

Towards Actionable Pedagogical Feedback: A Multi-Perspective Analysis of Mathematics Teaching and Tutoring Dialogue

Jannatun Naim
University of Colorado Boulder
Boulder, CO, United States
jannatun.naim@colorado.edu

Jie Cao
University of Oklahoma
Norman, OK, United States
jie.cao@ou.edu

Fareen Tasneem
University of Chittagong
Chittagong, Bangladesh
fareen.tasneem@gmail.com

Jennifer Jacobs
University of Colorado Boulder
Boulder, CO, United States
jennifer.jacobs@colorado.edu

Brent Milne
Saga Education, United States
bmilne@sagaeducation.org

James Martin
University of Colorado Boulder
Boulder, CO, United States
james.martin@colorado.edu

Tamara Sumner
University of Colorado Boulder
Boulder, CO, United States
sumner@colorado.edu

ABSTRACT

Effective feedback is essential for refining instructional practices in mathematics education, and researchers often turn to advanced natural language processing (NLP) models to analyze classroom dialogues from multiple perspectives. However, utterance-level discourse analysis encounters two primary challenges: (1) multi-functionality, where a single utterance may serve multiple purposes that a single tag cannot capture, and (2) the exclusion of many utterances from domain-specific discourse move classifications, leading to their omission in feedback. To address these challenges, we proposed a multi-perspective discourse analysis that integrates domain-specific talk moves with dialogue act (using the flattened multi-functional SWBD-MASL schema with 43 tags) and discourse relation (applying Segmented Discourse Representation Theory with 16 relations). Our top-down analysis framework enables a comprehensive understanding of utterances that contain talk moves, as well as utterances that do not contain talk moves. This is applied to two mathematics education datasets: TalkMoves (teaching) and SAGA22 (tutoring). Through distributional unigram analysis, sequential talk move analysis, and multi-view deep dive, we discovered meaningful discourse patterns, and revealed the vital role of utterances without talk moves, demonstrating that these utterances, far from being mere fillers, serve crucial functions in guiding, acknowledging, and structuring classroom discourse. These insights underscore the importance of incorporating discourse relations and dialogue acts into AI-assisted education systems to enhance feedback and create more responsive learning environments. Our framework may prove helpful for providing human educator feedback, but also aiding in the development of AI agents that can effectively

emulate the roles of both educators and students.

Keywords

Educational Data Mining, Talk Moves, Dialogue Acts, Discourse Relations, Classroom Analysis

1. INTRODUCTION

Research increasingly supports dialogic teaching (learning) - an approach that encourages student-driven academic discourse — as a means of enhancing motivation and learning outcomes [10, 36, 60, 71]. A critical element of dialogic instruction is accountable talk, which structures classroom discourse around three key dimensions: learning community, content knowledge, and rigorous thinking [53, 55]. Analyzing discourse in mathematic teaching and tutoring is essential for understanding its effectiveness.

Researchers and educators can assess dialogic teaching by examining who speaks, the nature of their contributions, and the extent to which discussions reflect multiple perspectives [45]. Classroom observations, whether conducted by educators or researchers, can document changes in discourse patterns over time, providing valuable insights for professional learning [15]. Historically, discourse analysis relied on labor-intensive qualitative methods, including detailed human annotation of classroom interactions [51]. Advances in recording technologies and natural language processing (NLP) have introduced scalable alternatives, enabling interested parties to capture and analyze classroom dialogue more efficiently [30, 58].

Recent advancements in educational technology have demonstrated the significant potential of automated, data-driven feedback derived from classroom recordings in fostering teacher development. The emergence of AI-driven tools has the potential to transform classroom discourse analysis, providing teachers with real-time, automated feedback on their instructional strategies. Improvements in machine learning, NLP, and automatic speech recognition have enabled researchers to develop models capable of detecting key discourse features, such as accountable talk [11, 18, 63], teacher questioning techniques [24, 28, 56], student participation [17, 29,

70], teacher uptakes [26, 27]. These findings suggest that AI-driven feedback mechanisms can serve as powerful tools for enhancing interactive teaching practices across various educational settings. A key challenge lies in delivering more actionable, context-specific feedback to educators to enhance their instructional effectiveness. One potential solution is to provide a detailed analysis from multiple perspectives and varying levels of granularity, ensuring insights are both comprehensive and practical.

1.1 Background and Related Work

Previous research on providing feedback to educators in multiple granularities or from a variety of perspectives often use different refined NLP models to solve multiple discourse analysis tasks on the same dialogue. This feedback ranges from information about turn-level discourse moves to session-level quality scores. For example, M-Powering Teachers [27, 26] has provided instructors with feedback in the context of online programming courses and 1:1 online tutoring via three metrics: uptakes of student contributions, talk time, and questioning practices. In addition, using a dataset collected by the National Center for Teacher Effectiveness (NCTE) [25] developed classifiers to annotate five turn-level indicators of effective mathematics instruction: student on task, teacher on task, student reasoning, high uptake, and focusing questions. Another recent study [69] developed models for a multi-dimensional assessment on the quality of classroom discussion based on two measures: a sentence-level Analyzing Teaching Moves discourse measure (ATM) [22, 23], and a session-level Instructional Quality Assessment (IQA) [50]. IQA scores use a 1-4 scale on 11 dimensions for modeling teacher and student contributions, and 4 of 11 dimensions' scoring depends on the ATM moves. Other multi-perspective pedagogical feedback systems like the above examples also exist. However, turn-level discourse analysis often faces two challenges: (1) *Multi-Functionalities*, where a sentence may serve multiple functions which may not be captured by a single tag [3, 20, 33]. Designing multiple phenomena as binary indicators could solve the multi-functionalities issue by allowing a single turn to activate multiple binary indicators. (2) *Many non-move utterances often get ignored*. Domain specific speech acts such as accountable talks [53] and ATMs are designed to identify **high leverage** discourse moves grounded in the corresponding theories, which leads to many utterances not being covered during discourse modeling (denoted as non-talk moves). For example, in our prior work developing classifiers for accountable talk moves, more than 50% of teachers' and students' utterances were classified as non-talk moves [18], and these non-talk moves have not been well-studied [38, 63].

1.2 Current Study and Contribution

Our work also falls in the multi-perspective feedback research, but specially focus on addressing the issue of multi-functionalities and understanding the nature of non targeted moves in utterance-level code classification (such as talk moves, ontask, and ATM, etc). Besides domain-specific talk moves (based on Accountable Talk Theory [53, 55] over 12 moves), our multi-perspective feedback use a version of broad-coverage dialogue acts designed for multi-functionalities, which is based on flattened multi-functional SWBD-MASL schema [41] with over 43 tags. We also propose to use graph-based discourse relation parsing (based on Segmented Discourse Representation Theory (SDRT)[5] over 16 relations) to naturally model the dependency relations among arbitrary utterance pairs. We separate out the global session-level analysis (such as IQA or MQI) to solely focus on fine-grained discourse analysis. Our contributions can be summarized as follows:

- We utilize two datasets TalkMoves (teaching) and SAGA22 (tutoring) annotated with 3-view discourse analysis: (1) domain specific talk moves, (2) dialogue acts, and (3) discourse relations to investigate pedagogical behaviors in mathematics education. We are interested in both classroom teachers and tutors because both work directly with students to support their mathematics learning. By using our proposed toolkits to automatically annotate the above three discourse views with the state-of-the-art models, we could easily conduct a comparative study on the pedagogical dynamics between teaching and tutoring domain.
- We propose a top-down analysis framework to provide a thorough understanding of utterances that contain talk moves, as well as utterances that do not contain talk moves. Our methodology involved distributional unigram analysis, documenting frequent pedagogical behaviors via sequential talk move analysis, and highlighting the intersection between talk moves/non-talk moves, dialogue acts and discourse relations. Dialogue acts offer insights into the role of non-talk move utterances, while discourse relations highlight the interplay between talk and non-talk moves within the discourse flow. This integrated analysis of talk moves, dialogue acts, and discourse relations helps address the research gap on the contributions of non-talk moves to the overall discourse structure. This multi-perspective framework offers more comprehensive explanations for providing feedback to human educators and for guiding the design of AI agents to mimic the role of educators and students.

2. METHODS

2.1 Dataset

We utilized two primary datasets from mathematics K-12 educational contexts, one teaching dataset TalkMoves and one tutoring dataset SAGA22, summarized in Table 1. Each session's transcript from the two datasets has been meticulously annotated by human experts, labeling instances of 7 distinct teacher talk moves and 5 student talk moves. The TalkMoves dataset, originally introduced by [64], comprises 567 mathematics classroom sessions spanning a diverse range of topics across elementary to high school levels. The SAGA22 dataset, originally introduced by [18], is derived from a high school tutoring dataset collected in 2022 in collaboration with Saga Education, a non-profit provider of tutoring services. Saga partners with school districts serving low-income and historically marginalized communities to offer high dosage mathematics tutoring. Different from classroom teaching, Saga's tutoring model operates in a hybrid format, where students participate in tutoring sessions within a classroom while paraprofessional tutors engage remotely using technology. The annotated Saga dataset includes 121 sessions, totaling 69.7 hours of video, with 33,695 teacher utterances and 11,115 student utterances labeled with talk moves.

