employees, allowing them to become more empathetic in their language, which would cultivate a more inclusive and uplifting environment.

2. LITERATURE REVIEW

Prior studies of identifying microaggressions have mostly fallen under the domain of interdisciplinary research. These studies mostly focused on identifying microaggressions through surveys, focus group discussions, or interviews with women, black/African Americans, and other minoritized groups in engineering. In these types of studies, participants would discuss their experiences where they faced a microaggression in a retrospective manner. These discussions/interviews focused their questions on the individual's experience in the STEM community and included some of the responses received in their work as well as their demographics. The outcomes of these studies gave better insight into the different types of microaggressions, how the academic and social climate of engineering education is shaped by microaggressions, and the different types of responses a victim gave when they encountered this discrimination.

In this study and related research, microaggressions were classified as either: interpersonal—where an individual from a majority group would ignore a person from the minoritized group, or diminish their success because they belong to that group; institutional—the marginalizing actions towards minoritized groups that contribute to a hostile working environment; or jokes—sexual remarks made by male students around their female classmates or sexist comments meant to be funny [8]. From the participants' responses in these studies, their level of tolerance to their peers behaviors ranged from lashing out to being almost indifferent and accepting them as "normal" behavior [9]. There is limited literature on microaggressions related to other personal characteristics such as gender identity, religious beliefs, and socioeconomic status, however this should still be thought of when considering one's speech. Studies of microaffirmations are also just beginning but their potential has been shown to foster inclusion in these groups and raise productivity by reinforcing and rewarding good behavior [10].

Current literature in the realm of AI has been limited to detecting abusive or hate speech mostly from social media. However, these methods do not work well for microaggressions and microaffirmations since these micro-behaviors are subtle in their manner and context sensitive. Due to these inherent challenges, creating and AI model that is tasked with detecting microaggressions and microaffirmations has been under-explored, making it a difficult task yet an actively researched problem [11, 12]. Previous work has explored creating datasets of microaggressions that were identified in American television shows and other web platforms [13]. Such datasets, such as the one we will be discussing in this paper, can be used by researchers to create AI models which would help in enhancing human communications.

In each of these scenarios, it was observed that these microaggressions create an unwelcoming environment and can have a serious negative affect on women and other minoritized groups. While some find other avenues to be successful such as joining organizations such as the Society of Women Engineers (SWE) or the Society of Hispanic Professional Engineers (SHPE), as an example, these negative experiences can be so discouraging as to limit or isolate these groups. For some women, they had to change their behavior to adapt to their group interactions with their male counterparts, but for other women they felt stressed and unmotivated to participate in the group altogether. Minorities are the most rapidly growing population in the United States and yet they are severely underrepresented in STEM which is even more prevalent in graduate education [14]. The inclusion of these groups in the STEM community is vital for the growth of STEM and creating well rounded products and solutions since the decisions being made would have a higher representation of minorities.

3. METHODS

3.1 Participant recruitment and study procedures

We conducted our research at Texas A&M University. Participants were recruited to our study via sending out invitation emails to campus-wide mailing lists and engineering departments, and displaying study advertisements on the displays in the engineering building. Students who were interested in the study would reach out through email to the research team. To be considered for the study, students had to meet these qualifications:

- Be a first or second year undergraduate.
- Be 18-30 years old of age.
- Be able to speak English.

In the recruiting materials that were sent out, students were only told they would be working with a team to solve programming questions and would receive a $50 amazon gift card upon completion. The true purpose of the study was disclosed. From the interested students that reached out to us, we set up 2 teams of 3 students (1 female, 2 male; and 1 male, 2 female) and 1 team of 4 students (2 female, 2 male). Our sample included various engineering majors and was not limited to a certain sub-discipline. Students ranged in age from 18 and 19 and were either freshman or sophomore classification. Out of the 10 total students, 6 were Asian, 3 were Hispanic, and 1 was a mix of 2 or more races. 9 out of the 10 students classified themselves as unemployed.

