Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #48016

PEER HELPER (Peer Engagement for Effective Reflection, Holistic Engineering Learning, Planning, and Encouraging Reflection) Automated Discourse Analysis Framework

yilin zhang, University of Florida
Dr. Bruce F. Carroll, University of Florida

Dr. Carroll is an Associate Professor of Mechanical and Aerospace Engineering at the University of Florida. He holds an affiliate appointment in Engineering Education. His research interests include engineering identity, self-efficacy, and matriculation of Latine/x/a/o students to graduate school. He works with survey methods and overlaps with machine learning using quantitative methods and sequential mixed methods approaches.

Jinnie Shin, University of Florida
Dr. Kent J. Crippen, University of Florida

Kent Crippen is a Professor of STEM education in the School of Teaching and Learning at the University of Florida.

PEER HELPER (Peer Engagement for Effective Reflection, Holistic Engineering Learning, Planning, and Encouraging Reflection) Automated Discourse Analysis Framework

Abstract

As peer mentoring increasingly complements professional advising in academic settings, ensuring effective mentor training remains challenging, particularly due to high turnover rates from student graduation. This study introduces a Talk-Move Framework that leverages Transformer models, specifically RoBERTa, to automate discourse analysis in peer mentor-mentee interactions. We call this framework PEER HELPER: Peer Engagement for Effective Reflection, Holistic Engineering Learning, Planning, and Encouraging Reflection. Building on established mentoring theories, our analysis framework categorizes dialogues into five key areas: Goal Setting and Planning, Problem Solving and Critical Thinking, Understanding and Clarification, Feedback and Support, and Exploration and Reflection. Using annotated mentoring data from the University of Florida and pre-trained insights from the DSTC7 dataset, the RoBERTa-based model achieved high classification performance, with an accuracy of 98.2%, an F1-score of 0.982, precision of 0.982. These results demonstrate the model's potential to accurately and systematically analyze mentoring dialogues, providing a reliable foundation for further development of AI-powered mentor training tools.

keywords

Discourse Analysis, Peer Mentoring, RoBERTa, Talk-Move Framework, Transformer

1 Introduction

Peer mentoring, where one person (i.e., the mentor) provides practical advice to the other (i.e., the mentee) given that they both are similar in age and share characteristics or experiences, has emerged as a cornerstone of engineering education, providing crucial academic, career, and emotional support to students navigating complex technical curricula. In engineering programs specifically, where students face challenging coursework and professional development requirements, peer mentors serve as invaluable guides who can relate to and support their peers through shared experiences. A comprehensive review of undergraduate mentoring programs has demonstrated that well-structured peer mentoring initiatives consistently yield positive outcomes across multiple domains [1]. In engineering education, these benefits include enhanced academic performance, strengthened leadership development, formation of engineering identity, and more effective career planning [2]. This peer-to-peer support system has proven particularly effective in helping students transition through different stages of their engineering education, from foundational courses to specialized technical subjects [3].

Despite their demonstrated importance, peer mentoring programs face several critical challenges that limit their effectiveness. Current programs often operate without standardized tools for analyzing mentor-mentee interactions, making it difficult to evaluate conversation effectiveness and identify areas for improvement [4]. High turnover rates due to student graduation create a constant need for effective mentor training, yet existing methods lack standardization and scalability [5]. Furthermore, the assessment of mentoring quality often remains subjective, lacking quantitative metrics that could guide program improvement and mentor development. The time-intensive nature of traditional mentor training and evaluation methods also limits programs' ability to scale while maintaining quality [6].

The Talk-Move Framework [7], originally designed for classroom teaching [8], offers a promising solution to these challenges through its structured approach to identifying the pedagogical roles of educational conversations. The framework provides a systematic way to analyze mentor-mentee interactions through well-defined categories [9]. The framework's adaptability makes it well-suited for engineering education mentoring, addressing both technical problem-solving and professional development aspects [10]. To address the scalability challenge, we propose integrating AI-supported implementation, specifically using transformer models, to enable the analysis of large volumes of mentoring conversations [11]. This automated approach ensures reliable evaluation criteria across different mentor-mentee pairs.

Our study introduces a Talk-Move Framework (named PEER HELPER: Peer Engagement for Effective Reflection, Holistic Engineering Learning, Planning, and Encouraging Reflection) specifically tailored for peer mentoring interactions and utilizes RoBERTa-based model to automate and scale dialogue classification. The study aims to develop and validate the framework, establish quantitative metrics for evaluating mentoring effectiveness, and demonstrate its application in engineering education.

This research makes several significant contributions. It represents a novel adaptation of the Talk-Move Framework from classroom contexts to personalized mentoring scenarios in engineering education. It combines publicly available datasets for pre-training with real-world, small-scale engineering mentoring data for testing and refinement. Additionally, the study demonstrates the development of specialized AI tools for automated analysis of mentoring conversations, incorporating engineering-specific context and terminology. These contributions support the sustainability and scalability of peer mentoring initiatives, enhancing their effectiveness and impact across educational institutions.

2 Literature Review

Peer mentoring plays a pivotal role in engineering education, with substantial research highlighting its contributions to academic, professional, and emotional development. However, current mentoring practices face significant challenges that hinder their effectiveness. This section explores the existing research on peer mentoring in engineering education, the challenges these programs face, and how the Talk-Move Framework provides a systematic solution to these issues.

2.1 Peer Mentoring in Engineering Education

Peer mentoring has been widely recognized for improving mentees' academic, professional, and emotional outcomes. Through fostering self-efficacy, goal-setting, and interpersonal skills, peer mentoring contributes to student success in demanding engineering programs. It enhances students' confidence in navigating complex curricula, managing workloads, and preparing for future career challenges. Research has consistently demonstrated the positive impact of peer mentoring in promoting leadership development, academic achievement, and engineering identity formation [2].

Despite its benefits, current mentoring practices face significant obstacles. Many programs lack structured tools to analyze and standardize mentor-mentee interactions. This absence of systematic analysis makes it difficult to ensure consistency across mentoring relationships or to train mentors effectively. Moreover, there is limited research investigating how specific mentoring conversations directly influence mentee outcomes. This lack of insight into the mechanisms driving effective mentorship leaves a critical gap in existing literature [1].

This study addresses these gaps by introducing a Talk-Move Framework adapted specifically for peer mentoring in engineering education. The framework enables systematic analysis of mentor-mentee conversations, offering insights into how these dialogues contribute to student development and mentorship effectiveness.

2.2 Self-Regulated Learning and Help-Seeking Behavior

Self-regulated learning (SRL) theory provides a comprehensive framework for understanding how students manage their own learning processes, particularly in challenging academic environments like engineering education. SRL encompasses the cognitive, metacognitive, behavioral, motivational, and emotional aspects of learning [12]. In engineering education, where students face complex technical curricula and professional development requirements, the ability to self-regulate becomes particularly crucial for academic success.

Help-seeking behavior emerges as a critical strategy within the self-regulated learning framework, representing students' ability to identify when they need assistance and effectively obtain it [13]. This behavior is particularly relevant in engineering education, where the technical complexity of coursework often necessitates seeking guidance from peers, mentors, or instructors. Help-seeking can be viewed as a proactive self-regulatory strategy that successful students employ to overcome academic challenges and enhance their learning outcomes [14].

The relationship between help-seeking behavior and academic advising is particularly significant in the context of peer mentoring. Research has demonstrated that academic advising fundamentally operates as a structured form of help-seeking, where students actively engage in obtaining guidance for academic, career, and personal development [15]. This connection provides a theoretical foundation for understanding why peer mentoring programs are effective: they create an accessible and non-threatening environment for help-seeking behavior, allowing students to develop and practice self-regulatory strategies.

This theoretical framework particularly illuminates the mechanisms through which peer mentoring supports student success in engineering education [16]:

1. Goal Setting and Planning: Self-regulated learning emphasizes the importance of setting appropriate goals and developing strategies to achieve them, which directly aligns with the mentoring process of helping students plan their academic and career paths.
2. Metacognitive Development: Through interactions with peer mentors, students develop metacognitive awareness of when and how to seek help effectively, a crucial component of self-regulated learning.
3. Strategic Help-Seeking: Peer mentoring provides a structured environment for students to practice and refine their help-seeking strategies, moving from dependent to independent problem-solving approaches.

Understanding peer mentoring through the lens of self-regulated learning and help-seeking behavior provides valuable insights for program design and implementation. This theoretical framework suggests that effective mentoring programs should not only provide direct assistance but also help students develop the self-regulatory skills necessary for long-term academic success.

2.3 PEER HELPER Talk-Move Framework in Educational Settings

The PEER HELPER Talk-Move Framework provides a structured approach to categorizing educational conversations. Originally developed for classroom teaching, the framework organizes dialogue into five categories that align closely with key aspects of self-regulated learning and help-seeking behaviors. Through this theoretical lens, each category of the framework represents different facets of self-regulatory processes in mentoring interactions [12-16].