2.2 Multi-Perspective Dialogue Analysis

Our approach integrates three perspectives of analysis with talk moves, dialogue acts, and discourse relations, as shown in Figure 1:

Talk Moves. Accountable Talk theory [52] encompasses various types of talk moves, each differing in usage frequency and application across classroom contexts. The TalkMoves and Saga datasets are annotated for seven teacher Talk Moves (with prefix "T" and in **dark cyan**) designed to foster student engagement, guide discussions,

Table 1: Summary of Datasets on Mathematic Teaching (TalkMoves) and Tutoring(SAGA22)

Dataset	Sessions	T-Utterances	S-Utterances	Domain	Students per Session	Session Length
TalkMoves	567	174,168	59,823	Mixed-Teaching	20	30-55 min
SAGA22	121	33,695	11,115	High School-Tutoring	2-5	35 min

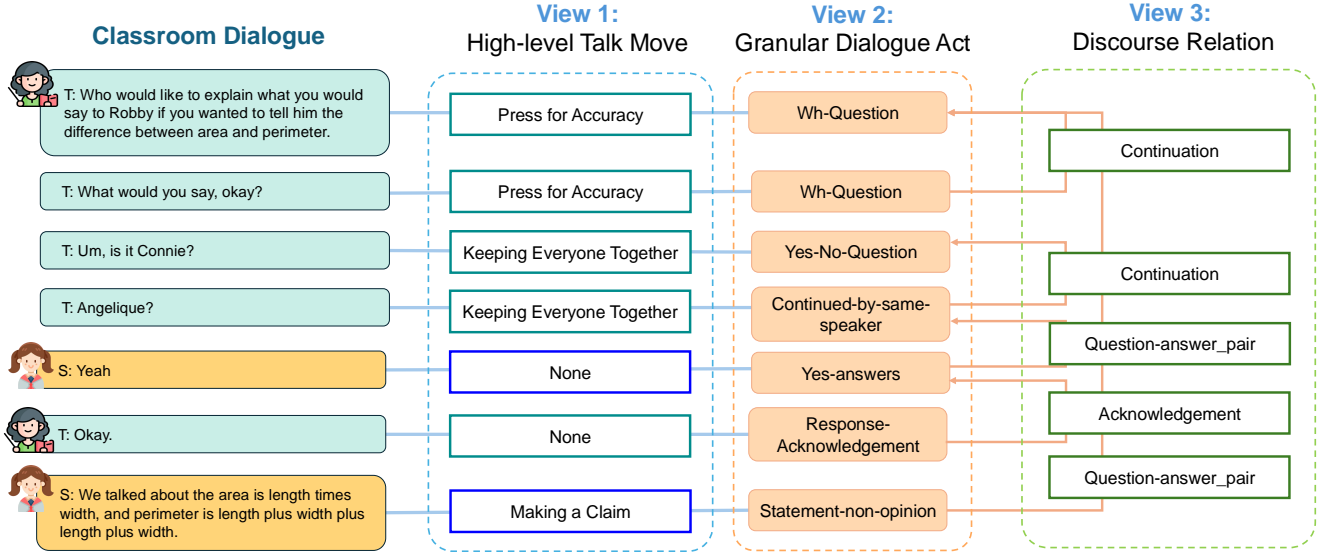


Figure 1: A running example for our multi-perspective analysis with talk moves, dialogue acts, and discourse relations.

and facilitate productive learning interactions: keeping everyone together (**T-KPTG**), encouraging participation (**T-GSTUR**), restating responses for clarity (**T-RESTAT**), revoicing ideas to highlight significance (**T-REVOIC**), pressing for reasoning (**T-PRSR**), pushing for accuracy (**T-PRSA**), and an unspecified category for utterances that do not fit these moves (**T-NONE**). Similarly, the datasets are annotated for five student Talk Moves (with prefix "S" and in blue) that capture how students actively contribute to discussions: making claims (**S-MCLAIM**), providing evidence and reasoning (**S-PROEVI**), reacting to and building on peers' ideas (**S-RELTO**), requesting clarification (**S-ASKMI**), and an unspecified category for utterances outside these moves (**S-NONE**). This set of talk moves closely corresponds to accountable talk theory but is not exhaustive; there are other important talk moves that were not included due to their low frequency in the dataset and/or minimal extant literature [38]. This study begins with an analysis of both teacher/tutor and student talk moves. Grounded in Accountable Talk theory, we aim to deepen our understanding of classroom discourse dynamics, the effectiveness of instructional strategies, and the role of structured dialogue in enhancing student learning and engagement. View 1 in Figure 1 shows the corresponding talk moves for each utterance in the example classroom dialogue. In our analysis, we used the human annotated talk moves. For future analysis, existing toolkits for modeling talk moves that include pretrained language models could be used [63, 65, 54, 18].

Dialogue Act. Talk moves can be viewed as a set of domain specific dialogue acts (DAs) tailored for the educational domain. The foundational idea that each utterance performs an action was first introduced by philosopher Wittgenstein [72]. Speech act theory, developed by Austin [6] and later expanded by his student Searle [61], was one of the earliest frameworks for categorizing communicative

actions. DAMSL (Dialogue Act Markup in Several Layers)[2, 21] was proposed to address the issue of multi-functionalities [3, 20, 33] by allowing utterances to serve multiple roles across independent layers, such as *Communicative-Status*, *Information-Level*, *Forward-Looking Function*, and *Backward-Looking Function*. Each layer contains multiple subcategories, resulting in a rich but complex tagging scheme. A major challenge with DAMSL is the vast number of possible tag combinations, which complicates both manual and automatic annotation. To mitigate this, SWBD-DAMSL [41], an adaptation of DAMSL for the Switchboard Corpus [32], was introduced. It reduces the tag set to 220 combinations (including 28 new tags not in DAMSL) and clusters them into 42 mutually exclusive categories, thereby flattening the multi-functional labels. This reduction significantly improves the feasibility of automatic dialogue act annotation and the development of dialogue-act-specific language models for speech recognition [62], while sacrificing the flexibility and expressiveness in capturing multi-functionality. Comparing to the talk moves in View 1, View 2 in Figure 1 demonstrates the more finegrained DAs for each utterance. In our work, we follow SWBD-DAMSL's 42 tags¹ with an extra tag "+", which means continued previous talk by the same speaker. Among those top-tier models, we mainly considered the open-sourced models as possible toolkits. The final model we selected is a hierarchical neural architecture with a Bi-GRU on top of trainable speaker-aware utterance encodings from RoBERTa-base [48], which jointly tags 196 utterances in a chunk window into their corresponding DAs [34]. We pretrained a model on the SWBD corpus with 43 tags (such as *Wh-Question*, *Statement-non-opinion*, *Yes-No-Questions* etc, denoted in italic and the color orange), obtaining 82.4 accuracy, which is approaching the state-of-the-art models with 83.1 on 42 tags. The model results

¹<https://web.stanford.edu/~jurafsky/ws97/manual.august1.html>

are replicable with its open-sourced code on the github ².

Discourse Relation. Beyond the utterance-level talk moves and dialogue acts, discourse relations as a structural dependency is able to capture the multi-functional nature of an utterance in relation to its neighboring utterances. It has been extensively studied across various discourse theories [7, 31, 14], such as Rhetorical Structure Theory (RST)[49], Discourse Representation Theory (DRT)[42], Hobbs’ theory of discourse [35], and the Penn Discourse Treebank (PDTB) framework [57]. Among those, Segmented Discourse Representation Theory (SDRT)[5], an extension of DRT, offers a hierarchical model of text organization with full discourse annotation. Several corpora, including DISCOR[59], ANNODIS [1] and STAC [4], implement SDRT using directed acyclic graphs (DAGs) that allow multiple parent nodes but prohibit crossing edges. The success of SDRT has made it as a popular framework to study discourse relations in various multiparty dialogues, such as Molweni for Ubuntu online forum [47] and Minecraft Structured Dialogue Corpus (MSDC) for jointly modeling conversation moves and builder moves [68]. In this work, we mainly focused on the SDRT analysis for mathematics dialogues. As shown in View 3 in Figure 1, beyond the utterance-level (Views 1 and 2), the SDRT provides the labeled discourse relations between pairs of utterances if one exists. Richer expressiveness often indicates harder annotation for both human and models. Among the top-tier SDRT parsing models [8, 19, 46, 67], we select the state-of-art model Llamipa [67] on automatic discourse parsing. Llamipa is a LLM (Llama3-8B) finetuned on Minecraft Structured Dialogue Corpus (MSDC), which has shown good generalization on MSDC (79.51 F1) and STAC (77.96 F1) datasets. The model we used is the public checkpoint on the huggingface model hub ³. As discussed in §4.2, finetuning the model with future in-domain annotation may further improve the performance.

2.3 Top-Down Analysis Framework

We propose a 3-stage top-down framework to provide a thorough analysis of the mathematics teaching and tutoring datasets. We cluster **T-NONE** and **S-NONE** as non-talk moves utterances throughout our analysis. For each of the following stages, we always started with an analysis with all talk moves, then we specifically studied the non-talk moves utterances.

2.3.1 High-level Analysis via Unigram

Since our datasets include manually annotated talk moves, beginning with a faithful unigram talk move analysis ensures a reliable high-level examination of discourse patterns. To achieve a more granular understanding, we further analyzed the DAs in conjunction with the talk moves. Specifically, we examined the top three DAs corresponding to each talk move and the top seven DAs for each non-talk move utterance. These thresholds were selected to ensure that at least 50% and 75% of the overall distribution are captured, respectively. Utterances labeled with the dialogue act *Continued-by-Same-Speaker* were excluded from this analysis, as they primarily function as extensions of preceding DAs and do not contribute meaningful standalone insights. All of our analyses compare the teaching and tutoring domains to identify similarities, differences, and domain-specific discourse patterns.

2.3.2 Behaviors Discovered via Sequential Analysis

²<https://github.com/zihaohe123/speak-turn-emb-dialogue-act-clf>

³<https://huggingface.co/linagora/Llamipa>

Unigram talk move analysis provides only a limited distributional perspective, making it insufficient for capturing detailed pedagogical behaviors. To address this, we conducted a second-level analysis focusing on the sequential patterns of talk moves, including non-talk moves, to identify highly frequent interaction patterns. We only considered sequences or "transitions" with a probability of 10% or higher between two talk moves. To better understand how different participants engaged and verbally responded, we categorized our transition analysis based on the actor (teacher or student) receiving the transition, allowing us to distinguish behavioral tendencies across different roles. We also analyzed sequences of talk moves excluding intervening non-talk moves, where we included all transitions without a probability threshold, ensuring a more comprehensive and accurate understanding of these meaningful interactions. For sequences that involved talk moves and non-talk move utterances, we analyzed the probability distribution of **T-NONE** occurring between two talk moves to capture how these **T-NONE** engage with a preceding talk move and influence the transition to the succeeding one. To determine the probability, we first identified instances where the talk move pairs are separated by zero or more **T-NONE** talk moves. For each possible number of **T-NONE** talk moves separating the pair, we calculated the frequency of such instances, excluding those with a frequency below 5%. We then utilized the following equation to calculate the probability:

$$\text{Probability}_{T\text{-}None}(tm_j, tm_k) = \begin{cases} \frac{\sum (TNoneCount_i \times Count_i)}{\sum Count_i} \times 100, & \text{if } \sum Count_i > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

In the above equation, $TNoneCount_i$ denotes the count of **T-NONE** utterances separating two talk moves, while $Count_i$ denotes the count of instances of two talk moves tm_j and tm_k separated by $TNoneCount_i$ **T-NONE**.