The teams met daily for 5 consecutive days through Zoom teleconferencing and were given 2 programming questions taken from Leetcode*, an online platform for practicing coding challenges designed to prepare software engineers for technical interviews, with 35 minutes to work on each problem. These questions were labelled medium to hard problems on leetcode. Due to the

*https://leetcode.com/

difficulty level of the questions, we did not expect teams to finish. Each student was given a codename to use during the meeting to protect confidentiality and were asked to keep their video on throughout the meeting. A research assistant was present at each meeting to prevent adverse events if, in the unlikely case, they were to occur. Each meeting was recorded, and transcript files and audio files were generated from Zoom. At the end of the study each participant was compensated with a $50 Amazon gift card and the zoom moderator disclosed the true objective of the study–to observe microaggressions toward minoritized groups in STEM in a group study session.

3.2 Questionnaires to Collect Student Emotional, Psychological, Cognitive Processes

Throughout the meetings, participants were asked to fill out forms so we can gauge their underlying emotion while they were completing the programming questions. Students were given a participant ID to use in their submissions of the questionnaires to keep their identity private. The questions that were included in each forum came from questionnaires from the bio-behavioral markers of human performance list of measures. Each questionnaire's (12 total) questions aim to identify certain psychological, cognitive, or behavioral processes in the individual. Each questionnaire has their unique way to calculate an overall score, or multiple subscale scores, from the responses collected.

On the first day of the team meetings, we gave them an enrollment forum to fill out to collect information about their background and feelings in stressful situations. The questionnaires that were included in this forum were: Demographics Survey, The Coping Inventory for Stressful Situations-short version [15], The College Resilience Questionnaire [16], Big Five Inventory [17], Internal-External Locus of Control Scale [18], and Stressful Life Events Screening Questionnaire (SLESQ) [19]. At the start of each meeting, each member of the team was asked to complete a form prior to solving the programming problems for pre-scaling of mood-this was the prior study forum. The questionnaires that were included in this forum were: Positive and Negative Affect Schedule (PANAS)-short version [20] and Perceived Stress Scale-short version [21]. At the conclusion of each programming problem, they were asked to fill out a form for post-scaling of mood-this was the post study forum. The PANAS, Perceived Stress Scale, and a third questionnaire-Team Conflict,

Conflict Management, Cohesion and Team Effectiveness [22], were included in this forum. On the last day of the team meetings at the conclusion of both problems they were asked to complete an exit forum. The questionnaires included in this forum were: The College Resilience Questionnaire, Gender Based Microaggression Scale [23], Flourishing Scale [24], and Defense Style Questionnaire [25].

From the responses that we received from the individuals across all the teams (3 total teams), we calculated, for each participant, their scores for each questionnaire which we will use later for analysis. For example, the PANAS questionnaire measures an individual's feelings by having the individual rank each word–interested, nervous, inspired, etc.-in a scale from 1 ("very slightly") to 5 ("extremely"). For the PANAS questionnaire scoring, there were 2 subscales that were measured–positive affect score and negative affect score–which were calculated by summing the responses of the questions that fell in its category. Each forum (4 total-Enrollment, Prior Study, Post Study, and Exit) was made up of a few questionnaires from the list of questionnaires. For each forum a spreadsheet was created that includes each participant's unique ID and their score(s) for each questionnaire.

From the post study analysis, the PANAS questionnaire was divided into 2 subscales, as previously mentioned, each with its own score, and the Team Conflict, Conflict Management, Cohesion and Team Effectiveness questionnaire was divided into 6 subscales: relationship conflict, task conflict, conflict management, cohesion, team satisfaction, and team viability. Higher scores indicate a higher association to the category except for the team viability subscale-a higher score in this category indicates less team viability based on the questions asked in this section.

3.3 Linguistic Measures Extracted from Audio Transcripts

We had 1 annotator import the transcript and audio files into Audacity †, an audio editing software application, and manually go through the audio recordings correcting any discrepancies in the transcripts. Erroneous translations and words or phrases that zoom did not catch, were corrected, or added to the transcript text. We then ran the Linguistic Inquiry and Word Count (LIWC)

†https://www.audacityteam.org/

toolbox on all the pre-processed transcript files to identify the various emotional, cognitive, and structural components present in the individuals' verbal language. LIWC is a text analysis program that calculates the number of words from a transcript file that fall into one or more of around 90 linguistic categories-psychological, punctuation, informal language, etc [26]. For example, if the word "cried" was identified, this would increase the count to the sadness, negative emotion, and past focus categories. After going through all the words, LIWC would calculate the percentage of words that fall into each category. After running this software, files were generated for each team and each day, as reflected in the transcripts. Each file contains the verbal statements captured from the interactions between the students and lists all the LIWC categories and the rates that each category was used in the given text.