The five categories of our framework demonstrate clear connections to self-regulated learning processes:

1. Goal Setting and Planning reflects the forethought phase of self-regulated learning, where students establish objectives and develop strategies. In mentoring contexts, this category captures how mentors guide mentees through the crucial self-regulatory process of setting achievable goals and creating actionable plans.
2. Problem-solving and Critical Thinking aligns with the performance phase of self-regulation, where learners actively engage in strategy implementation and monitoring. This category encompasses the strategic help-seeking behaviors that engineering students employ when facing academic challenges.
3. Understanding and Clarification represents the initial stage of adaptive help-seeking, where students identify knowledge gaps and articulate their needs. This category directly connects to the self-awareness aspect of self-regulated learning, capturing how students recognize when and what type of help they need.

4. Feedback and Support corresponds to the social aspects of self-regulated learning, highlighting how external guidance helps students develop their self-regulatory capabilities. This category is particularly crucial in engineering education, where complex technical concepts often require structured support for effective learning.
5. Exploration and Reflection aligns with the self-reflection phase of self-regulated learning, where students evaluate their progress and adjust their approaches. This category captures how mentoring dialogues facilitate the development of metacognitive skills essential for self-regulated learning.

The framework's effectiveness in analyzing help-seeking behavior stems from its ability to capture both the procedural and metacognitive aspects of mentoring interactions. It provides a systematic way to examine how students articulate their needs (Understanding and Clarification), receive guidance (Feedback and Support), and develop independent problem-solving strategies (Problem Solving and Critical Thinking). This structured analysis is particularly valuable in engineering education, where effective help-seeking behaviors are crucial for navigating complex technical curricula.

Moreover, the framework's categories naturally align with the developmental progression of self-regulated learning skills. As students move from dependent to independent help-seeking, their interactions typically progress from primarily Understanding and Clarification to increased engagement in Problem Solving and Exploration and Reflection. This progression mirrors the development of self-regulatory capabilities that engineering education aims to foster.

These theoretical connections between the PEER HELPER discourse analysis framework and self-regulated learning provide a robust foundation for analyzing and improving mentoring interactions. While the framework has proven effective in classroom contexts [8], its application to peer mentoring requires careful validation to ensure it effectively captures the unique dynamics of mentor-mentee interactions, particularly in the context of engineering education where both technical competence and self-regulatory skills are essential for success.

2.4 Research Questions

This study focuses on two key research questions:

1. **How can a Talk-Move Framework be developed to categorize mentor-mentee dialogues effectively?**
    - The framework must enable reliable and consistent categorization of statements into well-defined categories, allowing systematic analysis of mentoring conversations.
    - The framework must capture the full range of dialogue types commonly occurring in mentor-mentee interactions, ensuring comprehensive coverage of mentoring dynamics.

2. **To what degree can RoBERTa accurately classify dialog statements based on the Talk-Move Framework?**
    - This question evaluates whether RoBERTa, a Transformer-based AI model, can accurately detect the aspects of dialog statements and categorize them within the framework.
    - Acquiring a classification model with a high accuracy is critical for applying the framework at scale and ensuring consistency across diverse mentoring interactions.

By addressing these research questions, this study aims to establish a validated framework for analyzing peer mentoring dialogues and develop reliable AI-driven tools for automated analysis. Future research will build upon this foundation to investigate how the automated dialogue analysis system can be used to improve mentor training effectiveness and mentee outcomes.

2.5 Talk-Move Framework Structure and Categories

While self-regulated learning theory provides the overarching theoretical framework for our research, the specific categories within the PEER HELPER Framework are further supported by established research in mentoring and educational development. These theoretical perspectives complement the self-regulated learning framework by providing detailed insights into specific aspects of mentoring. The framework is designed to analyze the dynamic nature of mentoring interactions, with its structural foundation deeply rooted in established educational and developmental theories. A talk move consists of an initial statement by one participant and a corresponding response from another, which may fall within the same category or transition between different categories.

This framework integrates several key theoretical perspectives from mentoring research: Our Goal Setting and Planning category is grounded in Lent et al.'s [17] social cognitive model of engineering career choice, emphasizing the importance of structured support in academic and career planning. The Problem Solving and Critical Thinking category builds on Colvin and Ashman's [18] research on peer mentoring roles, particularly focusing on the learning coach role in developing analytical skills.

The Understanding and Clarification category draws from Young and Cates' [19] work on effective listening and communication in mentoring relationships. The Feedback and Support category is informed by Leidenfrost et al.'s [20] research on mentoring styles and their impact on academic success. Finally, the Exploration and Reflection category incorporates principles from Bunting and Williams' [21] study of transformative experiences in peer mentoring, emphasizing the importance of personal narrative and self-reflection in mentoring relationships.

Though empirical research does not prescribe exact proportions for each category, self-regulated learning theory suggests that effective mentoring should include meaningful representation across forethought (Goal Setting and Planning), performance (Problem Solving and Critical Thinking, Understanding and Clarification), and self-reflection (Exploration and Reflection, Feedback and

Support) phases. Rather than strict equality, a well-distributed pattern without extreme imbalances would theoretically better support comprehensive development of self-regulatory skills. These categorical distributions therefore serve as important signals for assessing the quality and comprehensiveness of mentoring interactions, offering insights into how effectively mentors are supporting all phases of self-regulated learning and help-seeking behavior development.

The adaptation of this theoretically-grounded framework for peer mentoring contexts reflects the unique requirements of these interactions, which often demand more personalized and empathetic approaches than traditional classroom settings. This adaptation particularly emphasizes the developmental aspects of mentoring relationships while maintaining the structural rigor necessary for systematic analysis. as detailed in Table 1. Each category is defined by specific characteristics and illustrated through representative example exchanges drawn from actual mentoring sessions.

Table 1: Talk-Move Framework Categories and Characteristics

Category	Definition	Key Features	Example Exchange
Goal Setting and Planning	Dialogue focused on establishing specific academic or career objectives and developing concrete steps.	• Future goals/timelines • Specific action steps • Planning strategies	M: Let's plan your course selection. S: I want to align with aerospace industry.
Problem Solving and Critical Thinking	Exchanges involving analysis of challenges, evaluation of options, and solution development.	• Problem analysis • Solution evaluation • Reasoning process	M: What approaches have you considered? S: I'm thinking of using Pomodoro.
Understanding and Clarification	Dialogue seeking to clarify concepts, requirements, or situations.	• Information seeking • Explanation requests • Concept clarification	S: How do prerequisites work? M: They ensure foundational knowledge.
Feedback and Support	Exchanges providing constructive feedback or emotional support.	• Reinforcement • Constructive criticism • Empathetic responses	M: Great progress on study habits. S: Thanks, working hard on it.
Exploration and Reflection	Dialogue promoting deeper thinking about experiences and decisions.	• Open-ended questions • Self-assessment • Reflective discussions	M: How does this align with goals? S: Need more practical projects.

Note: M = Mentor; S = Student/Mentee. Examples are abbreviated for space.

2.5.1 Framework Development and Validation

The adaptation of this framework for peer mentoring contexts involved several key considerations including category flexibility, where statements may exhibit characteristics of multiple categories; response dynamics, where mentor and mentee statements may fall into different categories; and context sensitivity, addressing unique aspects of peer mentoring such as emotional support and stress management in the Feedback and Support category.

Validation Process:

To ensure reliable categorization, we implemented a systematic validation process through inter-rater reliability assessment and category transition analysis. Two investigators independently categorized 200 randomly selected sample dialogues. Inter-rater reliability was measured using Cohen's Kappa coefficient, achieving $\kappa = 0.82$ after iterative refinements. Category definitions were iteratively refined based on areas of low agreement until reaching this sufficient reliability level. Initial analysis of dialogue pairs revealed that approximately 40% maintained category consistency, while 60% demonstrated natural category transitions, reflecting the dynamic nature of the conversations.

Example of Category Transition:

- **Starting in Goal Setting:** - Mentee: I want to secure an internship in robotics next summer. - Mentor: That's a great goal. Have you started looking at specific companies?
- **Transitioning to Problem Solving:** - Mentee: Yes, I've identified three companies, but I'm not sure how to approach them. - Mentor: Let's develop a strategy for each company. What information do you need to prepare?

2.5.2 Framework Application

The Talk-Move Framework serves as a practical and versatile tool for mentoring analysis and program enhancement. It provides a structured approach for categorizing mentoring conversations, enabling systematic evaluation of their effectiveness and creating a foundation for automated dialogue analysis using RoBERTa. Additionally, it supports the development of targeted mentor training programs by identifying effective communication patterns and areas for improvement. The clear definition and validation of categories are critical for ensuring both human classification accuracy and the success of automated analysis. As shown in Figure 1, the Peer Mentor Talk-Move Framework visually represents example dialogues categorized under each Talk-Move category, illustrating its capability to analyze and improve mentoring interactions systematically and at scale. This framework ultimately enhances training and program development in engineering education.

