Int J Artif Intell Educ (2025) 35:677–701
https://doi.org/10.1007/s40593-024-00436-8

ARTICLE

Recognizing Multi-Party Epistemic Dialogue Acts During Collaborative Game-Based Learning Using Large Language Models

Halim Acosta[1] · Seung Lee[1] · Haesol Bae[3] · Chen Feng[2] · Jonathan Rowe[1] · Krista Glazewski[4] · Cindy Hmelo-Silver[2] · Bradford Mott[1] · James C. Lester[1]

Accepted: 18 October 2024 / Published online: 14 November 2024

Abstract

Understanding students' multi-party epistemic and topic based-dialogue contributions, or how students present knowledge in group-based chat interactions during collaborative game-based learning, offers valuable insights into group dynamics and learning processes. However, manually annotating these contributions is labor-intensive and challenging. To address this, we develop an automated method for recognizing dialogue acts from text chat data of small groups of middle school students interacting in a collaborative game-based learning environment. Our approach utilizes dual contrastive learning and label-aware data augmentation to fine-tune large language models' underlying embedding representations within a supervised learning framework for epistemic and topic-based dialogue act classification. Results show that our method achieves a performance improvement of 4% to 8% over baseline methods in two key classification scenarios. These findings highlight the potential for automated dialogue act recognition to support understanding of how meaning-making occurs by focusing on the development and evolution of knowledge in group discourse, ultimately providing teachers with actionable insights to better support student learning.

Keywords Natural language processing · Game-based learning · Dialogue act recognition · Collaborative inquiry

Introduction

Problem-based learning is an instructional approach that emphasizes learning through collaborative problem solving in small groups. Problem based learning expects students to engage in complex problem solving by actively collaborating and negotiating through conversation (Hmelo-Silver, 2004). During these interactions students demonstrate a wide range of epistemic statuses while replying to each other to facilitate a continuous conversation flow (Heritage, 2012b). Epistemic status, an individual's beliefs about knowledge or knowing, is a principal factor in students' academic suc-

Extended author information available on the last page of the article

cess and their ability to learn (Hofer & Pintrich, 1997; Greene & Azevedo, 2007; Muis & Franco, 2009). For example, in a problem-based learning scenario, a student might express a high epistemic status when presenting a well-researched solution, while another student could exhibit a more uncertain epistemic status when exploring new ideas. By examining students' epistemic statuses researchers and educators can gain insight into how students approach collaborative tasks and uncover ways to promote effective collaborative behaviors. In conjunction with analyzing students' content-based dialogue contributions, identifying student epistemic statuses can help to improve the quality of interactions between individuals in a collaborative setting.

Computer-supported collaborative learning and game-based learning have been shown to be promising approaches to improving learning outcomes and promoting student engagement (Hainey et al., 2011; Connolly et al., 2012). Computer supported collaborative learning utilizes technology-mediated communication to facilitate collaborative discourse while game-based learning employs digital games to enhance motivation and learning. Current research shows that these techniques can enable researchers to collect and analyze student communication and interactions (Dillenbourg & Hong, 2008; Rienties & Alden, 2014; Tabuenca et al., 2015). Uncovering the patterns and dynamics that underlie these exchanges can provide a lens into the wide range of cognitive, affective, and social processes involved in learning. Game-based learning environments additionally provide an immersive and interactive environment in which to observe student behaviors while promoting engaging interactions. Recent efforts have been underway to explore how to combine the benefits of computer supported collaborative learning and game-based learning to create collaborative game-based learning environments (Sung & Hwang, 2013; de Jesus & Silveira, 2019). These frameworks are designed to help researchers understand student collaboration processes while creating engaging learning experiences.