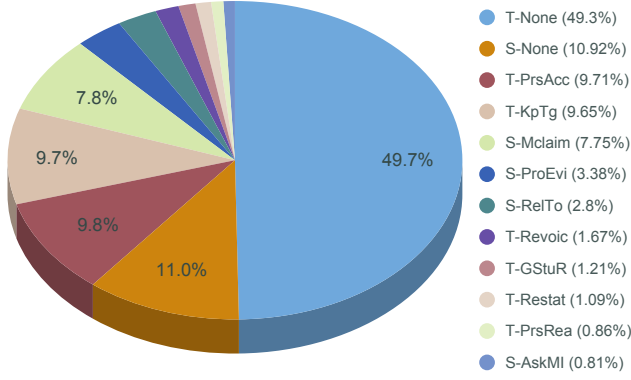
2.3.3 Deep-dive via Multi-view Analysis

Our analysis of sequential bigram pairs of talk moves provides insight into high-probability transitions that reveal key behavioral patterns of both teachers and students within classroom discourse. However, certain sequential dependencies between talk moves may remain implicit due to the presence of intervening utterances that do not contain talk moves. These non-talk moves elements can obscure direct talk move connections while still playing a crucial role in shaping the discourse dynamics. For example, **T-NONE** utterances might be used by the teacher to give directions, showing students how to solve a problem, or evaluating a student’s idea. To gain a deeper understanding of how utterances with and without talk moves are interwoven within these sequences and the pedagogical behavior patterns they reflect, we extend our analysis to examine the discourse relations underlying these transitions. Our investigation consists of two key components. First, we analyze direct transitions between talk move pairs to explore how they interconnect and contribute to the structured progression of discourse. Second, we examine transitions where non-talk moves utterances intervene, assessing their role in influencing dialogue flow and their potential role in shaping interactions.

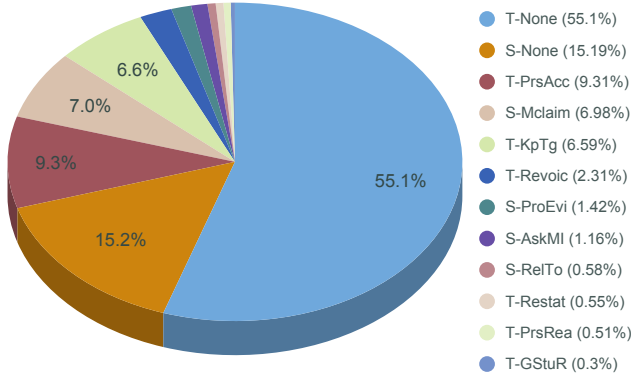
3. RESULTS

3.1 Results from Unigram Analysis

By examining the unigram distribution of teacher and student talk moves in both the teaching and tutoring datasets, Figure 2 shows



(a) Teacher/Student Talk Moves in the TalkMove Dataset.



(b) Tutor/Student Talk Moves in the SAGA Dataset.

Figure 2: Comparison of Teacher/Tutor and Student Talk Moves across TalkMove and SAGA Dataset.

that the tutoring dataset contains a higher proportion of utterances without talk moves (**T-NONE** and **S-NONE**) for both instructors (teachers/tutors) and students, with 5.8% more for instructors and 4.3% more for students compared to the teaching dataset. Consequently, the talk moves have a lower proportion in the tutoring dataset, except for **T-REVOIC** (2.31% > 1.67%) and **S-ASKMI** (1.61% > 0.81%). To gain further insights into the pedagogical behavior patterns and students' responses, we analyzed both the talk moves and the non-talk moves utterances with DAs to capture a more granular view. Among 43 DAs, only 30 are used in our two datasets as presented in Figure 3.

3.1.1 Talk Moves with DAs

Our findings on talk moves with their top three associated DAs are illustrated in Figure 4. As expected, **Wh-Question** naturally dominates the teacher talk moves **T-PRsACC**, **T-PRsREA**, and the student talk move **S-ASKMI**, which are related to asking questions. **Statement-non-opinion** is the predominant DA across the student talk moves: **S-MCLAIM**, **S-PROEVI**, **S-RELTO**, and the teacher talk moves: **T-RESTAT** and **T-REVOIC**. **Yes-No-Questions** is frequently used in **T-GSTUR**, **T-KPTG**, **T-PRsACC** and **S-ASKMI**. **Action-directive** is used more frequently in teacher talk moves than in student talk moves, highlighting their pivotal role in guiding and structuring classroom interactions. Besides that, students actively contribute to directing actions within classroom discussions.

More interestingly, as illustrated in Figure 5, **S-ASKMI** and **T-PRsACC** typically expressed as questions, can also appear in the

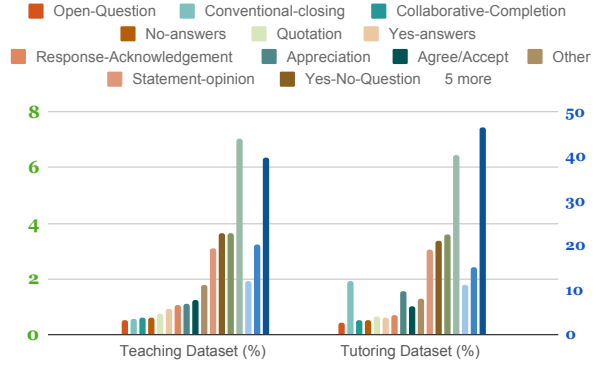
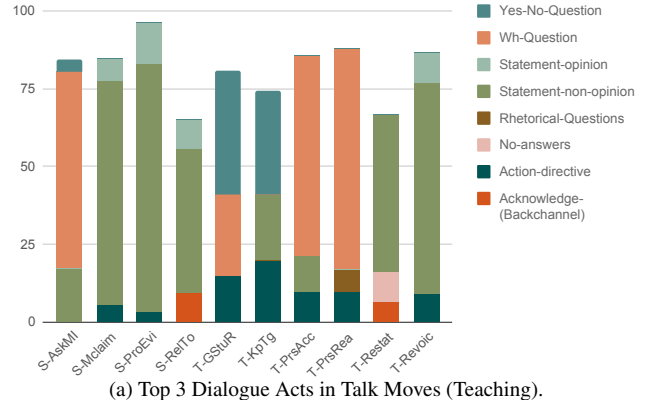
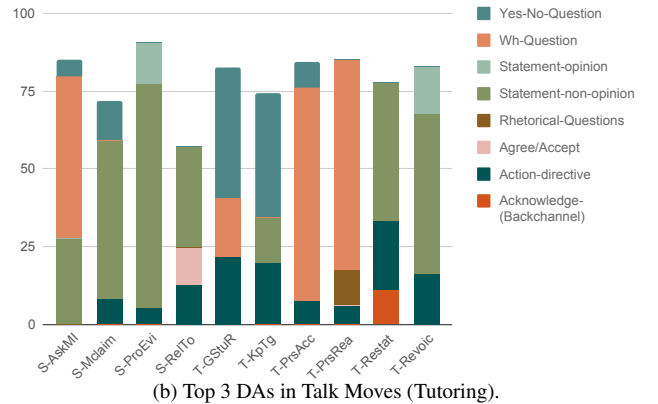


Figure 3: Distribution of DAs in the Teaching and Tutoring Datasets. The last three DAs: **Action-directive**, **Statement-non-opinion**, and **Continued-by-Same-Speaker** are represented on the right scale, while all others are on the left scale. Only DAs with a ratio greater than 0.5% are displayed.



(a) Top 3 Dialogue Acts in Talk Moves (Teaching).



(b) Top 3 DAs in Talk Moves (Tutoring).

Figure 4: Comparison of Talk Moves with DAs.

form of statements. For example, **S-ASKMI** being expressed as a **Statement** may reflect a student behavior of implicitly conveying their need for additional information by articulating their confusion rather than making a direct request. Recognizing this phenomenon can make future agent design for mimicing and comprehending teacher or students behavior more diverse. Furthermore, **S-MCLAIM** also exhibits a notable percentage of **Yes-No-Questions** rather than statements. Figure 6 illustrates instances of these cases. This figure reveals that students may frame claims as **Yes-No-Questions**. This pattern suggests that students might have lower confidence when making claims. The higher occur-



Figure 5: Interesting DA Use Cases for Talk moves

rence of this phenomenon in the tutoring dataset, as observed from Figure 4, could indicate an overall lower confidence level among students in the tutoring domain. Furthermore, the association of *Yes-No-Questions* with *S-MCLAIM* may serve as a valuable signal for teachers or dialogue agents to intervene and support students in developing greater confidence in their responses.

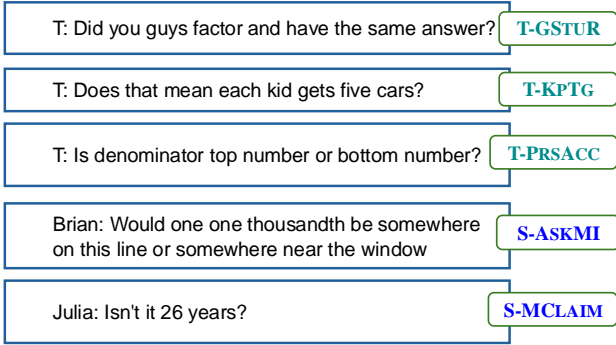


Figure 6: Examples of teacher talk moves *T-GSTUR*, *T-KPTG*, *T-PRSACC*, and the student talk moves *S-ASKMI*, *S-MCLAIM*, that aligns with the dialogue act *Yes-No-Questions*.

Our results suggest that DAs offer a more nuanced understanding of the behavioral patterns of students and teachers embedded within high-level talk moves. When integrated into instructor feedback or the training of dialogue agents, this detailed perspective may be helpful in more accurately interpreting student needs, suggesting specific responses and enhancing engagement.