3 Methodology

3.1 Research Context and Participants

This study focuses on mentoring conversations conducted within the University of Florida's (UF) Mechanical and Aerospace Engineering Department. Peer advisors assist students in addressing challenges related to academic planning, problem-solving, and emotional support. Conversations were recorded during structured peer advising sessions and simulated mentoring exercises to ensure a diverse range of interaction topics. Currently, the dataset comprises 668 dialogue exchanges, totaling 1,336 individual utterances collected from these sessions. Additional data

Peer Advisor Talk-Move Framework		
Talk-Move Category	**Mentee**	**Mentor**
Goal Setting and Planning	"I want to improve my grades, but I'm anxious about the workload." "What do employers look for beside grades? I'm not sure how to build my resume."	"Balancing workload and grades can be tricky. What's your strategy to do this?" "Thinking about building a resume is a good approach. What activities do you want to get involved in?"
Problem Solving and Critical Thinking	"I've been skipping class. Going makes me feel nervous. I get stressed out and stay home." "My tracking gpa is low and I'm not sure I can raise it."	"UF can be a pressure cooker. Are you familiar with the counseling and wellness center?" "The staff advisors are the best people to talk with about tracking gpa problems. We can make an appointment for you to see them."
Understanding and Clarification	"This is a tough semester and I want to drop a class." "I've heard that club doesn't have any women members. Do you think that's true?"	"What is making this semester tough? Is it just the courses or are other things coming up?" "Some clubs seem to attract more women than others. Have you had experiences where you were treated differently than the boys?"
Feedback and Support	"I've been thinking about joining the rocket team. Do you think I can handle that with these courses?" "I hear Dr. Hard is really tough. What have you heard? I'm nervous about taking EML4xxx from her."	"Joining the rocket team this semester makes sense. You will enjoy that group. What are your long-term goals?" "Some students like Dr. Hard and others don't. But that's true of most teachers. If you postpone taking that class what does that do to your overall study plan?"
Exploration and Reflection	"I haven't had any luck finding an internship. I wonder if my interview skills could use some help." "I'm thinking about going to graduate school but don't really know much about research."	"There are various resources at the CCC including mock interviews. What do you think about signing up for one of those?" "Graduate school can be a great option. What aspect of the major are you most interested in."

Figure 1: Peer Mentor Talk-Move Framework

collection is ongoing to further expand the study's scope and enhance the framework's validation. All conversations were recorded with informed consent, anonymized during transcription, and securely stored in compliance with ethical guidelines.

3.2 Data Collection Process

The mentoring sessions encompass three primary types of interactions that map onto our five-category Talk-Move Framework in different ways:

- **Academic Planning:** Students sought guidance on course selection, workload management, and long-term career goals. These interactions typically map to the Goal Setting and Planning category when focusing on objective-setting, and to the Feedback and Support category when providing guidance on academic decisions.
- **Problem-Solving:** Sessions involved identifying solutions to academic challenges and fostering critical thinking skills. These discussions primarily align with the Problem Solving and Critical Thinking category, often incorporating elements of Understanding and Clarification as mentors guide mentees through complex issues.
- **Emotional Support:** Discussions addressing stress management and building self-efficacy, which commonly map to both the Feedback and Support category through encouragement and the Exploration and Reflection category through guided self-assessment.

It is important to note that these interaction types are not discrete categories but rather common themes that can involve multiple framework categories within a single conversation. Our five-category Talk-Move Framework (shown in Figure 1) was developed to capture these dynamic interactions in a more structured way. During the annotation process, conversations were systematically categorized according to this framework, with categories being iteratively refined to ensure comprehensive coverage of all mentoring interaction patterns.

To enhance our model's training capabilities, we supplemented the UF dataset with the Dialog System Technology Challenges Task 1 (DSTC7) dataset [22]. The DSTC7 Task 1 dataset consists of goal-oriented dialogues collected from human-human conversations, including academic advising interactions from the University of Michigan. This dataset provides approximately 300,000 conversations that align well with our Talk-Move Framework's categories, particularly in academic advising contexts. Below are the label distributions for the training and test datasets from DSTC7.

Table 2: DSTC7 Training Dataset Label Distribution

Category	**Count**	**Percentage (%)**
Goal Setting and Planning	249,909	83.30
Feedback and Support	45,495	15.17
Problem Solving and Critical Thinking	2,757	0.92
Exploration and Reflection	351	0.12
Understanding and Clarification	1,488	0.50

Table 3: DSTC7 Test Dataset Label Distribution

Category	**Count**	**Percentage (%)**
Goal Setting and Planning	1,248	83.20
Feedback and Support	210	14.00
Problem Solving and Critical Thinking	6	0.40
Exploration and Reflection	12	0.80
Understanding and Clarification	24	1.60

As shown in Tables 2 and 3, the DSTC7 dataset exhibits significant class imbalance, with Goal Setting and Planning accounting for over 80% of the data, and categories such as Exploration and Reflection and Problem Solving and Critical Thinking being severely underrepresented. This imbalance underscores the importance of incorporating additional UF-specific data to ensure balanced representation of all framework categories. By addressing this imbalance, we aim to improve the model's ability to generalize across all dialogue categories and accurately reflect the diverse nature of real-world mentoring interactions.

We utilized this comprehensive dataset for pre-training our model, with the intention of transitioning to an exclusively UF-based dataset in future iterations as more local data is collected.

3.3 Analysis Framework and Plan

Our analysis framework consists of three main components: data preprocessing and feature extraction, model construction, and evaluation. Figure 2 provides an overview of this comprehensive pipeline.

3.3.1 Data Preprocessing and Feature Extraction

The preprocessing of dialogue data involves both manual and automated steps. Initial data annotation was performed manually by expert raters to establish ground truth labels for a subset of the data, ensuring accurate category assignment based on our framework definitions. This manually labeled dataset serves as both training data and a benchmark for evaluating automated classification performance.

The automated preprocessing includes text normalization and cleaning, tokenization using RoBERTa's built-in tokenizer, and sentence segmentation to identify complete dialogue units. Feature extraction incorporates linguistic features including sentence structure and dialogue patterns, semantic features utilizing RoBERTa's contextual embeddings, and structural features analyzing dialogue flow patterns.

3.3.2 Model Construction and Classification Strategy

Our classification approach leverages RoBERTa [23] for dialogue categorization, assigning each dialogue segment to one of the five predefined Talk-Move categories. We selected RoBERTa over other transformer models due to its enhanced training methodology and superior performance in contextual understanding tasks. Specifically, RoBERTa's dynamic masking, larger batch sizes, and longer training on more data make it particularly well-suited for analyzing the nuanced language patterns in mentoring dialogues. Additionally, its robust performance on sentence-level tasks and proven effectiveness in educational dialogue analysis [24] aligns well with our framework's requirements.

This single-label classification strategy was adopted to ensure clear and precise categorization of mentoring interactions, facilitating straightforward analysis of dialogue patterns and transitions. The approach extends beyond simple keyword matching by incorporating advanced linguistic features and context analysis to determine the primary category of each statement. Specifically, it utilizes contextual word embeddings to capture the semantic meaning of words within their context, leveraging RoBERTa's pre-trained language representations. The model analyzes sentence-level structure to understand the overall semantic structure and intent, considers sequential dialogue patterns to comprehend the flow and intent of interactions, and accounts for the broader dialogue context to enhance classification accuracy.

Classification Strategy and Model Development:

Our selected approach integrates contextual embeddings with structural features, demonstrating superior performance compared to simpler methods such as keyword-based classification or pure embedding-based classification. The model parameters were selected based on established practices in transformer-based dialogue classification [25], with learning rate at 2e-5, batch size at 8, and training epochs fixed at 10.

Evaluation Framework: Our evaluation integrates both technical metrics (e.g., accuracy, precision, recall, and F1-score) and quality indicators measuring engagement level, responsiveness, cognitive depth, and emotional atmosphere of dialogues.

The iterative refinement loop, as illustrated in Figure 2, ensures continuous performance improvement by addressing errors identified during evaluation. This phase employs feedback annotations to fine-tune the model further and enhance its overall effectiveness in dialogue classification.