Dialogue act recognition, the process of automatically identifying and categorizing the communicative intention within spoken or written conversation, has many potential applications in computer-supported collaborative learning where it can be employed to analyze and understand the nature of students' dialogue and collaboration (Katuka et al., 2021; Dascalu et al., 2022). Automatic dialogue act recognition can provide insight into the quality of group interactions and the effectiveness of learning activities. Further, dialogue act recognition has been used to identify patterns of interactions that promote positive collaborative problem solving and detect conflicts in communication (Stahl, 2006). However, applying dialogue act recognition models in scenarios with small amounts of training data can be challenging due to low availability of labeled data or the annotation costs associated with unlabeled data. Additionally, pretraining language models provide potential avenues to address these challenges by leveraging the wealth of pre-existing linguistic knowledge they encode. Prior work has shown success in incorporating pre-trained BERT and other transformer variants for the task of dialogue act recognition (Kumaran et al., 2023; Li & Chen, 2023). In contrast, there has been little work showing the effectiveness of the T5 model (Raffel et al., 2020), a transformer variant with a different set of pretraining tasks, for the dialogue recognition problem. Given this background we were motivated to explore the following research questions:

RQ1*: How effectively can pre-trained large language models discern student epistemic and topic-related contributions within the context of multi-party collaborative discourse, and what are the implications for AI-enabled learning environments?*

RQ2*: What qualitative and quantitative differences emerge when using traditional machine learning and BERT-based approaches versus the T5 model for recognizing students' dialogue acts in collaborative discourse?*

RQ3*: In what ways can the performance of T5 models be enhanced for student epistemic and topic related dialogue act recognition in collaborative discourse through methods such as dual contrastive fine-tuning, label-aware data augmentation, and incorporation of contextual information?*

In this work, we employ the T5 model on two tasks: (1) recognizing students' epistemic responses and (2) recognizing students topic related contributions, from group chat messages in a collaborative game-based learning environment, EcoJourneys. We compare their performance to baseline machine learning methods and BERT-based models. Further, we leverage a dual contrastive learning approach with label-aware data augmentation to enhance the performance of T5 by maximizing the interrelationship between input embeddings and their respective labels. An investigation was conducted to examine how well these language models perform on a set of epistemic and topic related contributions produced by student interactions, with only a modest amount of training and validation data. Results indicate that the T5 models show an improvement in performance over traditional machine learning techniques (random forest, logistic regression) and BERT-based approaches. Moreover, we find that providing context information along with using label-aware data augmentation while employing dual contrastive learning significantly outperforms baseline methods while enhancing the performance of the T5 model.

The primary contribution of this work is a T5-based model for recognizing students' epistemic responses and topic-related contributions in a collaborative game-based learning environment. We specifically demonstrate that employing the T5 model with dual contrastive fine tuning and label-aware data augmentation, shows improved performance over traditional machine learning and BERT-based approaches, particularly when provided with conversational context information. Our approach additionally achieves an improved agreement with human annotations highlighting its effectiveness for automatic dialogue act recognition.

Background

Dialogue Act Recognition

Dialogue act recognition, which is the task of identifying the communicative function of an utterance within a dialogue, has traditionally been performed using statistical methods (Stolcke et al., 2000; Chen et al., 2018a; Blache et al., 2020). However, recent advances have centered on deep learning for its ability to automatically extract meaningful features from data (Cai et al., 2022; Mezza et al., 2022; Vielsted et al., 2022). For instance, Liu and Lane (2016) developed an attention-based recurrent neural network (RNN) architecture that leverages bidirectional LSTMs and transformer-based mecha-

nisms for improved intent classification and slot label prediction. More recently, graph convolutional networks, utilizing heterogeneous user histories, have demonstrated superior performance compared to transformer architectures (Wang et al., 2020).

While many of these efforts have focused on deep learning techniques, our work distinguishes itself by fine-tuning large language models, particularly the T5 model, and integrating label-aware data augmentation, dual contrastive fine-tuning, and context awareness. This enables us to address challenges in low-resource settings, where traditional approaches often rely on substantial annotated data and computational resources (Wei et al., 2022; Pengfei & Yinglong, 2022; Tan et al., 2023). By leveraging dual contrastive fine-tuning and label-aware data augmentation, we augment the training data to enhance model performance without increasing computational overhead. This approach contrasts with methods such as multi-task learning (McLeod et al., 2019), which primarily improve performance through auxiliary tasks. Our dual contrastive fine-tuning framework further refines the model by introducing additional context from multi-party student dialogues, allowing for improved differentiation of epistemic stances in collaborative settings.