3.1.2 Non-talk moves Interactions

We also examined the utterances without talk moves (*T-NONE* and *S-NONE*) and their associated DAs, as illustrated in Figure 7. Notably, both *S-NONE* and *T-NONE* exhibit a similar distribution of DAs. The *Statement-non-opinion*, *Acknowledgment-(Backchannel)* and *Action-directive* are the most frequent DAs in all of these utterances. Moreover, *Acknowledgment-(Backchannel)* dominates more for *S-NONE* than *T-NONE*. In the tutoring domain, *None* talk moves also feature a reasonable portion of *Conventional-Closing*, which indicates that teachers and students are more engagingly saying "Bye" and "See you" in the online tutoring sessions.

In summary, as shown in Figure 2, non-talk moves utterances make up over 50% of the dialogue exchanges in both the teaching and tutoring datasets. Analyzing these utterances complements the study of talk moves by looking at the role of a broader range of DAs that can be used to define the conversational dynamics.

3.2 Results on Sequential Talk Move Analysis

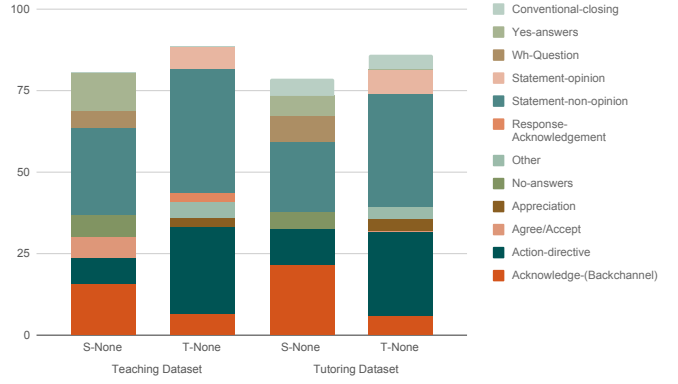


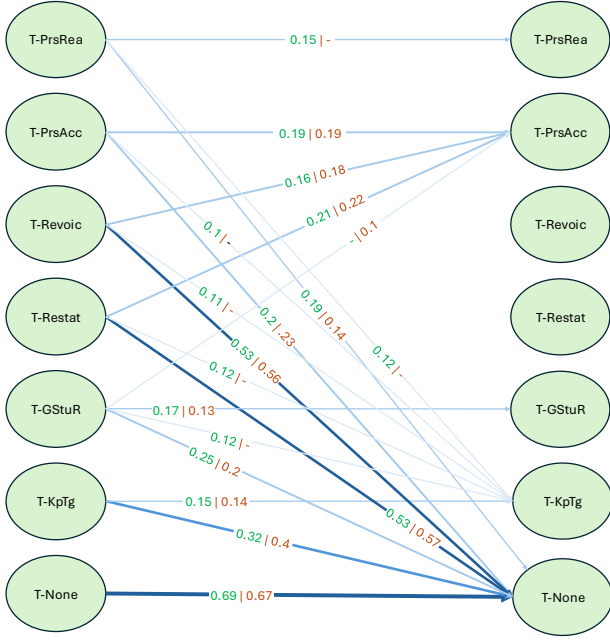
Figure 7: Comparison of *None* talk moves (*T-NONE* and *S-NONE*) with their associated DAs.

3.2.1 Transition Diagrams of Talk Moves

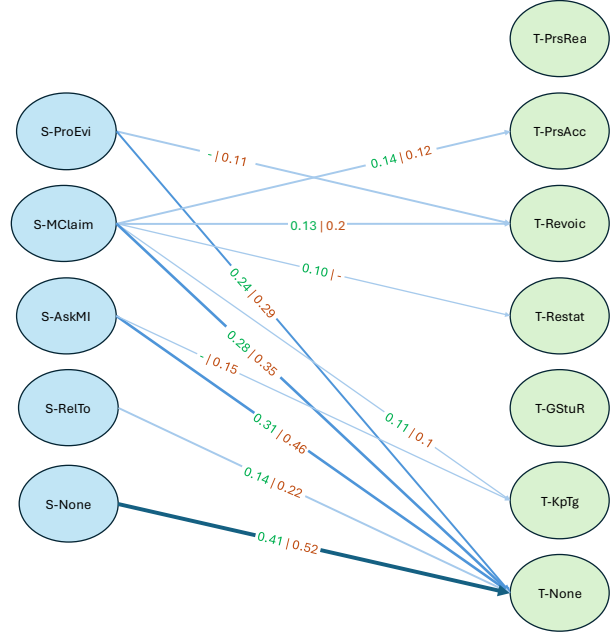
Next we expanded our unigram analysis to bigram analysis to capture frequently occurring pedagogical behaviors in both the teacher and tutoring domains. To examine the interaction pattern between talk moves (and utterances without talk moves), we determined the transition probability between any two utterances. Figure 8 focuses on the teacher or tutor's response after a given talk move, while Figure 9 focuses on the students' behavior after a talk move - either their own or made by other students or the teacher/tutor. The two transition probabilities on each edge in these figures represent the likelihood of one talk move being followed by another in teaching and tutoring respectively. Only transitions with a probability greater than 10% are displayed. This 10% threshold allows us to filter and focus on highly frequent instructional behaviors.

Transitions to Teacher/Tutor. Figure 8a shows that talk move pairs with the higher transition probabilities are consistent across both the teaching and tutoring datasets in the case of teacher-teacher and tutor-tutor talk move pairs. However, in the teaching dataset, all talk moves except for the *T-NONE* had a transition probability of 10% or higher to the *T-KPTG*, whereas this pattern does not hold in the tutoring dataset. This discrepancy aligns with the notable (3%) difference in the occurrence of the *T-KPTG* talk move between the two datasets, indicating that well-trained teachers may provide more group engagement in the classroom. As shown in Figure 8b, all student talk moves are often followed by *T-NONE*, which further demonstrate that important information resides in non-talk move utterances, especially for those replying to *S-ASKMI*, *S-MCLAIM*, and *S-PROEVI*. Additionally, the student talk move *S-MCLAIM* are often followed by *T-PRSACC*, *T-REVOIC* and *T-KPTG* in both domains, which highlights the potential need to suggest an instructional strategies based on the detailed content of students' claim. Furthermore, the observed transitions suggest a missed opportunity for educators to deepen student reflection and understanding by following a student's *S-MCLAIM* talk move with *T-PRSREA*.

Transitions to Students. As shown in Figure 9a, the conceptually aligned talk move pairs to students: *T-PRSREA* → *S-PROEVI*, *T-PRSACC* → *S-MCLAIM*, and *T-GSTUR* → *S-RELT* generally have high and similar transition probabilities across both the datasets. However, the transition probability for *T-PRSREA* → *S-PROEVI* pair is significantly higher in the tutoring dataset (41%), exceeding that of the teaching dataset by 13%. The pair *T-KPTG* → *S-NONE* also have moderately high transition probabilities in



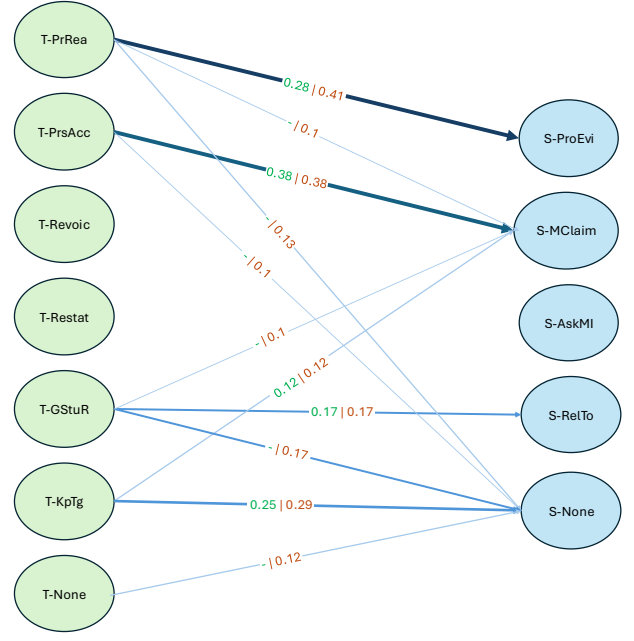
(a) Intra Teacher Talk Move Transition



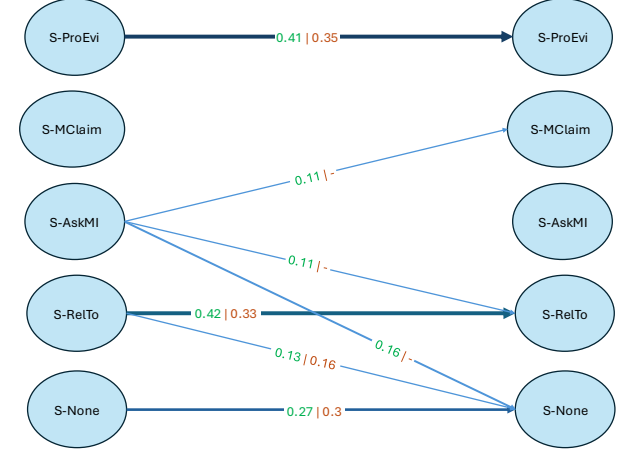
(b) Student-Teacher Talk Move Transition

Figure 8: Transition probabilities of one talk move being followed by another teacher talk moves in both the Teaching (numbers in green) and Tutoring (numbers in orange) dataset. The thickness of the edges represents the likelihood of transitions, with only probabilities above 10% shown for clarity.

both datasets. Moreover, tutor talk moves show a slight tendency to be followed by student utterances without talk moves (**S-NONE**), consistent with the higher percentage of **S-NONE**-tagged student utterances in the tutoring dataset. As illustrated through Figure 9b, student-student talk move pairs also have almost similar transition probabilities across both the datasets, though the teaching domain exhibits a greater variety of transition pairs.