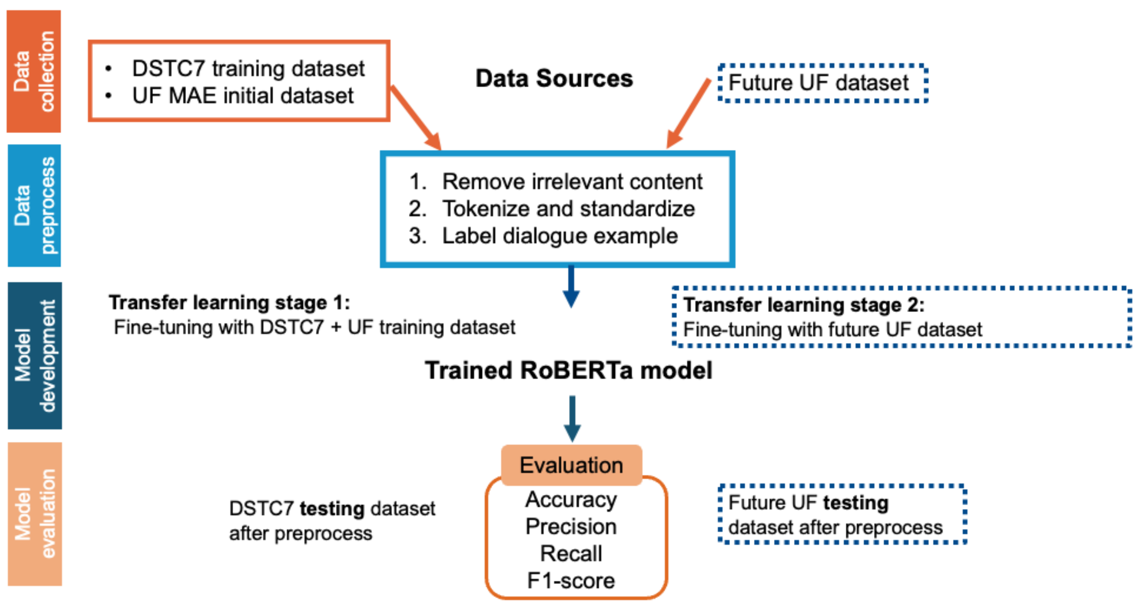

Figure 2: Two-Stage Transfer Learning Pipeline for Mentoring Dialogue Analysis: Fine-tuning RoBERTa with DSTC7 and UF Data

4 Results

Our RoBERTa model's performance in classifying dialogue segments across the five Talk-Move Framework categories was evaluated using the commonly adopted prediction model evaluation metrics [26]. These metrics provide a comprehensive evaluation of the model's ability to handle diverse dialogue categories, especially in the context of an imbalanced dataset.

Classification Accuracy measures the proportion of correctly classified instances, while **Precision** evaluates the reliability of predictions for a specific category. **Recall** assesses the model's ability to identify all relevant instances within a category, and **F1-score** balances precision and recall to provide a single performance metric. Table 4 presents the detailed performance metrics for each category.

The model achieved an overall accuracy of 98.2%, with high performance across well-represented categories such as *Feedback and Support* and *Goal Setting and Planning*. These categories demonstrated near-perfect precision, recall, and F1-scores, reflecting the model's ability to leverage sufficient training examples effectively. However, the results also reveal significant disparities in performance for underrepresented categories. For instance, the *Understanding and Clarification* category, which comprises only 1.6% of the dataset, exhibited notably lower precision (66.7%) and F1-score (36.4%), highlighting the challenges posed by data imbalance.

The imbalanced label distribution across categories biased the model toward majority classes, such as *Goal Setting and Planning*, which accounts for over 83% of the dataset. Consequently,

Table 4: Performance Metrics for Dialogue Classification (Test Dataset)

Category Name (% of total)	Accuracy	Precision	Recall	F1-score
Goal Setting and Planning (83.2%)	0.998	0.990	0.998	0.994
Feedback and Support (14.0%)	0.996	0.958	0.972	0.972
Problem Solving and Critical Thinking (0.4%)	1.00	1.00	0.750	1.00
Exploration and Reflection (0.8%)	0.750	0.750	0.750	0.750
Understanding and Clarification (1.6%)	0.250	0.667	0.250	0.364

minority classes, such as *Exploration and Reflection*, received less representation during training, limiting the model's ability to generalize effectively for these categories.

While the model demonstrates strong overall potential for dialogue classification, addressing data imbalance remains critical for ensuring equitable performance across all categories. Future work will focus on expanding the UF dataset to include more examples of underrepresented categories, along with employing strategies such as oversampling, data augmentation, or adaptive loss functions to improve performance consistency across all framework categories.

4.1 Label Category Distribution

The label category distribution reveals distinct patterns in the dialogue focus of UF mentors and mentees and serves as an indicator of the type and, potentially, quality of mentoring that occurred. Mentors primarily concentrated their contributions in Understanding and Clarification (39.13%), Goal Setting and Planning (28.34%), and Feedback and Support (22.79%), with comparatively lower involvement in Problem Solving and Critical Thinking (6.15%) and Exploration and Reflection (3.60%). This trend suggests that mentors prioritize clarifying concepts, setting clear objectives, and offering emotional support, while engaging less frequently in analytical or reflective dialogue.

In contrast, mentee dialogue is heavily weighted toward Understanding and Clarification (57.72%), underscoring their dominant role in seeking clarity and articulating challenges, indicating emergence of help-seeking behavior related to self-regulated learning. The remaining contributions are distributed across Goal Setting and Planning (15.89%), Feedback and Support (11.24%), Exploration and Reflection (9.45%), and Problem Solving and Critical Thinking (5.70%). This distribution indicates that while mentees are primarily focused on understanding, they also demonstrate meaningful participation in goal-setting, reflection, and feedback reception.

These distributions highlight the complementary yet asymmetrical nature of mentoring roles. Mentors balance multiple functions—guidance, structure, and support—whereas mentees act chiefly as help-seekers. Interestingly, both groups allocate nearly equivalent proportions of their dialogue to Problem Solving and Critical Thinking (mentors: 6.15%, mentees: 5.70%), suggesting moments of shared analytical engagement. Moreover, the greater representation of Exploration and Reflection in mentee contributions (9.45% vs. 3.60%) implies that mentees may initiate reflective inquiry more readily when space is provided within the conversation.

Figure 3 visualizes these patterns, particularly emphasizing the dominance of clarification-oriented dialogue and the relative scarcity of deep reflection or problem-solving exchanges.

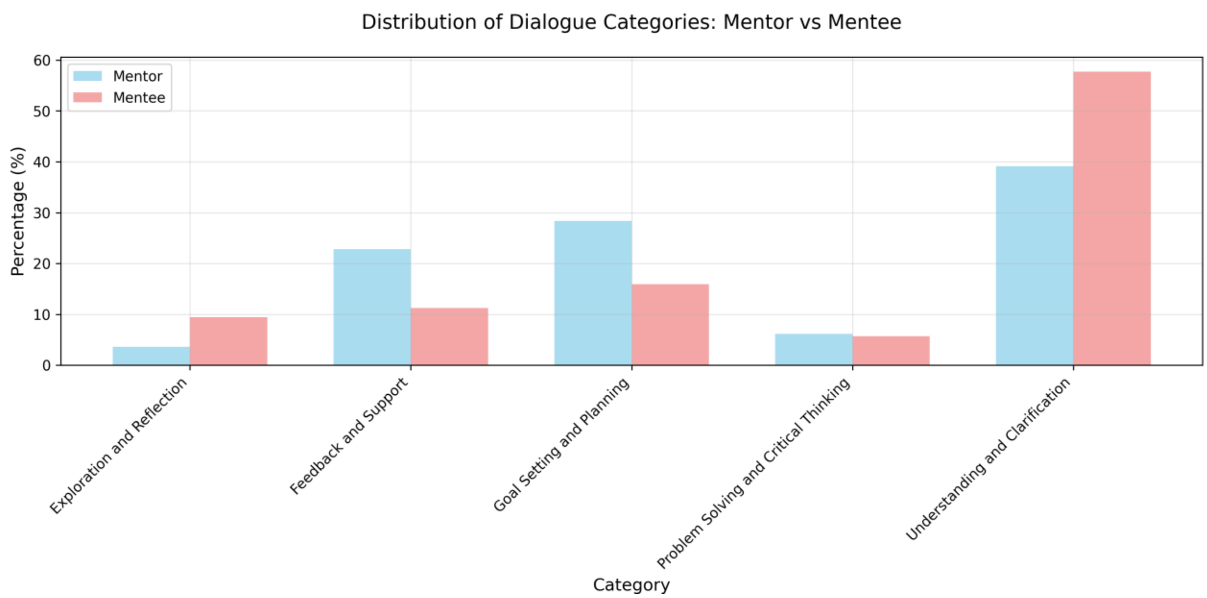

Figure 3: Mentor and Mentee Dialogues Classification

4.2 Dialogue Matching Analysis

The comparative analysis between the DSTC7 and UF datasets focused on dialogue alignment patterns – a critical aspect of mentoring effectiveness. We define dialogue alignment as the extent to which a mentor's statement and the corresponding mentee's response fall within the same Talk-Move category. This analysis serves multiple purposes in understanding and improving mentoring quality: it helps identify whether mentors maintain appropriate balance between consistency and flexibility in their responses, reveals patterns in conversation flow, and provides insights for mentor training.

Examining dialogue alignment is particularly important because both excessive consistency and excessive variation can indicate potential issues in mentoring approaches. High levels of category matching might suggest a mentor is not adapting their approach to the mentee's changing needs, while frequent category transitions might indicate a lack of focused guidance. For example, consistently remaining in the problem-solving category might miss opportunities for emotional support or reflection, while constant category shifts might fail to provide the sustained attention needed to address specific issues effectively.

As shown in Figure 4, there are notable differences in dialogue matching patterns between the two datasets. In the DSTC7 dataset, approximately 67% of dialogues showed category matching, with 33% displaying category transitions. In our expanded UF dataset, 41.08% (274 exchanges) demonstrated category matching, while 58.92% (393 exchanges) showed category transitions.

This represents a significant shift from our earlier analysis of a smaller UF dataset subset, which had indicated only 20% matching and 80% transitions.