The majority of dialogue act recognition research has largely focused on dyadic conversations, but the multi-party nature of much educational discourse presents unique challenges due to the variability in participant interactions. Previous approaches have addressed these challenges with approaches such as Bayesian networks (Dielmann & Renals, 2008), speaker-aware networks (Yu et al., 2022), and graph-based neural networks (Wang et al., 2021). Our model's ability to identify epistemic stance in these multi-party interactions-supported by our use of context-aware annotations and fine-tuning methods-enhances the interpretability and utility of student chat data analysis in educational settings.

Epistemic Dialogue

The concept of accountable talk involves classroom discursive practices that support students to engage in argumentation (O'Connor & Michaels, 2019), while the transactive discussion framework refers to reasoning that builds on previous contributions/dialogue acts from others (Berkowitz & Gibbs, 1983). Both frameworks focus on epistemics which involve how knowledge claims are asserted and negotiated through shared interactions (Heritage, 2012b). In conversation, participants demonstrate varying levels of knowledge, and their joint epistemic status establishes information ownership and rights. In this context, epistemic stance refers to how individuals position themselves relative to the knowledge discussed in their turn. While a student's epistemic stance is influenced by their epistemic status (Heritage, 2012a), it can change throughout a conversation due to group interaction dynamics. Research has shown that computer-supported collaborative learning environments are effective in supporting these dialogue act practices and impact instructional objectives (Chen et al., 2018b; Jeong et al., 2019). Thus, understanding student epistemic stance and status is a crucial aspect to comprehend when examining discourse within educational settings (Saleh et al., 2021).

In this study's chat data, we focus on epistemic stance, which reflects how individuals present their knowledge. While epistemic stance often correlates with epistemic status, they are not always equivalent. For example, a student may offer many suggestions but perform poorly on assessments, revealing a gap between perceived and actual knowledge. Thus, inferring epistemic status from stance alone can be unreliable. In education, epistemic stance helps evaluate students' contributions during discussions, revealing how they perceive their knowledge and interact with peers. Students asserting high knowledge may dominate group discussions, deterring participation, while more modest students might withhold valuable insights (Baker et al., 2013; Kääntä, 2014; Solem, 2016).

Automatically tagging and analyzing chat data can contribute to understanding how knowledge is exchanged during collaborative learning (Saleh et al., 2021). Topic-based annotations can assess the quality of discourse and guide teachers in facilitating discussions. By considering epistemic stance alongside traditional assessments, educators gain a more comprehensive view of students' learning. This fosters self-awareness, encourages collaboration, and enhances learning outcomes (Suthers & Hundhausen, 2003; Scardamalia & Bereiter, 2006; Chi & Wylie, 2014). We use this annotation scheme to label student chat messages from a collaborative game-based learning environment, offering insights into students' knowledge presentation and content-related contributions.

The research reported in this article builds on prior work on epistemic stance in educational dialogue (Heritage, 2012a), by extending it with automatic tagging and analysis of in-game chat data. By employing label-aware data augmentation, dual contrastive fine-tuning, in combination with context awareness our approach not only improves prediction accuracy but also provides a method to automatically identify how students present and negotiate knowledge during collaborative tasks. In doing so, we offer a novel framework for analyzing the relationship between students' epistemic stance and their learning outcomes, contributing to a more nuanced understanding of how knowledge is constructed and shared in computer-supported collaborative learning environments.

Dataset

In section, we provide a detailed overview of our data collection methods and preprocessing steps. First, we describe the EcoJourneys environment. We then provide an overview of demographics of the participants and the structure of the gameplay sessions. Next, we elaborate on the nature and volume of the chat data generated during these interactions. Finally, we outline the preprocessing techniques employed to clean the chat data as well as describe the annotation methodologies (Fig. 1).