(a) Teacher-Student Talk Move Transition

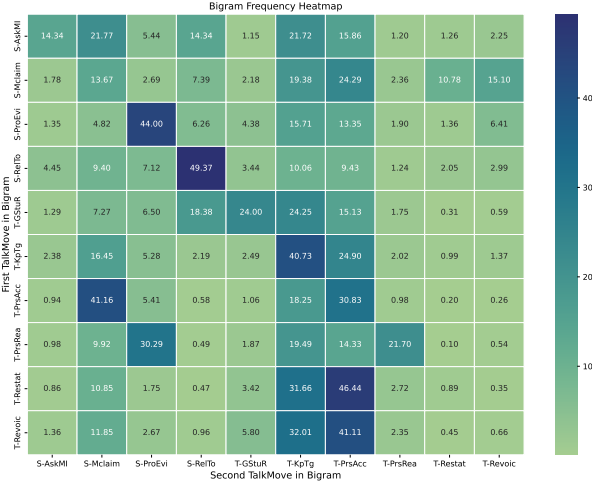


(b) Intra Student Talk Move Transition

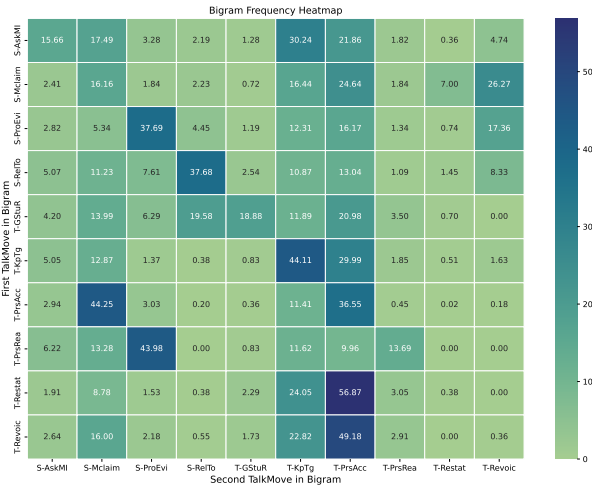
Figure 9: Transition probabilities of one talk move being followed by another student talk move in both the Teaching (numbers in green) and Tutoring (numbers in orange) dataset. The thickness of the edges represents the likelihood of transitions, with only probabilities above 10% shown for clarity.

3.2.2 Non-talk moves Utterances Interaction

From the unigram distribution in Figure 2 and the transition diagrams in Figure 8 and Figure 9, we can observe that interactions that don't involve talk moves (**T-NONE**) are the most prevalent in classroom discourse across teaching and tutoring domains. These interactions also have higher incoming transition rates from the talk moves. According to the procedure described in Section 2.3.2, Figure 11b shows that the tutoring domain contains more **T-NONE** interactions between talk move pairs compared to the teaching domain. In both datasets, self-transitions (indicated by the lighter shades along the diagonal in the heatmaps) have fewer **T-NONE** interactions between them than other transitions (see results on teaching domain in Figure 11a). Additionally, **T-RESTAT** and **T-REVOIC** exhibit a higher



(a) Teaching



(b) Tutoring

Figure 10: Transition between Talk Moves excluding Intermediary *None* Talk Moves

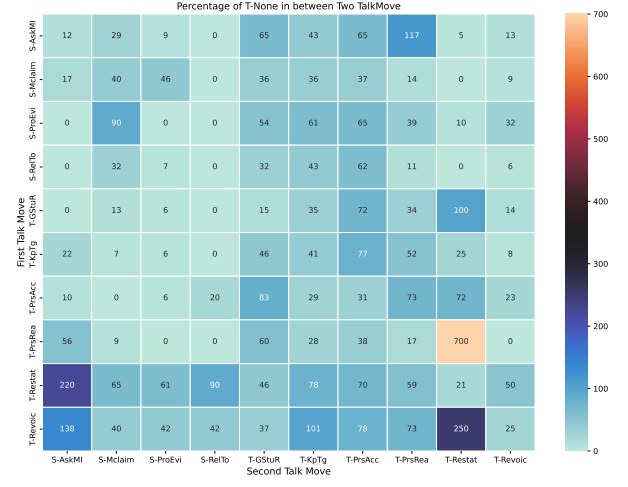
occurrence of **T-NONE** following them before transitioning to another talk move. Figure 10b shows a bigram analysis exclusively on talk moves by filtering out non-talk moves from the sequence of utterances in the tutoring settings. **T-KpTg** and **T-PrsAcc** have relatively high incoming transition (dark columns) indicates they are two frequent structural strategies without considering the None talk moves. Similar patterns are also shown in the teaching sessions (Figure 10a)

In summary, the above transition diagram highlights high-frequency behavior patterns in talk move pairs. The probability distribution of **T-NONE** separating talk-move pairs helps estimate the influence of other behaviors between these interactions.

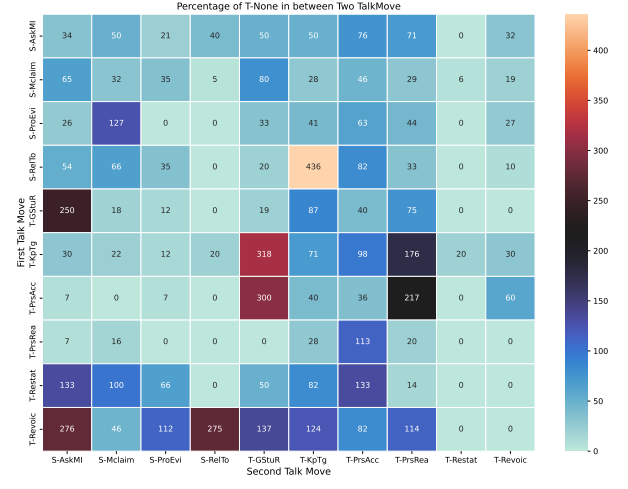
3.3 Results from Multi-View Deep-Dive

To gain deeper insights into the interactions uncovered in our previous sequential analysis §3.2, we turn to a multi-view analysis via discourse relations considering the transitions among utterances with and without talk moves.

3.3.1 Discourse Relations involving Talk Moves



(a) Teaching



(b) Tutoring

Figure 11: Comparison of Probability of **T-NONE** occurrence in between Talk Moves.

Table 2: Comparison of Bigrams in Teaching and Tutoring Datasets

Bigrams	Teaching Dataset		Tutoring Dataset	
	Trans Prob	Discourse Relations	Trans Prob	Discourse Relations
S-MCLAIM - T-PrsACC	14%	ClariQ.(21.1%) Cont. (19.69%) QElab.(10.49%)	12%	ClariQ.(33%) Cont. (8.37%) QElab.(6.65%)
S-MCLAIM - T-Revoic	13%	Cont.(33.53%) Elab. (14.56%) Ack.(11.22%)	20%	Ack.(31.60%) Corr. (14.88%) Cont.(10.28%)
S-RELTO - S-RELTO	42%	Cont. (19%) Corr. (10.18%) Ack. (8.32%)	33%	Cont. (11.96%) Corr. (7.61%) Ack. (7.61%)
S-PROEVI - S-PROEVI	41%	Elab. (46.81%) Cont. (40.51%) Corr. (3.10%) Contr. (3.01%)	35%	Cont. (28.57%) Elab. (21.43%) Corr. (6.30%)

Table 2 highlights the key findings from our analysis of discourse relations among talk moves. The first column 'Bigram' shows

the talk move pairs detected in our previous sequential talk move analysis. We selected 4 talk move pairs with a relatively high transitional probability ('Trans Prob') from the teaching and tutoring datasets. The 'Discourse Relations' column shows the portion of the top 3 discourse relations between each pair of the talk moves. We observe from this table that when replying to a students' claim, an actionable suggestion for a teacher or tutor could be **T-PRsACC** via asking a clarification question or simply revoicing the claim with some continuation or elaboration (Figure 14).

When designing a student bot to collaborate with students as they learn mathematics, developers seek to simulate the behaviors of a student, for example as they relate other students or provide evidence across multiple utterances. Figure 12 and Figure 13 elucidates instances of these types of talk move pairs. Further lexical analysis shows that 22% of the talk move pairs in the teaching dataset and 32% in the tutoring dataset use the conjunction word "so" to connect continuous utterances.

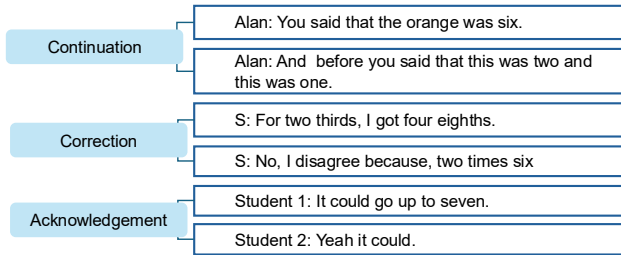


Figure 12: Examples of the **S-RELTo- S-RELTo** with the *Continuation*, *Contrast*, and *Acknowledgement* discourse relations.

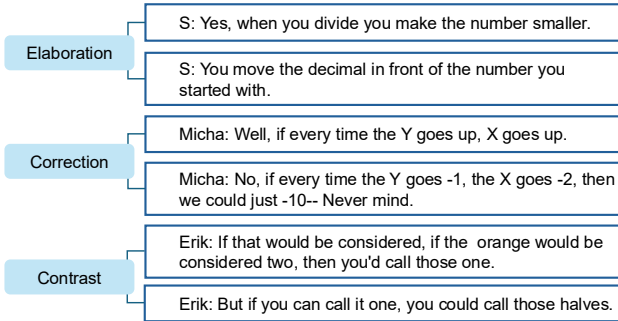


Figure 13: Examples of the **S-PROEVI- S-PROEVI** talk move pair with the *Elaboration*, *Continuation*, *Correction*, and *Contrast* discourse relations.

3.3.2 Importance of Utterances without Talk Moves

We conducted an in-depth analysis of discursive interactions not involving talk moves using two approaches. First, we examined bigram pairs consisting of one talk move and one non-talk move utterance. Second, we analyzed instances where utterances classified as **None** occurred between two talk moves that exhibit significant transition probabilities given our prior analyses. The frequent DAs associated with these significant transition patterns are detailed in Table 3.

Talk Moves → None. The teacher talk moves **T-RESTAT**, **T-KPTG**, and **T-PRsACC** exhibit high transition rates towards utterances classified as **T-NONE** in both the teaching and tutoring domains, as observed in Figure 8a. The predominant discourse relations for these transitions are *Continuation*, *Elaboration*, and

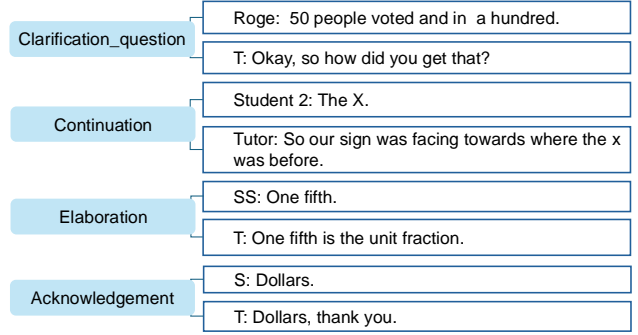


Figure 14: Example of the **S-MCLAIM → T-PRsACC** talk move pair with the *Clarification Question* discourse relation and the **S-MCLAIM → T-REVOIC** talk move pair with the *Continuation*, *Elaboration*, and *Acknowledgement* discourse relations.