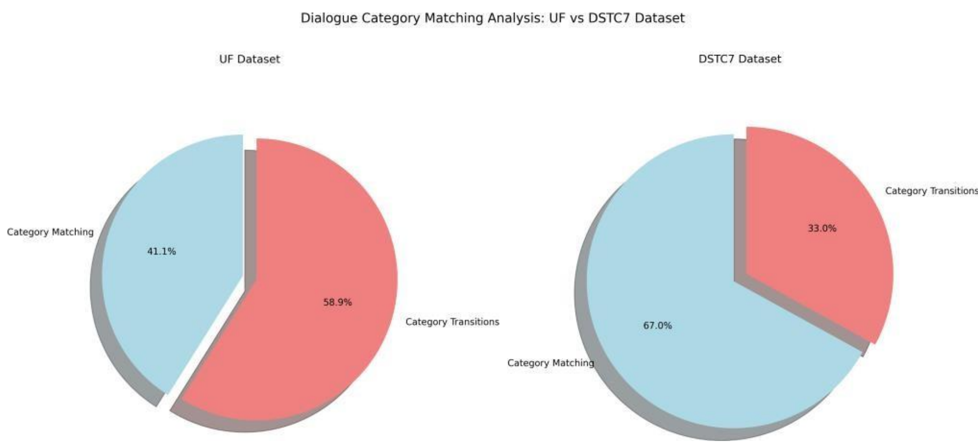

Figure 4: Mentor and Mentee Label Matching Distribution Across DSTC7 and UF Datasets

Further analysis of the UF dataset reveals important transition patterns in mentoring dialogues. The most frequent patterns include:

- Understanding and Clarification → Understanding and Clarification (26.69%): The most common matched pattern, indicating extended exchanges focused on information clarification
- Goal Setting and Planning → Understanding and Clarification (14.99%): A transition reflecting how goal-setting often prompts questions and clarification needs
- Feedback and Support → Understanding and Clarification (11.24%): Showing how supportive feedback frequently leads to further information seeking
- Goal Setting and Planning → Goal Setting and Planning (6.75%): Sustained focus on planning and objective development

- Understanding and Clarification → Goal Setting and Planning (5.40%): Illustrating how clarification often leads to concrete planning

These findings suggest a more balanced approach to dialogue consistency in the UF mentoring interactions than previously observed. The DSTC7 dataset still shows a stronger tendency toward category matching (67%), which might indicate more structured but potentially less flexible interactions. The UF mentoring sessions now demonstrate a more moderate distribution between matching (41.08%) and transitions (58.92%), suggesting a balance between thematic consistency and adaptive responsiveness.

This balance allows for both sustained focus on specific topics (through matching patterns) and adaptability to mentee needs (through category transitions). The predominance of transitions to and from Understanding and Clarification highlights the centrality of information exchange and clarification in effective mentoring relationships.

Future research will explore the relationship between specific dialogue matching patterns and mentoring effectiveness, identifying optimal transition sequences that support mentee development while maintaining coherent conversation flow.

4.3 Dialogue Length Analysis

Using the UF mentoring dataset, we analyzed dialogue length variability across Talk-Move categories to uncover how interaction types and speaker roles influence contribution patterns.

The analysis revealed distinct patterns in dialogue length between mentors and mentees. Mentor utterances were generally longer in *Goal Setting and Planning*, *Problem Solving and Critical Thinking*, and *Exploration and Reflection*—categories requiring detailed guidance, elaboration, or nuanced feedback. In particular, *Problem Solving and Critical Thinking* displayed the highest median length among mentors, reflecting the complexity of analytical discourse. *Goal Setting and Planning* also elicited extended mentor responses, often including step-by-step strategies and examples tailored to academic or career goals.

Mentee contributions followed a different distribution. Their longest utterances occurred in *Problem Solving and Critical Thinking* and *Exploration and Reflection*, suggesting deeper involvement when articulating challenges, exploring ideas, or reflecting on decisions. Conversely, mentee dialogue in *Understanding and Clarification* and *Feedback and Support* was notably shorter, often limited to focused questions, confirmations, or acknowledgments.

As illustrated in Figure 5, boxplots depict median length variation by category and speaker. Mentors exhibited relatively even distributions across categories, with moderate variation in interquartile ranges. In contrast, mentee dialogue length was more polarized, highlighting their adaptive engagement based on the nature of the conversation. Outliers—particularly in *Goal Setting and Planning*—underscore instances where either mentors or mentees elaborated at length, suggesting that planning discussions sometimes demand deeper contextualization.

These asymmetrical patterns reflect the complementary roles of mentors and mentees in dialogue. Mentors maintain a consistent depth across categories, while mentees vary their contribution length depending on the function of the exchange. This alignment suggests that while mentors provide structure and support across topics, mentees selectively expand their responses when reflection or problem-solving is encouraged.

Understanding these patterns can guide mentoring practice. For example, mentors might intentionally create space for mentee elaboration in reflective or analytical contexts, while maintaining clarity and conciseness in goal clarification or support exchanges. These findings also support the development of targeted training that adapts dialogue strategies to different conversational objectives.

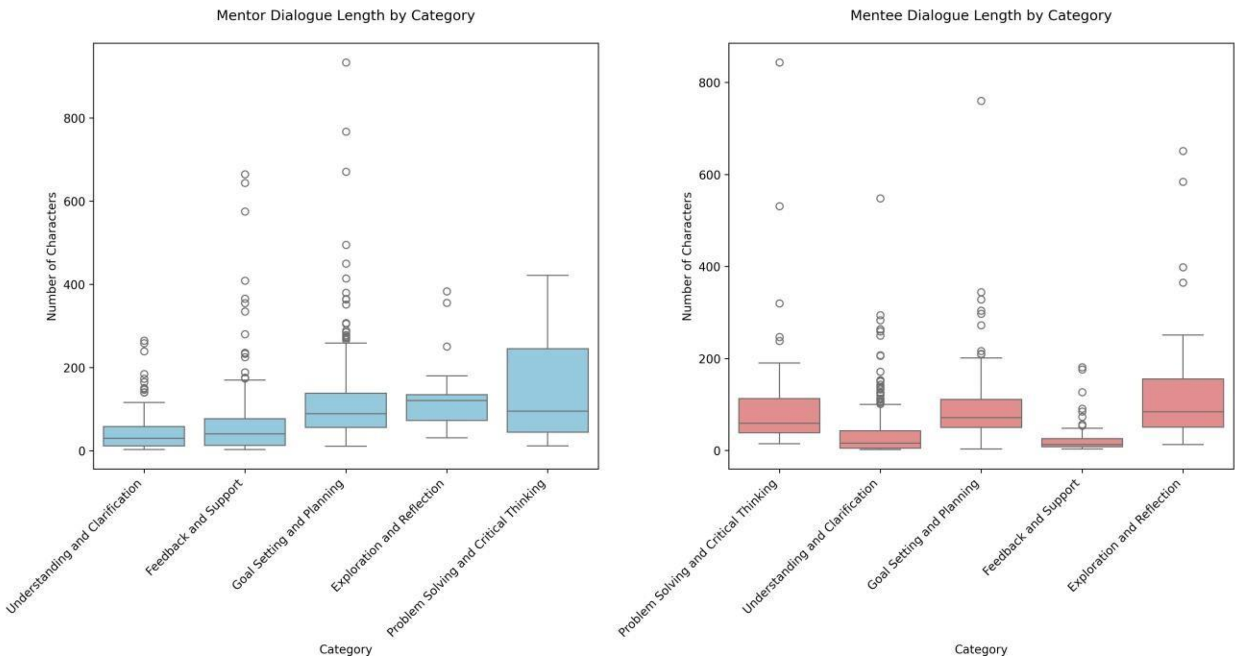

Figure 5: Mentor and Mentee Dialogue Length

4.4 Quality Indicators Analysis

Our analysis framework incorporates four key quality indicators to evaluate mentoring dialogue effectiveness. Engagement Level measures interaction depth by assessing the frequency and quality of exchanges between mentor and mentee. High engagement is characterized by detailed, elaborative responses and active question-asking behavior, while low engagement manifests as brief, minimal responses and limited interaction [27]. Responsiveness evaluates how effectively and promptly participants address each other's needs during the dialogue, with high responsiveness demonstrated through timely, relevant feedback and appropriate follow-up, in contrast to delayed or misaligned responses [28].

The framework also assesses Cognitive Depth, which examines the complexity and sophistication of dialogue content [29]. Conversations demonstrating high cognitive depth feature critical analysis, complex problem-solving approaches, and development of comprehensive solutions, whereas those with low cognitive depth remain at surface-level information exchange. The Emotional Atmosphere of interactions is evaluated through the presence and quality of emotional support elements [30], with positive atmospheres characterized by consistent encouragement, empathy, and supportive language, contrasting with negative atmospheres that lack these emotional support components.