The scores generated by LIWC give insight to the psychological states from the participants' language. The scores calculated from the questionnaires give us insight into the underlying emotion that was self-measured by the individual. We computed the Pearson's correlations between the LIWC measures and the scores from the participants' answers to the post study questionnaires to identify how the linguistic measures are related to the measures that describe each session. We decided to focus on correlating the LIWC measures with just the scores from the questionnaires in the post study forum. This was because this forum gives us the most valuable information for students' perceived mood, stress, and team functioning behaviors while they were solving the programming problems with their teammates. Since each participant was asked to fill out this form twice at each meeting (after each programming problem) we split the post study questionnaire scores from the spreadsheet created into 2 categories–Task 1 and Task 2. Since the LIWC software assigns a score to each category for every sentence captured, participants had more than one score for each LIWC category. Because of this, we calculated an average score, for each participant and for each day in their study, for each LIWC measure.

3.4 Gather Underlying Emotion through Correlating LIWC Feature Scores with Questionnaire Scores

Each average LIWC measure score was then correlated with each questionnaire score for Task 1 and Task 2 separately. By separating correlations between Task 1 and Task 2 we are better able to analyze how their sentiment might have changed while working with their team over the course of the meeting. These correlations give valuable insight into the underlying emotions of the participants by associating emotional context to specific language categories through the relationships identified between the linguistic measures and self-reported measures. This will provide context to the teams' interactions where the microaggressions/microaffirmations will be identified and labelled for our dataset.

4. RESULTS

4.1 Correlation Results between Linguistic Features and Self-Reported Measures

To obtain the correlations between the linguistic features and the self-reported measures of individual emotion and team interaction quality, I calculated the Pearson's correlation coefficient (also known as r value) and the p-value which indicates the significance of the correlation. From running the LIWC software on the pre-processed transcripts, we collected over 90 different linguistic features that describe the content and tone of the conversation. Each feature was aggregated at the dialog-level, representing an overall linguistic measure for a given day of the study. Similarly, each self-reported score represents a specific day. Thus, the correlations were computed for each pair of LIWC feature and self-reported measure, each presenting a granularity on a daily-level. Toward this, I created a file which contained the feature name, the questionnaire name, the correlation coefficient, and the p value. For questionnaires whose scores were split into subscales, such as the PANAS and Team Conflict, Conflict Management, Cohesion and Team Effectiveness questionnaires, each subscale score had its own correlation.

Table 1: Correlation results between linguistic features and self-reported scores of individual emotion and team interaction quality (Task 1)

Linguistic Feature	Self-reported Score	Correlation	P-value
nonflu	Perceived Stress score	0.595057	0.000032
Analytic	Perceived Stress score	-0.562534	0.000106
achieve	PANAS Positive score	0.517706	0.000446
you	Team Viability score	0.595057	0.000032
Dash	Task Conflict score	0.505854	0.000632
Tone	Conflict Management score	0.504971	0.000648
focuspresent	PANAS Positive score	0.496921	0.000815
friend	Perceived Stress score	0.496125	0.000834
male	Perceived Stress score	0.494376	0.000875
swear	Conflict Management score	-0.477532	0.001384

Our findings from Task 1 indicate that 109 out of the 837 computed correlations had a p-value of less than 0.05, meaning that the corresponding features were significantly associated with the self-reports–indicating that there is a clear relationship between the two that is not attributed to chance. From Table 1, we can see that non-fluencies in the participants' dialog (e.g., "uh" and "um") had a high positive correlation to feelings of stress that the participant indicated in their post study analysis. The "analytic" feature, which referred to increased use of words related to analytical thinking, had a high negative correlation with feelings of stress. Inspection of the data suggests that there were many times during the sessions where students would discuss their strategies and thoughts on how to solve the problem out loud, so this association depicts that this type of dialog contributes to less feelings of stress from the individual. Another linguistic feature that had a high positive correlation to the PANAS positive affect emotion and was found to be statistically significant was the "achieve" category. Words that attributed to feelings of achievements, such as references indicating a successful implementation of the team's ideas, had a high positive correlation with positive emotions such as excited, inspired, and enthusiastic (i.e., categories from the PANAS questionnaire). These top features that we identified from our analysis are very useful in indicating the type of dialog and the underlying emotion that it instilled in the students while they were working on their task.