EcoJourneys Game-Based Learning Environment

EcoJourneys is a collaborative game-based learning environment designed to teach middle school students about ecosystems (Saleh et al., 2019). In this environment,

Fig. 1 Student preparing for collaborative interaction in EcoJourneys

students visit a virtual remote island to investigate the cause of a mysterious illness affecting the local fish population. Students work in groups of four, each on their own laptops, interacting with the virtual environment by talking to non-player characters (NPCs) who act as local experts, collecting and analyzing information, and discussing their findings via a persistent in-game text chat interface.

During their interactions, students gather information about ecosystem concepts from NPCs and explore the island to uncover clues about the illness. At predefined intervals, students come together at a virtual whiteboard within the game to collaboratively share and categorize the collected information, discuss potential causes of the fish illness, and develop hypotheses. This whiteboard session is facilitated through the game's chat interface, encouraging students to negotiate, share ideas, ask questions, and engage in meaningful discourse.

Each group includes a facilitator, who is either a researcher or a teacher, tasked with supporting the students' learning activities. Facilitators monitor the students' activities and conversations using a separate in-game view and can intervene to guide the learning process. They encourage student engagement by asking questions and selecting from a set of pre-authored messages or providing free-form messages via the in-game chat interface. This structured yet interactive setup ensures that students are actively involved in problem-solving and learning about ecosystem science throughout their gameplay.

Data Collection

Data was collected, with informed consent, from eighteen groups, each consisting of 3-4 middle school students (aged 11-13) totaling 72 students (31 female, 41 male) while they interacted with the EcoJourneys collaborative game-based learning environment. Students engaged with EcoJourneys over multiple class periods, with the average game play time of 32.33 minutes (min = 8.48, max = 57.32, SD = 13.78) over

an average of 6 classes total. With a total of 8,781 chat messages collected during the student interactions with the game, there was substantial variation in the number of chat messages sent across groups. The largest number of chat messages sent by a group was 1,091 messages (12.9% of total), and the smallest number of messages sent by a group being 138 messages (1.6% of total). The average number of chat messages per group was 485 (min = 138, max = 1,091, SD = 267). The average number of chat messages sent per student was approximately 60 (min = 0, max = 1072, SD = 151.52) over the course of all game play sessions. We preprocessed the data by removing messages that did not contain words (e.g. ":)", "!!!!", etc.). Additionally, we replaced words with many repeated characters with their original counterparts (e.g. "NOOOOOOOO!" replaced with "No") in order to have clear representation of the semantic meaning of an input utterance. After applying preprocessing for text cleaning, a total of 8,740 chat messages were used for training, testing, and validation. Appendix A (Table 6) provides additional descriptive statistics for the students multi-party chat data.

Annotations

Overall, 8,938 chat lines from 18 groups were distributed for annotation by two subject matter experts. Initially, the annotators were trained using 377 lines. Out of the total chat lines, 20% (1,788 lines) were used to assess Inter-rater Reliability (IRR) after two training periods. For the Epistemics annotation, a Cohen's Kappa of .909 was achieved, and for Topic annotation, a Cohen's Kappa of .925 was reached. Following the IRR phase, the remaining 6,773 lines were roughly equally distributed between the two annotators for further labeling. Of these 8,938 utterances approximately 157 lines were removed since they represented system generated messages.

The annotations fall under two broad categories: epistemic stance and topic-based labels. Epistemic stance labels include 5 categories. These include "K-" (17.2%) utterances, which are hedges or requests; "K+" (13.5%) utterances offering an index for information; "Reply" (29.4%) messages consisting of basic and simple replies to questions; "Reply-knowledge" (18.5%) messages, which provide content-based information. "Social-organization", referring to greetings and other social references, and "Other" (21.4%) referring to any other acts that maintain discussion, are condensed into a single category of "Other". Tables 1 & 2 show examples of both epistemic stance and topic-based annotations respectively.

Topic labels include "Content" (27.4%), which refers to any utterance describing cognition or science related content; "Task" (42.3%) labels, which cover utterances related to procedural actions or directions and often describe ongoing actions or tasks; "Socio-emotional" (15.7%) labels, which are chat messages that are concerned with building social rapport within a group and are considered on-task but do not fall explicitly under "Content" or "Task" labels; and "Other" (14.6%) labels, which refer to any utterance that does not fall under the prior three categories. Under this annotation scheme two trained annotators achieved a high inter-rater reliability score (Kappa = 0.87).