Comment. The most frequent DAs associated with **T-NONE** s in this pair are presented in Table 3. Figure 15 presents examples of the **T-RESTAT → T-NONE** pair, showcasing various discourse relations and DAs. From this figure, we can observe that utterances classified as **T-NONE** can actively contribute to the classroom discourse by introducing new information, guiding the flow of conversation, or acknowledging and appreciating student contributions. These acts can play a crucial role in fostering productive classroom dialogue.

Similarly, the frequently associated DAs and discourse relations with the pair **T-KPTG → T-NONE** suggests that the purpose of the **T-KPTG** in these examples is to draw the attention of students and encourage active listening. Utterances classified as **T-NONE** following **T-KPTG** talk moves may provide students with more direction, helping to better engage and understand the flow of the discussion. Our analysis also reveals that the **T-NONE** talk move in the **T-PRsACC → T-NONE** can support **T-PRsACC** by offering additional information and directions, ensuring that a question is clearly conveyed and effectively understood by the students.

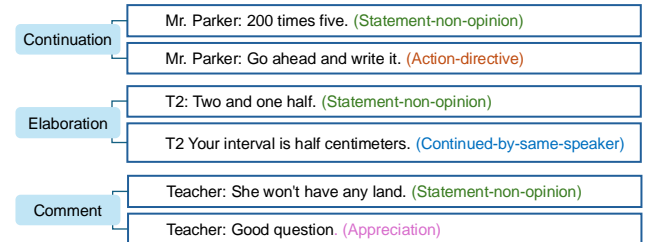


Figure 15: Examples of the **T-RESTAT → T-NONE** pair with different discourse relations and DAs.

Talk Moves → None → Talk Moves. We further analyzed the role of (**T-NONE**) utterances occurring between two consecutive talk moves, focusing first on cases where the talk moves belong to the same category. Based on the qualitative examples illustrated in Figure 16 and the importance of these transitions supported by Table 3, we can see that **T-NONE** between same-category student talk moves help to acknowledge, correct, and guide students, enabling them to refine their understanding while encouraging further contributions to the classroom discourse. Meanwhile, **T-NONE** between same-category teacher talk moves can act as a bridge, seamlessly linking the two moves and creating the effect of a multi-utterance spanning talk move.



Figure 16: Examples of the same-category talk moves separated by a teacher utterance classified as **T-NONE**.

When exploring the role of teacher utterances classified as **T-NONE** in transitions between different talk moves, we observe that these utterances help with elaborating on student contributions, guiding discourse, and fostering engagement. Additionally, **T-NONE** utterances can strengthen the coherence of teacher talk moves, ensuring a fluid and connected sequence of strategic interactions. Illustrative examples of these transitions are presented in Figure 17.

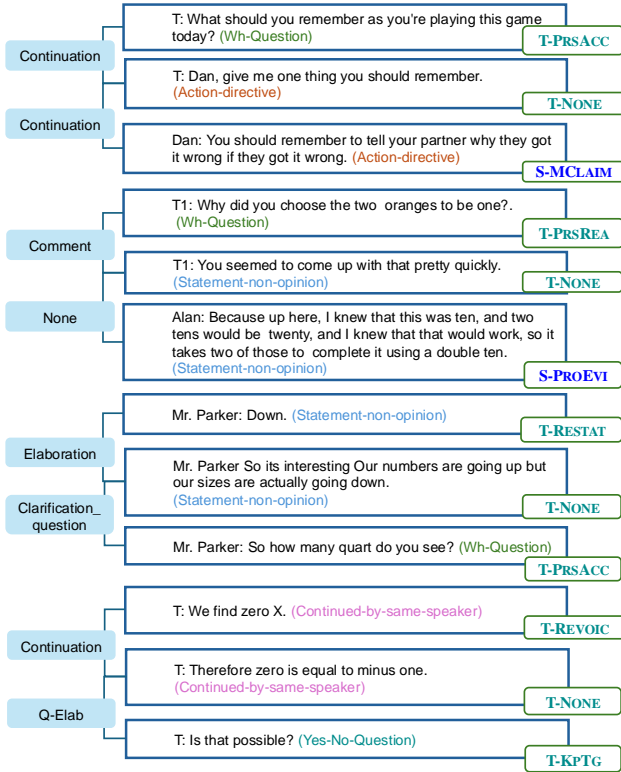


Figure 17: Examples of different-category talk moves separated by a teacher utterance classified as **T-NONE**.

Table 3: Dialogue Acts Associated with Non-Talk Moves in Interaction with Talk Moves. (Dialogue Acts: ContS. = Continued-by-same-speaker, StatNO = Statement-non-opinion, AcD. = Action-directive, AckB. = Acknowledge (Backchannel), R-Ack. = Response-Acknowledgement, Apr. = Appreciation, StatO. = Statement-opinion, Thanking = Thanking, Other = Other).

TalkMove Pairs	Dialogue Acts (of the intervening T-NONE)
Talk Move to Non-Talk Move Transition	
T-RESTAT → T-NONE	ContS. (43.74%), StatNO (19.88%), AcD. (8.39%), AckB. (7.29%), Apr. (3.76%)
T-KpTG → T-NONE	StatNO (26.65%), ContS. (24.05%), AcD. (19.43%), StatO. (4.94%)
T-PRsACC → T-NONE	StatNO (35.93%), ContS. (23.67%), AcD. (20.34%)
S-MCLAIM → T-NONE	ContS. (61.66%), AckB. (14.83%), R-Ack. (5.69%), StatNO (4.18%), Apr. (3.90%)
S-PROEVI → T-NONE	ContS. (54.62%), AckB. (23.15%), R-Ack. (4.57%), Apr. (4.52%), StatNO (4.11%)
S-RELTO → T-NONE	AckB. (14.50%), R-Ack. (7.14%), StatNO (5.09%), AcD. (4.00%)
Intra-Talk Move Transition with Intervening Non-Talk Move	
S-RELTO → T-NONE → S-RELTO	ContS. (55.17%), AckB. (17.24%), StatNO (17.24%), R-Ack. (3.45%)
S-PROEVI → T-NONE → S-PROEVI	ContS. (47.47%), AckB. (33.33%), StatNO (6.06%), AcD. (5.05%)
T-KpTG → T-NONE → T-KpTG	StatNO (27.28%), ContS. (24.30%), AcD. (17.48%), Other (6.24%), StatO. (5.28%)
T-PRsACC → T-NONE → S-MCLAIM	StatNO (39.35%), ContS. (29.60%), AcD. (14.80%), StatO. (3.25%)
T-PRsREA → T-NONE → S-PROEVI	StatNO (28.57%), AcD. (21.43%), ContS. (21.43%), Thanking (7.14%)
T-RESTAT → T-NONE → T-PRsACC	ContS. (42.59%), StatNO (16.67%), AckB. (8.33%), R-Ack. (8.33%), AcD. (6.94%)
T-REVOIC → T-NONE → T-KpTG	ContS. (39.81%), StatNO (19.91%), AcD. (13.89%), AckB. (6.94%), StatO. (3.70%)

4. DISCUSSION

We analyzed pedagogical behaviors in two mathematics education datasets, **TalkMoves** (teaching) and **SAGA22** (tutoring). Using manually annotated talk moves and two state-of-the-art models for dialogue act (DAs) and discourse relation (DRs) prediction, we conducted a top-down analysis on discursive behaviors looking at (1) unigram patterns, (2) sequential patterns, and (3) a deep dive via multi-view analysis.

4.1 Main Findings

Our unigram analysis of utterance-level talk moves and dialogue acts revealed similar overarching distributional patterns across the teaching and tutoring datasets. However, the tutoring dataset (**SAGA22**) exhibited a slightly higher prevalence of utterances without talk moves, with **T-NONE** occurring 5.8% more frequently and **S-NONE** 4.3% more frequently relative to the teaching (**TalkMoves**) dataset. This discrepancy may stem from the fact that well-trained teachers are more likely to employ talk moves, leading to a more structured and intentional discourse in classroom settings. A deeper joint analysis of DAs with talk moves provides a more nuanced perspective on discursive interactions and also helps us further differentiate between the teaching and tutoring datasets. While similar patterns were observed for utterances with and without talk moves in both domains, *Conventional-Closing* played a more significant role in utterances without talk moves during the tutoring sessions. This highlights how the differences (e.g., amounts of students, different training of the teachers/tutoring, and in-person vs online) between teaching and tutoring influences discourse flow, potentially altering communication dynamics. Moreover, the higher ratio of student talk moves followed by another student talk move in the teaching

domain and the higher frequency of student talk moves followed by non-talk move utterances by the instructor in the tutoring domain may indicate that the teachers were more adept at fostering student-driven discussions than the tutors. Additionally, the lower transition ratio to **T-RESTAT** in tutoring sessions suggests a potential area for actionable feedback to the tutors.

Finally, through a detailed analysis of frequent behaviors, we identified meaningful discourse patterns in talk move sequences. We discovered that, beyond talk moves and dialogue acts, the dependency relations within the discourse provide valuable insights that can inform instructor feedback and guide action policies for future AI agents in mathematics education. Our analysis shows how students' discursive participation patterns may reveal important information about their needs and state of mind, enabling teachers and AI agents to adjust their strategies for more targeted support. Additionally, we uncovered patterns in teachers' behavior beyond their talk moves, such as their role as the primary action director in the classroom discourse and their use of various strategies "under the hood" of talk moves to address different situations. Moreover, despite being not being categorized as include a talk move, **T-NONE** can significantly contribute to classroom discourse by guiding discussions, acknowledging student input, and maintaining coherence in teacher talk moves. These utterances help to structure transitions, offering clarification and direction. Between student moves, they can play the role of acknowledging and refining ideas. Between teacher moves, they can act as a bridge and foster continuity. These findings underscore the nuanced role of talk moves and dialogue acts, offering insights to enhance teacher training and the development of intelligent tutoring systems.