Table 2: Correlation results between linguistic features and self-reported scores of individual emotion and team interaction quality (Task 2)

Linguistic Feature	Self Reported Score	Correlation	P-value
family	PANAS Negative score	0.512226	0.000274
you	Team Viability score	0.496736	0.000446
family	Team Viability score	0.480001	0.000736
achieve	PANAS positive score	0.454869	0.001493
focuspresent	PANAS positive score	0.440849	0.002166
female	Relationship Conflict score	0.438068	0.002328
we	Team Viability score	-0.431869	0.002727
WC	Cohesion score	-0.427861	0.003016
Analytic	Perceived Stress score	-0.420114	0.003652
sad	PANAS Positive score	-0.411945	0.004447

Results from Task 2 suggest that 65 out of the 837 correlations were statistically significant. Some of the associations between features and self-reports that were identified from Task 1 are also observed in Task 2 (Table 2). Examples include the negative association between analytic use of language and perceived stress, and the positive association between achievement-related language and PANAS positive score. A positive association between the use of second-person pronouns (e.g., "you") and team viability was also found in both Task 1 and Task 2, which is an interesting metric. A high positive score in team viability (ability to work successfully) from our questionnaire represents less team viability based on the questions that were asked in this section. So, this correlation would indicate that when students referred to each other as "you" this had a negative effect on the teams' dynamics. Prior literature on romantic relationships studying the affect of spouses' pronouns use during marital interaction on their martial satisfaction found that You-focus pronouns had a high positive correlation to negative interaction behavior and a negative correlation to marital satisfaction [27]. In this context, the increased use of second-person pronouns could indicate more blaming. Additionally, they found that We-focus pronouns had a high negative correlation to negative interaction behavior and a positive correlation to marital satisfaction. This can also be seen from our results, with the high negative correlation between "we" and team viability. This suggests that teams with increased viability will include more first-person plural pronouns (e.g., "we") that refer to the entire team, encouraging effective collaboration.

The positive association between the use of female-specific references and relationship conflict is another interesting metric that was observed. This suggests that teams who indicated an increased use of gender-specific words (specifically female-specific words) depicted increased conflict. A theory from this relationship could be that gender-specific language might not be good for team functioning. Gender neutral language, on the other hand, could be more inclusive and avoid gender bias in speech altogether.

Some other relationships that were observed to be statistically significant such as words per sentence (WPS) and team viability, and dash and task conflict score do not contribute much to our understanding of the sentiment of the individual. While these features should not be ignored in our analysis due to the high significance of the correlation, it is important to keep in mind that words per sentence and dash are dependent upon when the annotator was going through the audio and how they decided to format the sentences while correcting discrepancies. For the other features that were identified as significant and whose relationship is indicative of the individuals' emotional state in their dialog with their teammates, the entire set of features can be further analyzed by human labelers when deciding which statements to classify as a microaggression or microaffirmation. For features that are not as obvious but still significant to the analysis, this can be further explored by research to determine what is useful to the context and should be considered by the human labelers when classifying the statements. Playing back the recorded interactions to hear the speakers' tones and closely observe the facial expressions would also provide better insight into the context of the setting.

5. LIMITATIONS

This study should be considered in the light of the following limitations. Data collection of recruiting participants began in the summer and continued into the academic year where it was harder to retain participants across each consecutive meeting day. Our sample was comprised of 3 teams-2 teams comprising of 3 students (1 female, 2 male; and 1 male, 2 female) and 1 team comprising of 4 students (2 female, 2 male)-so our data was limited. A larger sample size would be needed to better effectively train a model to detect these micro-behaviors since it would include a wider range of interactions and contexts.