Table 1 Epistemic response label examples

Code	Description	Example
K+	Utterances that index information	"We know cyanobacteria makes its own food"
K-	Utterances that are hedges or requests	"What does being intolerant mean"
Reply	Basic and simple replies to questions	"Agree"
Reply-knowledge	Replies that provide content-based information	"Because that means there isn't enough dissolved oxygen"
Other	Other acts that maintain discussion	"Thanks:)"

Methods

We employed three types of language models, the T5 model (Raffel et al., 2020) and two BERT-based models, to perform dialogue act recognition in group chats collected in EcoJourneys. We investigated the performance of these models for inferring epistemic and topic-based dialogue acts from group text chat messages as students have collaborative dialogue within the learning environment. We compared BERT-based and T5 models against baseline machine learning techniques that rely on manually generated natural language features. Our baseline models consisted of logistic regression and random forest. These utilized a feature representation consisting of TF-IDF vectors of student utterances as well as the cosine similarity of student's utterances to content-related text in the EcoJourneys learning environment.

Bidirectional Encoder Representations from Transformers (BERT) is a transformer-based model trained on a large corpus for language modeling and next sentence

Table 2 Topic-based label examples

Code	Description	Example
Content	Utterances about science content and/or reasoning behind a hypothesis	"Do the fish need high temp or low temp? Can anyone explain?"
Task	Utterances related to in-game activities	"I am also waiting at the whiteboard"
Socio-emotional	Utterances focused on building social rapport	"Welcome back!"
Other	Utterances that do not fall within the other three categories	"I love tea"

prediction (Devlin et al., 2018). These encoder transformer-based models can learn latent representations of the data that preserve contextual and semantic information. We employed the base BERT model with a fully connected feed-forward classification head for dialogue act recognition. Additionally, we experimented with RNN based heads in the form of LSTMs and BiLSTMs. The "Text-to-Text Transfer Transformer" (T5) model is an encoder-decoder transformer that has been pre-trained on a mixture of supervised and unsupervised tasks (Raffel et al., 2020). T5 formulates all tasks as text-to-text generation tasks that take in the input as a prefix and autoregressively samples the output to generate a prediction. The model has achieved state of the art results on a broad range of NLP tasks, including lexical semantics, and entailment prediction.

We employed the T5 and BERT-based models for dialogue act recognition because of their advanced capabilities in understanding and generating nuanced text. T5's encoder-decoder architecture and text-to-text framework allow it to capture comprehensive representations of dialogue, making it well-suited for this task. BERT's bidirectional encoder efficiently captures contextualized representations, enhancing its ability to understand complex interactions within group chats. These transformer-based models outperform traditional machine learning techniques by leveraging deep contextual and semantic information, reducing the need for manual feature engineering and providing a more robust understanding of student dialogues.

T5

For the T5 models we experimented with various window sizes ($k = [0, 3, 5]$) of included utterances to consider when recognizing the current dialogue act which we describe further below. Additionally, we include three new learnable tokens "$< cont >, < acont >, < usr >$" to delineate contextual, additional context, and user information, respectively. The input to the model contains whether it is a student or facilitator utterance and the utterance itself as in the following example:

```
<usr> Student
<cont> "Hello:)"
<acont>
```

When the context window is greater than zero the previous k utterances are concatenated to the input with each utterance following the above format. The final input to the model utilizing these context windows employs the following format, "$< usr >$ user type $< cont >$ utterance $< acont >$ **[previous k utterances]**" as shown in the following:

```
<usr> Student
<cont> "Are you guys there yet?"
<acont>
 <usr> Student <cont> "what is cyanobacteria",
 <usr> Facilitator <cont> "You're group is at the whiteboard",
 <usr> Student <cont> "heloooo"
```

The input to these models utilizes the standard tokenization and preprocessing procedures for the T5 model. Additionally, the labels are encoded using the standard T5 tokenizer in contrast to being label encoded like the BERT based methods. This is part of the "Text-to-Text" aspect of the model whereby the model will attempt to autoregressively generate a label instead of utilizing SoftMax predictions.