4.2 Limitations and Future Work

A limitation of this work is that, due to a lack of annotated dialogue acts and discourse relation data, we only selected two of the available state-of-art schema with public accessible models, SWBD-DAMSL and SDRT for dialogue acts and discourse relation respectively. Designing the most suitable schema for educational dialogue still requires future investigation. Further more, the two off-the-shelf models on SWBD-DAMSL and SDRT were not fine-tuned in our mathematics education datasets **TalkMoves** and **SAGA22**, which may lead to suboptimal results. Future work on annotating a larger amount of in-domain training data will enable potential finetuning or multi-task learning for jointly modeling all three tasks, which could further increase the accuracy of our proposed toolkits for the multi-perspective analysis. In this paper, our analysis mainly focused on educational theory-grounded talk moves analysis with both well-annotated datasets and finetuned models. We mainly focused on bigram talk moves sequences with or without non-talk moves utterances in between. Jointly considering a connected multi-view sub-graph could lead to the discovery of other interesting patterns of pedagogical behaviors. Finally, we did not associate the dialog interactions with an analysis of the instruction quality or student learning outcomes. Future work combining these kinds of data may provide more insights into the nature of high-quality instruction.

4.3 Potential Applications

This work offers an in-depth analysis of mathematics instructional dialogue in both the teaching and tutoring domains, which may enable educators and future AI agents to facilitate more effective discussions with students. Our empirical analysis suggests that dialogue acts and discourse relations could offer a more nuanced understanding of the behavioral patterns of students and teachers

embedded within theory-driven talk moves. A multi-view analysis may lead to actionable feedback provided to educators that goes beyond information based solely on talk moves [38, 39, 40], including important dialogue captured by **T-NONE** and **S-NONE** utterances. Moreover, a deeper understanding of these dialogue structures can inform the development of AI-driven tutoring systems in mathematics education [9, 44, 66]. By integrating these findings with advancements in generative AI and domain-specific datasets, this research paves the way for broader applications beyond mathematics. Potential areas of impact include collaborative learning [13, 17, 16, 43], creative problem solving [12, 73], and fostering meaningful student engagement [37]. This work lays the foundation for AI-driven educational tools that enhance explainable instructional quality, promote controllable learning experiences, and support the evolving needs of both educators and students.

5. CONCLUSION

Integrating talk moves, dialogue acts, and discourse relations, our multi-perspective study reveals key insights into the nature of teaching and tutoring discourse across two mathematics education datasets: **TalkMoves** and **SAGA22**. Our analysis reveals that while teaching and tutoring datasets share overarching distributional similarities in their teacher and student talk moves, tutoring interactions exhibit a slightly higher prevalence of utterances not classified as containing a talk move, suggesting a need for targeted tutor training. Joint dialogue act analysis underscores the nuanced role of diverse dialogue acts in enhancing strategic communication. Notably, transition analysis highlights tutors' greater reliance on utterances that do not contain a talk move and a reduced tendency to restate students' ideas, suggesting that actionable feedback in these areas might be appropriate. Furthermore, frequent teacher-student interaction patterns align with core educational clusters, emphasizing the structured nature of pedagogical discourse. Beyond utterance-level talk moves, discourse dependency relations offer insights for optimizing AI-driven educational agents. Our deeper exploration into utterances without talk moves reveals their essential function in shaping classroom dialogue. Rather than mere fillers, they can serve as pivotal elements in guiding, acknowledging, and structuring discourse. Whether linking teacher talk moves for coherence or scaffolding student contributions, utterances with and without talk moves likely play a crucial role in engagement and comprehension. These findings underscore the importance of integrating discourse relations and dialogue acts into AI-assisted education to foster more effective and responsive learning environments.

6. ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable feedback. This research was supported by the National Science Foundation grant #2222647 and the NSF National AI Institute for Student-AI Teaming (iSAT) under grant DRL #2019805. All opinions are those of the authors and do not reflect those of the funding agencies.

7. REFERENCES

- [1] S. Afantenos, N. Asher, F. Benamara, M. Bras, C. Fabre, M. Ho-dac, A. L. Draoulec, P. Muller, M.-P. Péry-Woodley, L. Prévot, J. Rebeyrolles, L. Tanguy, M. Vergez-Couret, and L. Vieu. An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odiijk, and S. Piperidis, editors, *Proceedings of the Eighth International*

- Conference on Language Resources and Evaluation (LREC'12)*, pages 2727–2734, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA).
- [2] J. Allen and M. Core. Draft of damsl: Dialog act markup in several layers, 1997.
 - [3] J. Allwood. An activity based approach to pragmatics. 1995.
 - [4] N. Asher, J. Hunter, M. Morey, B. Farah, and S. Afantenos. Discourse structure and dialogue acts in multiparty dialogue: the STAC corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2721–2727, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).
 - [5] N. Asher and A. Lascarides. *Logics of conversation*. Cambridge University Press, 2003.
 - [6] J. L. Austin. *How to do things with words*. Oxford university press, 1975.
 - [7] F. Benamara and M. Taboada. Mapping different rhetorical relation annotations: A proposal. In M. Palmer, G. Boleda, and P. Rosso, editors, *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 147–152, Denver, Colorado, June 2015. Association for Computational Linguistics.
 - [8] Z. Bennis, J. Hunter, and N. Asher. A simple but effective model for attachment in discourse parsing with multi-task learning for relation labeling. In A. Vlachos and I. Augenstein, editors, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3412–3417, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics.
 - [9] M. Z. bin Mohamed, R. Hidayat, N. N. binti Suhaizi, M. K. H. bin Mahmud, S. N. binti Baharuddin, et al. Artificial intelligence in mathematics education: A systematic literature review. *International Electronic Journal of Mathematics Education*, 17(3):em0694, 2022.
 - [10] R. Böheim, K. Schnitzler, A. Gröschner, M. Weil, M. Knogler, A.-K. Schindler, M. Alles, and T. Seidel. How changes in teachers' dialogic discourse practice relate to changes in students' activation, motivation and cognitive engagement. *Learning, Culture and Social Interaction*, 28:100450, 2021.
 - [11] B. M. Booth, J. Jacobs, J. B. Bush, B. Milne, T. Fischhaber, and S. K. DMello. Human-tutor coaching technology (htct): Automated discourse analytics in a coached tutoring model. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pages 725–735, 2024.
 - [12] L. Boussioux, J. N. Lane, M. Zhang, V. Jacimovic, and K. R. Lakhani. The crowdless future? generative ai and creative problem-solving. *Organization Science*, 35(5):1589–1607, 2024.
 - [13] T. Breideband, J. Bush, C. Chandler, M. Chang, R. Dickler, P. Foltz, A. Ganesh, R. Lieber, W. R. Penuel, J. G. Reitman, et al. The community builder (cobi): Helping students to develop better small group collaborative learning skills. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, pages 376–380, 2023.
 - [14] J. Cai, B. D. King, M. Perkoff, S. Dudy, J. Cao, M. Grace, N. Wojarnik, G. Ananya, J. Martin, M. Palmer, M. Walker, and J. Flanagan. Dependency dialogue acts — annotation scheme and case study. *The 13th International Workshop on Spoken Dialogue Systems Technology*, 2022.
 - [15] E. Calcagni, F. Ahmed, A. L. Trigo-Clapés, R. Kershner, and S. Hennessy. Developing dialogic classroom practices through supporting professional agency: Teachers' experiences of using the t-seda practitioner-led inquiry approach. *Teaching and Teacher Education*, 126:104067, 2023.
 - [16] J. Cao, R. Dickler, M. Grace, J. B. Bush, A. Roncone, L. M. Hirshfield, M. A. Walker, and M. S. Palmer. Designing an ai partner for jigsaw classrooms. Los Angeles, California., 2023.
 - [17] J. Cao, A. Ganesh, J. Cai, R. Southwell, E. M. Perkoff, M. Regan, K. Kann, J. H. Martin, M. Palmer, and S. D'Mello. A comparative analysis of automatic speech recognition errors in small group classroom discourse. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, pages 250–262, 2023.
 - [18] J. Cao, A. Suresh, J. Jacobs, C. Clevenger, A. Howard, C. Brown, B. Milne, T. Fischhaber, T. Sumner, and J. H. Martin. Enhancing talk moves analysis in mathematics tutoring through classroom teaching discourse. In O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, and S. Schockaert, editors, *Proceedings of the 31st International Conference on Computational Linguistics*, pages 7671–7684, Abu Dhabi, UAE, Jan. 2025. Association for Computational Linguistics.
 - [19] T.-C. Chi and A. Rudnicky. Structured dialogue discourse parsing. In O. Lemon, D. Hakkani-Tur, J. J. Li, A. Ashrafzadeh, D. H. Garcia, M. Alikhani, D. Vandyke, and O. Dušek, editors, *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 325–335, Edinburgh, UK, Sept. 2022. Association for Computational Linguistics.
 - [20] P. R. Cohen, J. L. Morgan, and M. E. Pollack. *Intentions in communication*. MIT press, 1990.
 - [21] M. G. Core and J. Allen. Coding dialogs with the damsl annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, volume 56, pages 28–35. Boston, MA, 1997.
 - [22] R. Correnti, L. C. Matsumura, M. Walsh, D. Zook-Howell, D. D. Bickel, and B. Yu. Effects of online content-focused coaching on discussion quality and reading achievement: Building theory for how coaching develops teachers' adaptive expertise. *Reading Research Quarterly*, 56(3):519–558, 2021.
 - [23] R. Correnti, M. K. Stein, M. S. Smith, J. Scherrer, M. McKeown, J. Greeno, and K. Ashley. Improving teaching at scale: Design for the scientific measurement and learning of discourse practice. *Socializing Intelligence Through Academic Talk and Dialogue*. AERA, 284, 2015.
 - [24] D. Datta, J. P. Bywater, M. Phillips, S. Lilly, J. L. Chiu, G. S. Watson, and D. E. Brown. Classifying mathematics teacher questions to support mathematical discourse. In *International Conference on Artificial Intelligence in Education*, pages 372–377. Springer, 2023.
 - [25] D. Demszky and H. Hill. The ncte transcripts: A dataset of elementary math classroom transcripts. *arXiv preprint arXiv:2211.11772*, 2022.
 - [26] D. Demszky and J. Liu. M-powering teachers: Natural language processing powered feedback improves 1: 1 instruction and student outcomes. In *Proceedings of the Tenth ACM Conference on Learning@ Scale*, pages 59–69, 2023.
 - [27] D. Demszky, J. Liu, H. C. Hill, D. Jurafsky, and C. Piech. Can automated feedback improve teachers' uptake of student ideas? evidence from a randomized controlled trial in a large-scale online course. *Educational Evaluation and Policy Analysis*, 46(3):483–505, 2024.
 - [28] D. Demszky, J. Liu, H. C. Hill, S. Sanghi, and A. Chung.