Some of the students, in addition, had more programming experience than their teammates so in some instances they took charge and did most of the work. This caused a lack of communication and inclusion between teammates which limited the amount of text we were able to look at to classify micro-behaviors. Since our study focuses on micro-behaviors that are targeted to women it is important to consider that our study and our results do not include or represent a diverse pool of human characteristics or race/ethnicities.

Another limitation was that we recruited students from the engineering school at Texas A&M University so our study cannot be generalized across different campuses. Our study was focused on observing teams interact while working on a set of programming problems through Zoom, but microaggressions can occur even beyond the classroom and into other scientific professional contexts such as faculty meetings, conferences, stand up calls, etc. Further research would be needed to be able to generalize these findings to other schools at other regions of the country and in various group settings. To obtain context to the microaggressions and microaffirmations that are to be classified, we must also rely on the students' self-reporting metrics when answering the questionnaire forms-which gives us context to their underlying emotions throughout the study.

When computing the correlations between the LIWC features and self-reported measures, we separated the post study questionnaire scores into Task 1 and Task 2, and calculated the average scores of the LIWC features for each participant/day in the study. We then correlated these averaged feature scores with each questionnaire score for the matching participant/day for both Task 1 and Task 2 separately. However, to obtain more accurate results, additionally separating the generated LIWC files into Task 1 and Task 2 based on the timestamp of the captured dialog, averaging the feature scores for Task 1 and Task 2 separately, and then running the correlations across would have been better representative of the relationships determined at the correct segment of the study. The correlation analysis was conducted at the team-level, while it would be also important to examine associations between linguistic features and team functioning scores at the individual-level, and also separately from participants belonging to majority groups and minoritized groups.

Lastly, to successfully identify a microaggression or microaffirmation, speech to text translation must be highly accurate. Since the author of this thesis was the only annotator going through the audio recordings and manually correcting any discrepancies, even though each snippet was inspected 2-3 times before editing, it could be possible that there are discrepancies in the data that have not been found yet. When classifying microaggressions and microaffirmations, it is essential to have a diverse pool of annotators to eliminate any gender or racial bias. It is also essential to carefully construct a coding manual of micro-behaviors that relies on theoretical underpinnings and empirical evidence from prior work, as well as to revise this manual based on the collected data. In addition, when classifying these micro-behaviors, due to the subtlety of their nature, it would be vital to listen to the playback audio and even watch the recording at the timestamp of the interaction, in addition to reading the verbal text, to consider the tone of the speaker and facial expressions.

6. FUTURE WORK

My research was focused on developing a dataset that contains context into the underlying sentiment of the individuals with the goal of classifying verbal statements, taken from interactions between STEM students, as either a microaggression or microaffirmation. This dataset is in the process of being made and will require human annotators to label the statements from the audio transcripts. Classification of micro-behaviors from purely verbal statements alone would be challenging, so having context to the underlying emotions of the students during their interactions would help annotators make more accurate conclusions in their decision making. Carefully designing a coding manual with the help from researchers from psychology and communication sciences would further contribute to accurately characterizing the collected data in the context of micro-behaviors.

This dataset can be used across research disciplines. Researchers in the field of psychology or sociology can utilize the dataset to further expand upon microaggression research and the impact that microaggressions can have on a student pursuing a STEM career. Since our dataset would also identify microaffirmations, for which there is limited literature in this area, this would also aid in the development of a research question of how microaffirmations can encourage students to be more outspoken, improve team dynamics, and increase the likelihood of individual contribution when trying to solve a problem or task. The study of microaffirmations can further have useful implications into the design of interventions.

Ideally, this dataset will include the captured verbal statement, the speaker of the statement, the recipient, the context into the underlying emotions of the individuals based on the language category identified in the statement, the background of the speaker and recipient (i.e., their gender, ethnicity, etc.), and the label of the statement (“microaffirmation” or “microaggression”). This dataset would also contain the timestamp of the audio recording where the microaggression/microaffirmation was identified. This timestamp can be used to extract the video clip of the interaction and use it as input into facial recognition software which would provide more context of how the individuals in the interaction were feeling. The context that's already provided in our dataset from the correlation of the linguistic measures and the self-reported measures and this additional context can be used in Natural Language Processing (NLP) to train an AI model to be able to detect micro-behaviors in real time. This model would be trained to detect language depicting a micro-behavior while also considering the underlying emotion of the individuals in the interactions and how their language was influenced by this sentiment, which would improve model accuracy.