T5 Enhancement

To improve upon the T5 model's performance for dialogue act recognition we propose a method (Fig. 2) that follows the work of Chen et al. (2022) and Ni et al. (2021). Specifically, we adapt the label aware data augmentation and dual contrastive learning approach to enhance the model's representation of chat messages. Chen et al. (2022) propose a contrastive learning approach that takes a dataset X and learns "dual" representations of the discriminative features for the classification task z and the label aware input representation θ in a shared representation space with the goal to minimize the SoftMax transform of the normalized features with their respective label aware representations.

We chose dual contrastive learning with label-aware data augmentation to enhance the T5 model's performance in dialogue act recognition as this approach improves representation learning by combining supervised and unsupervised training, leading to better class separability and generalization. By leveraging the T5 architecture, it effectively uses contextual information and label-aware augmentation to enrich training

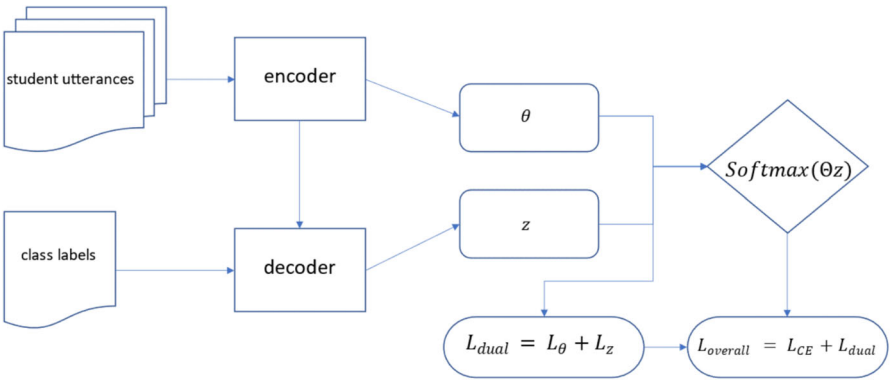

Fig. 2 Dialogue act recognition architecture

data without generating additional samples. These techniques, showing prior success, optimize the model's ability to capture nuanced dialogue acts accurately.

Dual Contrastive Fine-Tuning

Formally, dual contrastive loss is formulated as:

$$L_{Dual} = L_z + L_\theta \tag{1}$$

The full loss function utilizes a weighted combination of the cross entropy and the dual contrastive loss:

$$L_{Overall} = L_{CE} + \lambda\, L_{Dual} \tag{2}$$

Where L_{CE} represents the cross-entropy loss and λ is a hyperparameter controlling the effect of the dual contrastive term. Predictions are made using $\hat{y} = argmax\,(\theta z)$. Label aware data augmentation is achieved without creating any additional data points by getting unique views of the input sample as it relates to its corresponding labels.

Given T5's encoder-decoder model, unlike the encoder-only BERT models, we extended the label-aware data augmentation procedure to leverage the T5 architecture more effectively. In Fig. 2, all inputs share the same encoder-decoder structure. The current input is compared against positive samples, serving as an anchor to pull the underlying embeddings closer together. Simultaneously, negative samples, represented by all other input-label pairs, are repelled from the anchor to achieve higher class separability in their representations.

Dual contrastive fine-tuning differs from other contrastive learning methods by combining instance-level and label-aware objectives rather than solely focusing on the instance-level discrimination. By utilizing cross-entropy in conjunction with contrastive loss can make use of both supervised and unsupervised model training. Finally, traditional contrastive learning approaches utilize general augmentations (synonym replacement, back-translation, etc.) while dual contrastive learning relies primarily on label-aware augmentation.