- Automated feedback improves teachers' questioning quality in brick-and-mortar classrooms: Opportunities for further enhancement. *Computers & Education*, 227:105183, 2025.
- [29] P. J. Donnelly, N. Blanchard, B. Samei, A. M. Olney, X. Sun, B. Ward, S. Kelly, M. Nystran, and S. K. D'Mello. Automatic teacher modeling from live classroom audio. In *Proceedings of the 2016 conference on user modeling adaptation and personalization*, pages 45–53, 2016.
- [30] R. K. Franklin, J. O'Neill Mitchell, K. S. Walters, B. Livingston, M. B. Lineberger, C. Putman, R. Yarborough, and L. Karges-Bone. Using swivl robotic technology in teacher education preparation: A pilot study. *TechTrends*, 62:184–189, 2018.
- [31] Y. Fu. Towards unification of discourse annotation frameworks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 132–142, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [32] J. J. Godfrey, E. C. Holliman, and J. McDaniel. Switchboard: Telephone speech corpus for research and development. In *Acoustics, speech, and signal processing, IEEE international conference on*, volume 1, pages 517–520. IEEE Computer Society, 1992.
- [33] M. Hancher. The classification of cooperative illocutionary acts1. *Language in society*, 8(1):1–14, 1979.
- [34] Z. He, L. Tavabi, K. Lerman, and M. Soleymani. Speaker turn modeling for dialogue act classification. In M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, editors, *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2150–2157, Punta Cana, Dominican Republic, Nov. 2021. Association for Computational Linguistics.
- [35] J. R. Hobbs. *Literature and cognition*. Number 21. Center for the Study of Language (CSLI), 1990.
- [36] C. Howe, S. Hennessy, N. Mercer, M. Vrikki, and L. Wheatley. Teacher–student dialogue during classroom teaching: Does it really impact on student outcomes? *Journal of the learning sciences*, 28(4-5):462–512, 2019.
- [37] A. Y. Huang, O. H. Lu, and S. J. Yang. Effects of artificial intelligence-enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, 194:104684, 2023.
- [38] J. Jacobs, K. Scornavacco, C. Harty, A. Suresh, V. Lai, and T. Sumner. Promoting rich discussions in mathematics classrooms: Using personalized, automated feedback to support reflection and instructional change. *Teaching and Teacher Education*, 112:103631, 2022.
- [39] E. Jensen, M. Dale, P. J. Donnelly, C. Stone, S. Kelly, A. Godley, and S. K. D'Mello. Toward automated feedback on teacher discourse to enhance teacher learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.
- [40] E. Jensen, M. Dale, P. J. Donnelly, C. Stone, S. Kelly, A. Godley, and S. K. D'Mello. Toward Automated Feedback on Teacher Discourse to Enhance Teacher Learning. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.
- [41] D. Jurafsky. Switchboard swbd-damsl shallow-discourse-function annotation coders manual. *Institute of Cognitive Science Technical Report*, 1997.
- [42] H. Kamp, J. v. Genabith, and U. Reyle. Discourse representation theory. In *Handbook of philosophical logic*, pages 125–394. Springer, 2011.
- [43] J. Kim, H. Lee, and Y. H. Cho. Learning design to support student-ai collaboration: Perspectives of leading teachers for ai in education. *Education and Information Technologies*, 27(5):6069–6104, 2022.
- [44] A. Latham. Conversational Intelligent Tutoring Systems: The State of the Art. In A. E. Smith, editor, *Women in Computational Intelligence: Key Advances and Perspectives on Emerging Topics*, Women in Engineering and Science, pages 77–101. Springer International Publishing, Cham, 2022.
- [45] A. Lefstein and J. Snell. *Better than best practice: Developing teaching and learning through dialogue*. Routledge, 2013.
- [46] C. Li, C. Braud, M. Amblard, and G. Carenini. Discourse relation prediction and discourse parsing in dialogues with minimal supervision. In M. Strube, C. Braud, C. Hardmeier, J. J. Li, S. Loaiciga, A. Zeldes, and C. Li, editors, *Proceedings of the 5th Workshop on Computational Approaches to Discourse (CODI 2024)*, pages 161–176, St. Julians, Malta, Mar. 2024. Association for Computational Linguistics.
- [47] J. Li, M. Liu, M.-Y. Kan, Z. Zheng, Z. Wang, W. Lei, T. Liu, and B. Qin. Molweni: A challenge multiparty dialogues-based machine reading comprehension dataset with discourse structure. In D. Scott, N. Bel, and C. Zong, editors, *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2642–2652, Barcelona, Spain (Online), Dec. 2020. International Committee on Computational Linguistics.
- [48] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.
- [49] W. C. Mann and S. A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse*, 8(3):243–281, 1988.
- [50] L. C. Matsumura, H. E. Garnier, S. C. Slater, and M. D. Boston. Toward measuring instructional interactions “at-scale”. *Educational Assessment*, 13(4):267–300, 2008.
- [51] N. Mercer, R. Wegerif, and L. C. Major. *The Routledge international handbook of research on dialogic education*. Routledge Abingdon, 2019.
- [52] S. Michaels, M. W. Hall, and L. B. Resnick. *Accountable talk sourcebook: For classroom conversation that works*. University of Pittsburgh Pittsburgh, PA, 2013.
- [53] S. Michaels, M. C. O'Connor, M. W. Hall, and L. B. Resnick. *Accountable talk® sourcebook. Pittsburg, PA: Institute for Learning University of Pittsburgh. Murphy, PK, Wilkinson, IAG, Soter, AO, Hennessey, MN, & Alexander, JF*, 2010.
- [54] B. Moreau-Pernet, Y. Tian, S. Sawaya, P. Foltz, J. Cao, B. Milne, and T. Christie. Classifying tutor discursive moves at scale in mathematics classrooms with large language models. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale, L@S '24*, page 361–365, New York, NY, USA, 2024. Association for Computing Machinery.
- [55] C. O'Connor and S. Michaels. Supporting teachers in taking up productive talk moves: The long road to professional learning at scale. *International Journal of Educational Research*, 97:166–175, 2019.
- [56] E. M. Perkoff, A. Bhattacharyya, J. Cai, and J. Cao. Comparing neural question generation architectures for reading comprehension. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational*

- Applications (BEA 2023)*, pages 556–566, 2023.
- [57] R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. The penn discourse treebank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, 2008.
 - [58] J. L. Ramos, A. A. Cattaneo, F. P. de Jong, and R. G. Espadeiro. Pedagogical models for the facilitation of teacher professional development via video-supported collaborative learning. a review of the state of the art. *Journal of Research on Technology in Education*, 54(5):695–718, 2022.
 - [59] B. Reese, J. Hunter, N. Asher, P. Denis, and J. Baldridge. *Reference manual for the analysis and annotation of rhetorical structure*. PhD thesis, University of Texas at Austin, 2007.
 - [60] L. B. Resnick, C. S. C. Asterhan, S. N. Clarke, and F. Schantz. *Next Generation Research in Dialogic Learning*, chapter 13, pages 323–338. John Wiley & Sons, Ltd, 2018.
 - [61] J. R. Searle and J. R. Searle. *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge university press, 1969.
 - [62] A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. V. Ess-Dykema, and M. Meteer. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373, 2000.
 - [63] A. Suresh, J. Jacobs, C. Clevenger, V. Lai, C. Tan, J. H. Martin, and T. Sumner. Using ai to promote equitable classroom discussions: The talkmoves application. In *Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part II*, pages 344–348. Springer, 2021.
 - [64] A. Suresh, J. Jacobs, C. Harty, M. Perkoff, J. H. Martin, and T. Sumner. The talkmoves dataset: K-12 mathematics lesson transcripts annotated for teacher and student discursive moves. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4654–4662, 2022.
 - [65] A. Suresh, J. Jacobs, M. Perkoff, J. H. Martin, and T. Sumner. Fine-tuning transformers with additional context to classify discursive moves in mathematics classrooms. In *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, pages 71–81, 2022.
 - [66] A. Tack and C. Piech. The ai teacher test: Measuring the pedagogical ability of blender and gpt-3 in educational dialogues. *arXiv preprint arXiv:2205.07540*, 2022.
 - [67] K. Thompson, A. Chaturvedi, J. Hunter, and N. Asher. Llamipa: An incremental discourse parser. In Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, editors, *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 6418–6430, Miami, Florida, USA, Nov. 2024. Association for Computational Linguistics.
 - [68] K. Thompson, J. Hunter, and N. Asher. Discourse structure for the Minecraft corpus. In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, and N. Xue, editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 4957–4967, Torino, Italia, May 2024. ELRA and ICCL.
 - [69] N. Tran, B. Pierce, D. Litman, R. Correnti, L. C. Matsumura, et al. Multi-dimensional performance analysis of large language models for classroom discussion assessment. *Journal of Educational Data Mining*, 16(2):304–335, 2024.
 - [70] D. Wang, D. Shan, Y. Zheng, K. Guo, G. Chen, and Y. Lu. Can chatgpt detect student talk moves in classroom discourse? a preliminary comparison with bert. 2023.
 - [71] N. M. Webb, M. L. Franke, M. Ing, N. C. Johnson, and J. Zimmerman. The details matter in mathematics classroom dialogue. In L. M. Neil Mercer, Rupert Wegerif, editor, *The Routledge international handbook of research on dialogic education*, pages 530–546. Routledge, 2019.
 - [72] L. Wittgenstein. *Philosophical investigations*. John Wiley & Sons, 2010.
 - [73] Z. Wu, D. Ji, K. Yu, X. Zeng, D. Wu, and M. Shidujaman. Ai creativity and the human-ai co-creation model. In *Human-Computer Interaction. Theory, Methods and Tools: Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part I 23*, pages 171–190. Springer, 2021.