To achieve a reliable micro-behavior detection model, it is essential that the dataset continues to be expanded since it is not an exhaustive collection of all microaggressions and microaffirmations. Only once the dataset is refined in this manner, with NLP, this could advance the realm of artificial intelligence with models being able to detect subtle aggressive or affirmative words and tones in language since this has not been able to be done before. An AI technology such as this could be deployed in team settings at classrooms and in the workforce where it can be used for training purposes on emotional intelligence, or it could even monitor an interaction to ensure successful team dynamics.

6.1 Ethical considerations

Opportunities for this dataset include creating an AI model, using NLP, that can detect microaggressions and microaffirmations in group settings in this field in real time and can also be used to advance research in micro-behaviors in interdisciplinary fields, such as psychology. This can help tremendously in enhancing interactions and improving students' communication so teams can work better–a core feature in engineering that carries on into the workforce. As with any human-centered technology, there are a lot of ethical considerations before such technology can be designed and deployed to ensure societal benefits and prevent any potentially adverse effects. An important factor to consider when creating this technology is to ensure that errors are minimized, and the AI is deployed in a manner that benefits the team. How the AI communicates that a microaggression was detected is an important consideration. It is important that those in leadership roles in the learning environment or the workforce ensure that the microaggression would be acknowledged and taken seriously and the victim's feelings are validated. It is also important to consider the intent of the perpetrator and give them the benefit of the doubt, especially if prior awareness training has not been in place. Confidence in AI is difficult to achieve with 100% accuracy. In this scenario, it could be that a microaggression was detected but it was not the person's intention.

7. CONCLUSION

Artificial intelligence has been used more and more to enhance interactions and facilitate tasks but there has not been much research done on detecting microaggressions especially at universities and in the STEM field. In this paper we presented different metrics that will be included in the creation of a dataset of microaggressions and microaffirmations obtained from real-time group interactions between STEM students. Relationships determined from the linguistic features of the dialog between students and their self-reported metrics of their team dynamics, emotions, and stress levels can help us determine the underlying emotions that were present when students were working through the problems with their teammates. This underlying emotion gives context to the creation of our dataset which would allow for higher accuracy when labelling statements as either a microaggression or microaffirmation.

REFERENCES

[1] D. N. Beede, T. A. Julian, D. Langdon, G. McKittrick, B. Khan, and M. E. Doms, "Women in stem: A gender gap to innovation," *Economics and Statistics Administration Issue Brief*, no. 04-11, 2011.

[2] Y. J. Xu, "Attrition of women in stem: Examining job/major congruence in the career choices of college graduates," *Journal of Career Development*, vol. 44, no. 1, pp. 3–19, 2017.

[3] J. J. Park, J. Zheng, K. Lue, C. Salazar, A. M. Liwanag, R. M. Parikh, and J. L. Anderson, "Looking beyond college: Stem college seniors on entering the workforce and the impact of race and gender.," *Journal of Diversity in Higher Education*, 2022.

[4] D. W. Sue and L. Spanierman, *Microaggressions in everyday life*. John Wiley & Sons, 2020.

[5] C. Harrison and K. D. Tanner, "Language matters: Considering microaggressions in science," *CBE—Life Sciences Education*, vol. 17, no. 1, p. fe4, 2018.

[6] M. T. Williams, "Psychology cannot afford to ignore the many harms caused by microaggressions," *Perspectives on Psychological Science*, vol. 15, no. 1, pp. 38–43, 2020.

[7] R. Boyatzis, K. Rochford, and K. V. Cavanagh, "Emotional intelligence competencies in engineer's effectiveness and engagement," *Career Development International*, 2017.

[8] M. M. Camacho and S. M. Lord, "" microaggressions" in engineering education: Climate for asian, latina and white women," in *2011 Frontiers in Education Conference (FIE)*, pp. S3H–1, IEEE, 2011.

[9] C. R. Diaz-Espinoza, "You kind of have to prove it: Gender microaggressions within the lived experiences of women in engineering," 2015.

[10] M. Rowe, "Micro-affirmations and micro-inequities," *Journal of the International Ombudsman Association*, vol. 1, no. 1, pp. 45–48, 2008.