4.4.1 Measurement Framework and Methodology

Each quality indicator is assessed through a systematic approach combining manual evaluation and the potential for automated methods:

- **Engagement Level:** This indicator quantifies mentee participation quality and frequency. Assessment involves expert annotators rating dialogues on a 1-5 Likert scale, examining specific markers such as question frequency, response elaboration, and topic initiation. The evaluation considers both the quantity of participation (frequency of contributions) and quality (depth and relevance of responses). Our framework includes potential for linguistic

  feature analysis, such as tracking response length and question density, to support more efficient assessment in the future.
- **Responsiveness:** This indicator measures how effectively and promptly participants address each other's needs during dialogues. Evaluators assess responsiveness by examining whether responses directly address questions or concerns raised, the timeliness of these responses, and the appropriateness of the guidance provided. This bidirectional metric considers both mentor responsiveness to mentee needs and mentee receptiveness to mentor guidance, creating a comprehensive view of dialogue effectiveness.
- **Cognitive Depth:** Drawing on Anderson and Krathwohl's revised Bloom's Taxonomy, this indicator assesses the complexity and sophistication of dialogue content. Structured rubrics identify evidence of higher-order thinking skills, including analysis, evaluation, and creation. The assessment examines whether conversations demonstrate critical thinking, complex reasoning, and reflective inquiry beyond surface-level information exchange.
- **Emotional Atmosphere:** This indicator evaluates the presence and quality of emotional support elements, examining verbal encouragement, empathetic responses, and supportive language. The assessment focuses on identifying positive reinforcement, expressions of empathy, and the creation of a psychologically safe environment conducive to open communication and learning.

4.4.2 Implementation Process

In future work, the quality indicators assessment will be implemented through a systematic multi-perspective process:

1. **Post-Session Mentor Evaluation:** Following each mentoring session, mentors complete a structured self-assessment form rating their performance on each quality

indicator and providing specific dialogue examples that demonstrate their effectiveness.

2. **Peer Assessment:** Selected mentoring sessions undergo peer review, where experienced mentors evaluate the dialogues using the same quality indicators framework, offering external validation and identifying additional improvement opportunities.
3. **Mentee Feedback:** Mentees complete parallel evaluations assessing their experience across the same quality dimensions, creating a comprehensive view of the mentoring interaction. This triangulated approach ensures balanced assessment.
4. **Data Analysis:** Combined quantitative ratings and qualitative examples undergo both statistical analysis and thematic coding to identify patterns, strengths, and improvement areas, informing both individual mentor development and program-level enhancements.

4.4.3 Connection to Talk-Move Framework

Our theoretical analysis reveals meaningful relationships between quality indicators and Talk-Move categories, demonstrating how these complementary frameworks enhance mentoring assessment:

- *Goal Setting and Planning* effectiveness appears closely linked with Engagement Level, as setting clear, motivating goals often increases mentee participation and commitment in mentoring interactions.
- *Problem Solving and Critical Thinking* dialogues typically demonstrate higher Cognitive Depth due to their analytical nature and focus on complex problem resolution.
- *Understanding and Clarification* quality is directly reflected in Responsiveness metrics, as effective clarification sequences involve timely and relevant exchanges between mentor and mentee.
- *Feedback and Support* demonstrates a strong connection to Emotional Atmosphere, highlighting how effective feedback creates psychological safety and mentee receptivity.
- *Exploration and Reflection* often exhibits strength in both Cognitive Depth and Engagement Level dimensions, as reflective activities typically promote both deeper thinking and active participation.

Table 5: Quality Indicators Assessment Framework

Indicator	Definition	Example Markers
Engagement Level	Measures interaction depth through frequency and length of exchanges	High: Detailed responses, frequent questions Low: Minimal responses, limited interaction
Responsiveness	Evaluates timeliness and relevance of responses	High: Prompt, relevant feedback Low: Delayed or misaligned responses
Cognitive Depth	Assesses complexity of dialogue content	High: Critical analysis, solution development Low: Surface-level information exchange
Emotional Atmosphere	Quantifies emotional support and interaction tone	Positive: Encouragement, empathy Negative: Limited emotional support

This integrated approach enables both robust assessment of individual mentoring interactions and scalable program-level evaluation. By combining the descriptive power of the Talk-Move Framework with the evaluative capacity of quality indicators, we provide a comprehensive system for enhancing mentoring effectiveness in engineering education. Future research will focus on implementing real-time feedback mechanisms based on these indicators and exploring their relationship with engineering student success outcomes.

4.5 Case Studies

The case studies presented here build upon the findings from the Dialogue Matching Analysis, illustrating how the Talk-Move Framework captures both category consistency and transitions in mentoring dialogues. As noted in the Dialogue Matching Analysis section, UF mentoring dialogues exhibit more dynamic interaction patterns compared to DSTC7, with frequent category transitions reflecting the complexity and multifaceted nature of mentor-mentee conversations. By examining real-world cases, we demonstrate how these transitions are managed effectively and highlight the framework's practical application in identifying areas for improving mentoring strategies.

Case Study 1: Goal Setting and Feedback Transition. In this case, the mentee expressed uncertainty about their academic goals, stating: I want to improve my GPA next semester to qualify for internships. The mentor responded with constructive feedback: That's a great goal. Have you considered scheduling study blocks or joining a study group for support?

The mentee's statement was categorized as Goal Setting and Planning, while the mentor's response transitioned to Problem Solving and Critical Thinking. This dynamic reflects the findings from the Dialogue Matching Analysis, where UF conversations frequently shift between categories to address mentees' evolving needs. By moving from goal-setting to problem-solving, the mentor provided actionable steps to help the mentee achieve their objective, demonstrating the flexibility and responsiveness highlighted in UF mentoring interactions.

Case Study 2: Emotional Support in Understanding and Clarification In this case, the mentee shared concerns about stress management: I feel overwhelmed by the course load. I'm not sure if I can manage everything. The mentor offered empathetic reassurance: It's normal to feel that way. Let's break down your workload into smaller steps, so it feels more manageable.

The mentee's statement was categorized as Understanding and Clarification, while the mentor's response transitioned to Feedback and Support. This transition highlights the UF dataset's dynamic interaction patterns, where mentors often shift between categories to address mentees' academic and emotional challenges holistically. This aligns with the Dialogue Matching Analysis, reflecting the framework's capacity to identify and analyze these nuanced interactions.

Figure 6 is Emotional Support Flow provides a timeline-based visualization of the dialogue in Case Study, illustrating how the interaction progresses across Talk-Move categories. The figure highlights the transition from Understanding and Clarification to Feedback and Support, emphasizing the dynamic nature of UF mentoring dialogues.

Insights from Case Studies These case studies underscore the Talk-Move Framework's utility in analyzing and enhancing mentor-mentee interactions. By categorizing statements and responses into well-defined categories and tracking transitions between them, the framework offers actionable insights for mentor training and program development. The dynamic transitions observed in UF mentoring dialogues demonstrate the importance of adapting to mentees' evolving needs, further validating the framework's relevance and practical applicability.

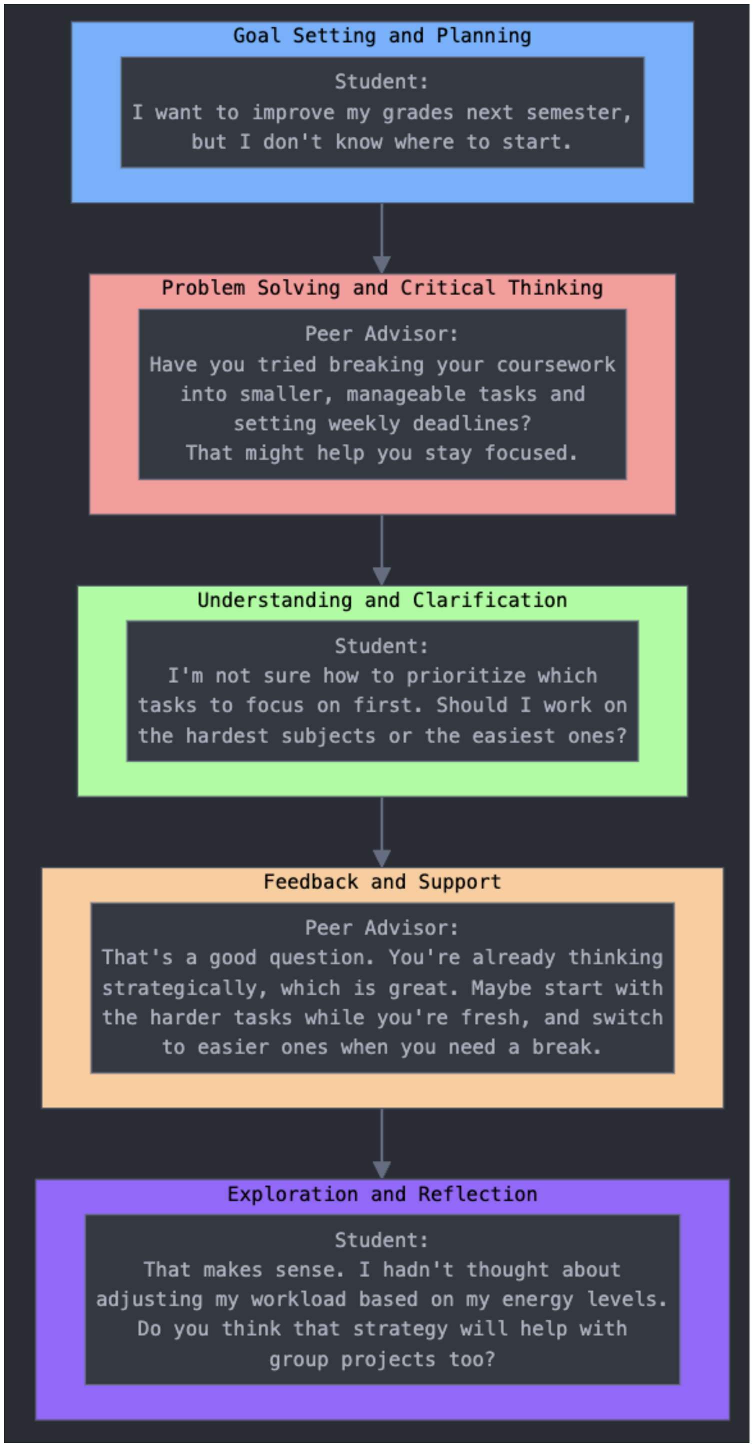

Figure 6: Dynamic Transition between Categories

5 Discussion

5.1 Addressing the Research Questions

This study was guided by two fundamental research questions that explore the development of the Talk-Move Framework and its automated classification using RoBERTa. Our findings provide strong empirical support for both aspects while identifying key areas for further enhancement.