Label-Aware Data Augmentation

Label-aware data augmentation seeks to generate unique views of the input data and their corresponding labels such that the underlying representations are brought closer together and opposing labels are pushed farther apart in a shared representation space. In our proposed approach, we obtain the $k + 1$ views proposed by Chen et al. (2022) of the input sample with the kth label by utilizing the k label names as input to the T5 decoder architecture that will be processed with the output of the encoder's last hidden state. The k label names represent the label aware information that is processed with each input sample. In conjunction input utterances along with its contextual information is processed by the encoder to produce an embedding representation for all the tokens in the input. We follow the work of Ni et al. (2021) to generate a total sentence representation by taking the average of the encoder's final hidden state. The representation from the decoder outputs, the encoded list of label names,

and the encoder outputs, the final sequence representation of the input, are utilized to compute the dual contrastive loss. Formally, given an input utterance and its contextual information $x \in x_0, \ldots, x_n$ we obtain the feature representation z with the following equation:

$$z = Average\left(Softmax\left(\frac{QK^T + \Omega}{\sqrt{d_k}}\right)V\right) \tag{3}$$

With this formulation we can obtain the label aware data augmentation without the addition of any additional samples.

Model Evaluation

To evaluate the predictive performance of the baseline methods and language models on dialogue act recognition, we performed group-level leave-one-out cross-validation (LOOCV). In each cross-validation fold, sixteen groups were used for training, while one group was reserved for validation and another for testing. For each fold, the validation group alternated to ensure that the chosen model parameters generalize well.

During training, the models were fine-tuned using specific optimization strategies. The BERT-based models utilized the AdamW optimizer with a learning rate of 3e-5. Both the LSTM and Bidirectional LSTM models included one recurrent layer with 512 units, followed by dropout at a rate of 0.3, and a fully connected classification layer. Meanwhile, the T5 models were optimized using the AdaFactor optimizer with a learning rate of 1e-3 based on empirical results. Additionally, the hyperparameter λ, responsible for controlling the influence of the dual contrastive term, was determined through a grid search 0.01, 0.05, 0.1 whereby we find that a value of 0.05 results in the best performance. Similarly, τ, representing the temperature parameter for the dual contrastive loss, is set to 0.1 following the work of Chen et al. (2022).

The training process was managed with PyTorch Lightning's Trainer, which was configured to run on GPUs with early stopping and model checkpointing to avoid overfitting. Early stopping was set with a patience of 2 epochs, and the maximum number of epochs was capped at 10. After training, the best model checkpoint was selected based on validation performance, and the model was subsequently tested on the reserved test group to assess its predictive performance. This evaluation ensured that the models were effectively trained and tested across different data splits, providing a good measure of performance for dialogue act recognition.

We utilized key performance metrics such as precision, recall, accuracy, and F1-score to understand the model's performance and to discern which types of dialogue scenarios present challenges for the model. We employ Cohen's kappa score as an additional metric to gauge the model's agreement with human annotations. This metric offers valuable insights into the model's ability to align with human judgments, shedding light on its capacity for meaningful analysis. Finally, as part of the validation process, we also compare our results to majority class baselines and classical machine learning methods for our classification scenario.

Table 3 Performance of the language models on topic related annotations, with majority class baseline as reference

Language models topic					
Model	Precision	Recall	F1-Score	Accuracy	Kappa
Majority classifier	NA	NA	NA	0.45	NA
BERT	0.70	0.68	0.68	0.68	0.55
BERT LSTM	0.69	0.68	0.68	0.68	0.54
BERT BiLSTM	0.70	0.69	0.69	0.69	0.55
T5	0.71	0.70	0.69	0.70	0.55
T5 W3[1]	0.74	0.72	0.72	0.72	0.59
T5 W5	0.72	0.69	0.69	0.69	0.55
T5 W3 DC[2]	0.79	0.74	0.74	0.74	0.63
T5 W3 DC low temp	**0.80**	**0.76**	**0.76**	**0.76**	**0.65**

Results

The findings presented in Tables 3, 4, 5 and 6 provide an overview of the models' performance, encompassing their average performance metrics across all cross-validation folds. Tables 3 & 4 shows each of the model's predictive performance on the topic related labels while Tables 5 & 6 displays the results for the epistemic stance labels.