[11] L. Breitfeller, E. Ahn, D. Jurgens, and Y. Tsvetkov, "Finding microaggressions in the wild: A case for locating elusive phenomena in social media posts," in *Proceedings of the 2019*

conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), pp. 1664–1674, 2019.

[12] P. Paromita, A. Khader, S. Begerowski, S. T. Bell, and T. Chaspari, "Linguistic and vocal markers of microbehaviors between team members during analog space exploration missions," *IEEE Pervasive Computing*, 2023.

[13] G. J. Washington, G. Mance, S. K. Aryal, and M. Kengni, "Abl-micro: Opportunities for affective ai built using a multimodal microaggression dataset.," in *AffCon@ AAAI*, pp. 23–29, 2021.

[14] R. Congleton, ""it's like a rhetorical groundhogs day": Microaggressions, marginality, and the graduate student experience in stem," *EOL 574, Lorenzo Baber, Instructor, Spring 2013*, 2013.

[15] I. M. Brands, S. Köhler, S. Z. Stapert, D. T. Wade, and C. M. van Heugten, "Psychometric properties of the coping inventory for stressful situations (ciss) in patients with acquired brain injury.," *Psychological assessment*, vol. 26, no. 3, p. 848, 2014.

[16] J. Sun and D. Stewart, "Development of population-based resilience measures in the primary school setting," *Health education*, 2007.

[17] M. N. Shiota, D. Keltner, and O. P. John, "Positive emotion dispositions differentially associated with big five personality and attachment style," *The journal of positive psychology*, vol. 1, no. 2, pp. 61–71, 2006.

[18] J. B. Rotter, "Generalized expectancies for internal versus external control of reinforcement.," *Psychological monographs: General and applied*, vol. 80, no. 1, p. 1, 1966.

[19] L. A. Goodman, C. Corcoran, K. Turner, N. Yuan, and B. L. Green, "Assessing traumatic event exposure: General issues and preliminary findings for the stressful life events screening questionnaire," *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, vol. 11, no. 3, pp. 521–542, 1998.

[20] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the panas scales.," *Journal of personality and social psychology*, vol. 54, no. 6, p. 1063, 1988.

[21] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *Journal of health and social behavior*, pp. 385–396, 1983.

[22] A. G. Tekleab, N. R. Quigley, and P. E. Tesluk, "A longitudinal study of team conflict, conflict management, cohesion, and team effectiveness," *Group & organization management*, vol. 34, no. 2, pp. 170–205, 2009.

[23] J. A. Lewis and H. A. Neville, "Construction and initial validation of the gendered racial microaggressions scale for black women.," *Journal of counseling psychology*, vol. 62, no. 2, p. 289, 2015.

[24] E. Diener, D. Wirtz, W. Tov, C. Kim-Prieto, D.-w. Choi, S. Oishi, and R. Biswas-Diener, "New well-being measures: Short scales to assess flourishing and positive and negative feelings," *Social indicators research*, vol. 97, pp. 143–156, 2010.

[25] G. Andrews, M. Singh, and M. Bond, "The defense style questionnaire.," *Journal of Nervous and mental Disease*, 1993.

[26] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of liwc2015," tech. rep., 2015.

[27] R. A. Simmons, P. C. Gordon, and D. L. Chambless, "Pronouns in marital interaction: What do "you" and "i" say about marital health?," *Psychological science*, vol. 16, no. 12, pp. 932–936, 2005.

APPENDIX: LIWC FEATURE INFORMATION

Table A.1: LIWC features identified and their meanings

LIWC Feature	**Explanation**
WC.	word count, incremented for each word in statement.
Analytic.	summary language variable that examines analytical thinking.
tone.	summary language variable that examines tone.
nonflu.	nonfluency, informal language such as "hm", "umm".
swear.	swear words.
achieve.	words related to achievement such as "win", "success".
sad.	words related to sadness such as "crying", "grief".
you.	second person pronouns such as "you", "your".
we.	first person plural pronouns such as "we", "our".
dash.	dashes (punctuation).
focuspresent.	words related to present focus such as "today", "now".
male.	male references such as "his", "boy".
female.	female references such as "her", "girl".
family.	words such as "dad", "daughter".