5.1.1 RQ1: Development and Validation of the Talk-Move Framework

The first research question asked: *How can a Talk-Move Framework be developed to categorize mentor-mentee dialogues effectively?* Our study successfully constructed and validated a structured dialogue classification framework by integrating self-regulated learning theory and help-seeking behaviors. The Talk-Move Framework consists of five well-defined categories—*Goal Setting and Planning, Problem Solving and Critical Thinking, Understanding and Clarification, Feedback and Support,* and *Exploration and Reflection*—each with precise operational definitions and linguistic markers. These categories were developed through an iterative annotation process, ensuring clear differentiation and applicability across various mentoring scenarios.

To establish the framework's reliability, we conducted an inter-rater reliability analysis, achieving a Cohen's Kappa of 0.82, indicating strong agreement among annotators. The framework's effectiveness was further demonstrated by its ability to capture key mentoring dynamics, including shifts in mentee help-seeking behaviors and mentor scaffolding strategies. Our findings show that successful mentoring dialogues exhibit clear transitions between categories, often progressing from *Understanding and Clarification* to *Problem Solving and Critical Thinking*, reflecting the mentee's increasing self-regulation. This structured categorization not only enhances the systematic analysis of mentoring conversations but also provides a valuable tool for mentor training and program evaluation.

5.1.2 RQ2: RoBERTa's Performance in Automated Dialogue Classification

The second research question investigated: *To what degree can RoBERTa accurately classify dialogue statements based on the Talk-Move Framework?* Our empirical analysis revealed that the RoBERTa-base model achieved an overall classification accuracy of 98.2%, demonstrating strong performance in identifying and categorizing mentoring dialogues. The model exhibited particularly high precision and recall in well-represented categories such as *Problem Solving and Critical Thinking* (100% F1-score), *Goal Setting and Planning* (99.8%), and *Feedback and Support* (99.6%). However, performance varied for underrepresented categories, with *Understanding and Clarification* achieving a lower F1-score of 36.4%, indicating that data imbalance affected classification accuracy.

Further analysis suggests that while RoBERTa effectively recognizes patterns in structured dialogues, its classification confidence is significantly influenced by the distribution of training samples. Categories with more frequent occurrences, such as *Exploration and Reflection* (75.0%

F1-score), were classified with high reliability, whereas those with fewer training instances showed greater variability. This highlights the need for dataset expansion, particularly for minority classes, to improve model robustness. Additionally, the model's ability to capture dialogue progression suggests its potential for tracking mentee development over time, making it a valuable tool for large-scale mentoring assessment.

Overall, our findings confirm that RoBERTa provides a scalable and efficient method for automated mentoring dialogue analysis. However, addressing dataset limitations through targeted data collection and augmentation remains a critical next step in refining the classification model. The combination of a theoretically grounded framework with AI-driven classification offers promising opportunities for enhancing peer mentoring programs, enabling more structured mentor training, real-time feedback mechanisms, and scalable analysis of mentoring effectiveness in engineering education.

5.2 Framework and Model Effectiveness

The integration of self-regulated learning theory and help-seeking behavior with the PEER HELPER Framework provides valuable insights into the effectiveness of peer mentoring practices. Our analysis reveals how this theoretical foundation magnifies the framework's practical utility in several key aspects:

First, understanding mentoring interactions through the lens of self-regulated learning helps explain why certain dialogue patterns are more effective than others. For instance, when mentors guide mentees through goal-setting processes, they are not merely helping with planning, but actively supporting the development of self-regulatory skills [12]. This theoretical understanding helps mentors move beyond simple advice-giving to fostering independent learning capabilities.

Second, the help-seeking perspective illuminates some mentees benefit more from mentoring than others. Successful mentoring interactions often demonstrate a progression from dependent to strategic help-seeking behaviors [13]. Our framework captures this progression through the transition patterns between categories, particularly from Understanding and Clarification to Problem Solving and Critical Thinking.

Third, in the engineering education context, the theoretical framework helps explain why certain mentoring approaches are particularly effective. The complex technical nature of engineering coursework requires sophisticated help-seeking strategies [14], which our framework helps identify and analyze. This understanding enables more targeted mentor training and support.

5.2.1 Model Performance Analysis

The RoBERTa-based classification model demonstrated varying performance across different categories, achieving an overall accuracy of 98.2%, with a weighted-average F1-score of 0.982 and precision of 0.982. Performance analysis revealed distinct patterns across categories: Problem Solving and Critical Thinking achieved perfect performance (100% across all metrics), while Goal Setting and Planning demonstrated near-perfect results (99.8% accuracy, 99.0% precision, 99.4%

F1-score). Feedback and Support maintained strong performance with 99.6% accuracy, 95.8% precision, and 97.2% F1-score. Exploration and Reflection showed consistent but lower performance at 75.0% across all metrics. The most challenging category was Understanding and Clarification, which exhibited notably lower performance with 25.0% accuracy, 66.7% precision, and an F1-score of 36.4%, primarily due to limited representation in the training data.

5.2.2 Error Analysis and Improvements

Analysis of classification errors revealed that misclassifications frequently occur during subtle contextual shifts between categories, particularly at transition points from Understanding to Problem Solving. Mentor statements showed higher classification accuracy than mentee statements, indicating a potential bias in the model's performance. The impact of data imbalance on minority classes emerged as a significant challenge. To address these issues, we propose a comprehensive improvement strategy: implementing data augmentation techniques including back-translation and paraphrasing for minority classes, generating synthetic data for underrepresented categories, and applying class-weighted loss functions during model training. Additionally, we plan to enhance context modeling for transition detection and fine-tune the model specifically for mentee statement classification.

5.2.3 Practical Applications

The framework's effectiveness extends beyond classification to practical applications in mentoring program development. Real-time analysis enables monitoring of mentoring relationship development, identification of successful interaction patterns, and systematic evaluation of mentoring quality. Through data-driven mentor training programs and immediate feedback mechanisms, the framework supports continuous improvement in mentoring practices. Future work will focus on expanding the dataset, particularly for underrepresented categories, refining classification methods, and integrating real-time feedback mechanisms to enhance mentoring quality in engineering education contexts.

5.3 Challenges from Data Imbalance and Dataset Expansion

The evaluation revealed significant challenges posed by data imbalance in both the DSTC7 and UF datasets. Categories such as *Goal Setting and Planning* and *Feedback and Support*, which dominate the dataset, achieved near-perfect performance. In contrast, minority classes like *Exploration and Reflection*, representing less than 1% of the dataset, suffered from low recall and F1-scores. This imbalance biases the model towards majority classes, limiting its ability to generalize effectively for underrepresented categories.

To mitigate these challenges, expanding the UF-specific mentoring dataset is a critical next step. Future work will focus on recording additional mentoring sessions within the UF engineering context, ensuring that all Talk-Move categories are adequately represented. Data augmentation techniques such as back-translation, paraphrasing, and contextual substitutions will be employed to enrich minority categories like *Exploration and Reflection*. Generating synthetic dialogues through NLP-based methods will further enhance the dataset. Collaborations with other institutions for data sharing and collection will also be explored to increase dataset diversity. This comprehensive approach will improve the RoBERTa model's robustness, enabling more accurate classification across all categories and enhancing its applicability to real-world mentoring scenarios in engineering education.

5.4 Limitations and Future Work

Despite promising outcomes, this study encountered limitations that guide our future research direction. First, the relatively small UF mentoring dataset limited model validation and generalizability. Although the DSTC7 dataset provided extensive dialogues, its differing domain introduced potential classification noise. Second, significant data imbalance affected model accuracy, notably impacting the minority category *Exploration and Reflection* (36.4% F1-score), highlighting a clear need for balanced datasets to reliably recognize reflective dialogue.

To address these limitations, future work will prioritize expanding the UF-specific dataset within 12 months through targeted recordings designed to elicit underrepresented categories, particularly emphasizing reflective dialogues. We will apply advanced transformer-based synthetic data generation techniques that maintain linguistic authenticity and context relevance. Additionally, refining RoBERTa's ability to leverage dialogue context, particularly at category transition points, will enhance classification precision. Finally, we aim to implement and evaluate a real-time mentoring feedback system, providing immediate actionable insights to mentors to optimize dialogue quality and effectiveness.

5.5 Implications for Engineering Education

This research holds significant implications for engineering education, particularly in mentor training and mentoring program evaluation. The structured Talk-Move Framework provides an evidence-based method to systematically analyze mentor-mentee interactions, enabling mentors to enhance their mentoring quality through personalized training and objective assessments. Specifically, mentor training programs can leverage dialogue pattern analyses to design targeted interventions, such as modules focused explicitly on developing reflective questioning skills and adaptive problem-solving strategies.

Furthermore, integrating AI-driven dialogue analysis offers scalable, data-driven solutions for assessing mentoring quality at large scales. Real-time feedback mechanisms allow mentors to adjust their mentoring strategies dynamically, addressing diverse mentee needs efficiently and effectively. This scalable implementation model supports consistent quality monitoring across large-scale and remote mentoring initiatives, promoting accessibility and enhancing the inclusivity of mentorship programs.

Lastly, by addressing critical challenges such as student retention, self-efficacy development, and fostering diversity in engineering, the structured application of our framework contributes directly to improving overall student outcomes. The adoption of systematic, evidence-based mentoring supported by robust AI analytics aligns with current educational priorities, providing powerful tools to enhance student engagement and success in engineering education contexts.

5.6 Summary of Results and Future Impact

The RoBERTa-based automated classification model, combined with the structured Talk-Move Framework, has demonstrated significant promise for analyzing mentoring dialogues, achieving strong overall classification accuracy (98.2%). However, notable challenges persist in accurately classifying underrepresented dialogue types, primarily due to data imbalance. These preliminary findings clearly answer our research questions, confirming both the feasibility and practical utility of the developed framework.

Our outlined strategy for dataset expansion and methodological refinements, including synthetic data generation and contextual enhancements, will substantially improve model robustness and classification reliability. As the framework and model continue to evolve, their integration into mentoring practices will likely result in transformative impacts for engineering education, significantly improving mentorship quality, scalability, and student outcomes."

6 Conclusion

The PEER HELPER Talk-Move Framework has been successfully adapted and validated for use in peer mentoring scenarios, demonstrating its capability to systematically analyze and improve mentor-mentee interactions. The integration of RoBERTa-based AI classification has further enhanced the framework's effectiveness, enabling scalable and consistent evaluation of mentoring dialogues. While the current study is limited by the size of the UF mentoring dataset and reliance on publicly available data, future efforts to expand data collection and refine the model will facilitate broader applications and improve peer mentoring outcomes. This research underscores the transformative potential of combining structured frameworks with AI-driven tools to enhance mentor training and support the development of more effective mentoring programs across diverse educational contexts.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 2314772. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] S. Gershenfeld, "A review of undergraduate mentoring programs," *Review of Educational Research*, vol. 84, no. 3, pp. 365–391, 2014.

[2] G. Crisp and I. Cruz, "Mentoring college students: A critical review of the literature between 1990 and 2007," *Research in higher education*, vol. 50, pp. 525–545, 2009.

[3] M. Estrada, P. R. Hernandez, and P. W. Schultz, "A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into stem careers," *CBE—Life Sciences Education*, vol. 17, no. 1, p. ar9, 2018.

[4] C. Pfund, A. Byars-Winston, J. Branchaw, S. Hurtado, and K. Eagan, "Defining attributes and metrics of effective research mentoring relationships," *AIDS and Behavior*, vol. 20, pp. 238–248, 2016.

[5] V. Cornelius, L. Wood, and J. Lai, “Implementation and evaluation of a formal academic-peer-mentoring programme in higher education,” *Active Learning in Higher Education*, vol. 17, no. 3, pp. 193–205, 2016.

[6] D. Yomtov, S. W. Plunkett, R. Efrat, and A. G. Marin, “Can peer mentors improve first-year experiences of university students?” *Journal of College Student Retention: Research, Theory & Practice*, vol. 19, no. 1, pp. 25–44, 2017.

[7] S. Michaels and C. O’Connor, “Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion,” *Socializing intelligence through talk and dialogue*, pp. 347–362, 2015.

[8] K. Miel, R. D. Swanson, M. Portsmore, K. M. Paul, E. A. Moison, J. Kim, and A. V. Maltese, “Characterizing engineering outreach educators’ talk moves: An exploratory framework,” *Journal of Engineering Education*, vol. 112, no. 2, pp. 337–364, 2023.

[9] R. Balyan, T. Arner, K. Taylor, J. Shin, M. Banawan, W. L. Leite, and D. S. McNamara, “Modeling one-on-one online tutoring discourse using an accountable talk framework.” *International Educational Data Mining Society*, 2022.

[10] C. Poor and S. Brown, “Increasing retention of women in engineering at wsu: A model for a women’s mentoring program,” *College Student Journal*, vol. 47, no. 3, pp. 421–428, 2013.

[11] X. Zhai, Y. Yin, J. W. Pellegrino, K. C. Haudek, and L. Shi, “Applying machine learning in science assessment: a systematic review,” *Studies in Science Education*, vol. 56, no. 1, pp. 111–151, 2020.

[12] B. J. Zimmerman, “Becoming a self-regulated learner: An overview,” *Theory into practice*, vol. 41, no. 2, pp. 64–70, 2002.

[13] R. S. Newman, “How self-regulated learners cope with academic difficulty: The role of adaptive help seeking,” *Theory into practice*, vol. 41, no. 2, pp. 132–138, 2002.

[14] K. G. Nelson, D. F. Shell, J. Husman, E. J. Fishman, and L.-K. Soh, “Motivational and self-regulated learning profiles of students taking a foundational engineering course,” *Journal of Engineering Education*, vol. 104, no. 1, pp. 74–100, 2015.

[15] S. A. Karabenick and M. H. Dembo, “Understanding and facilitating self-regulated help seeking,” *New directions for teaching and learning*, vol. 2011, no. 126, pp. 33–43, 2011.

[16] M. K. DiBenedetto and M. C. White, “Applying the model of development of self-regulatory competence to mentoring.” 2013.

[17] R. W. Lent, S. D. Brown, J. Schmidt, B. Brenner, H. Lyons, and D. Treistman, “Relation of contextual supports and barriers to choice behavior in engineering majors: Test of alternative social cognitive models.” *Journal of counseling psychology*, vol. 50, no. 4, p. 458, 2003.

[18] J. W. Colvin and M. Ashman, “Roles, risks, and benefits of peer mentoring relationships in higher education,” *Mentoring & tutoring: partnership in learning*, vol. 18, no. 2, pp. 121–134, 2010.

[19] R. W. Young and C. M. Cates, “Listening, play, and social attraction in the mentoring of new teachers,” *Mentoring & Tutoring: Partnership in Learning*, vol. 18, no. 3, pp. 215–231, 2010.

[20] B. Leidenfrost, B. Strassnig, A. Schabmann, C. Spiel, and C.-C. Carbon, "Peer mentoring styles and their contribution to academic success among mentees: A person-oriented study in higher education," *Mentoring & Tutoring: Partnership in Learning*, vol. 19, no. 3, pp. 347–364, 2011.

[21] B. Bunting and D. Williams, "Stories of transformation: Using personal narrative to explore transformative experience among undergraduate peer mentors," *Mentoring & Tutoring: Partnership in Learning*, vol. 25, no. 2, pp. 166–184, 2017.

[22] L. F. D'Haro, K. Yoshino, C. Hori, T. K. Marks, L. Polymenakos, J. K. Kummerfeld, M. Galley, and X. Gao, "Overview of the seventh dialog system technology challenge: Dstc7," *Computer Speech & Language*, vol. 62, p. 101068, 2020.

[23] Y. Liu, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, vol. 364, 2019.

[24] D. Wang, L. Gao, D. Shan, G. Chen, C. Zhang, and B. Kao, "When tutors simultaneously instruct students from the primary, middle, and high school levels in online one-on-one tutoring: Investigating the interaction dynamics using ai, ena, and lsa methods," *Journal of Science Education and Technology*, pp. 1–15, 2024.

[25] V. Ashish, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, p. I, 2017.

[26] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.

[27] N. Zepke and L. Leach, "Improving student engagement: Ten proposals for action," *Active learning in higher education*, vol. 11, no. 3, pp. 167–177, 2010.

[28] R. A. Glazier, "Building rapport to improve retention and success in online classes," *Journal of Political Science Education*, vol. 12, no. 4, pp. 437–456, 2016.

[29] L. O. Wilson, "Anderson and krathwohl–bloom's taxonomy revised," *Understanding the new version of Bloom's taxonomy*, 2016.

[30] R. Pekrun, T. Goetz, W. Titz, and R. P. Perry, "Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research," *Educational psychologist*, vol. 37, no. 2, pp. 91–105, 2002.