Table 3 showcases the performance of various language models on topic-related annotations. The T5 model variants generally outperform the BERT-based models (Table 3) and traditional machine learning baselines (Table 4). Notably, the T5 W3 DC low temp model achieves the highest performance across all metrics, with a precision of 0.80, recall of 0.76, F1-score of 0.76, accuracy of 0.76, and a kappa of 0.65. This indicates the effectiveness of dual contrastive fine-tuning with label-aware data augmentation. In contrast, the highest performing BERT model only achieved an accuracy of 0.69.

The traditional machine learning models, as shown in Table 4, perform generally worse than the language models. The logistic regression model achieves an F1-score of 0.67, accuracy of 0.68, and kappa of 0.51. The random forest model lags behind with an F1-score of 0.40, accuracy of 0.50, and kappa of 0.16.

Table 5 provides the performance metrics for the language models on the epistemic labels. Similar to the topic-related annotations, the T5 model variants show superior performance. The T5 W3 DC model achieves the best balance of metrics with a

Table 4 Performance of the baseline models on the topic related annotations

Baseline models topic					
Model	Precision	Recall	F1-Score	Accuracy	Kappa
Logistic regression	0.69	0.68	0.67	0.68	0.51
Random forest	0.50	0.50	0.40	0.50	0.16

Table 5 Performance of the language models on the epistemic labels

Language models epistemic					
Model	Precision	Recall	F1-Score	Accuracy	Kappa
Majority classifier	NA	NA	NA	0.29	NA
BERT	0.63	0.60	0.60	0.60	0.49
BERT LSTM	0.65	0.61	0.61	0.61	0.54
BERT BiLSTM	0.63	0.60	0.60	0.60	0.49
T5	0.66	0.60	0.60	0.60	0.50
T5 W3[1]	0.65	0.60	0.60	0.60	0.50
T5 W5	0.67	0.62	0.62	0.62	0.52
T5 W3 DC[2]	**0.65**	**0.63**	**0.62**	**0.63**	**0.52**
T5 W5 DC	0.66	0.62	0.61	0.62	0.51

[1] W3 and W5 represent context windows of length 3 and 5 respectively
[2] DC represents dual contrastive fine-tuning with label-aware data augmentation

precision of 0.65, recall of 0.63, F1-score of 0.62, accuracy of 0.63, and a kappa of 0.52. This is an improvement over the BERT models where the highest accuracy is 0.61.

Traditional machine learning models, detailed in Table 6, under perform compared to the T5 variants. Logistic regression achieves a precision of 0.58, recall of 0.57, F1-score of 0.56, accuracy of 0.57, and kappa of 0.45. The random forest model again performs less effectively with a precision of 0.61, recall of 0.50, F1-score of 0.46, accuracy of 0.50, and kappa of 0.33.

It should be noted that the T5 model with dual contrastive fine-tuning, data aware label augmentation, and contextual information consistently outperform all baseline measures. This not only underscores the robustness of our proposed enhancement but also emphasizes the significance of the improvements it brings to the predictive performance of the T5 model. These results collectively contribute to a more thorough understanding of the model's ability to discern epistemic and topic related dialogue acts and its effectiveness in handling diverse dialogue scenarios.

Discussion

We discuss how these results explicitly answer our three research questions and provide a more detailed assessment of the model's performance in recognizing students dialogue acts in the context of multi-party collaborative discourse.

Table 6 Performance of the baseline models on the epistemic annotations

Baseline models epistemic					
Model	Precision	Recall	F1-Score	Accuracy	Kappa
Logistic regression	0.58	0.57	0.56	0.57	0.45
Random forest	0.61	0.50	0.46	0.50	0.33

Efficacy of LLMs for Dialogue Recognition

RQ1: *How effectively can pre-trained large language models discern student epistemic and topic-related contributions within the context of multi-party collaborative discourse, and what are the implications for AI-enabled learning environments?*

In addressing the first research question (**RQ1**), our analysis demonstrates the efficacy of pre-trained large language models in discerning student epistemic and topic-related contributions within multi-party collaborative discourse. This finding has significant implications for AI-enabled learning environments, particularly in enhancing the analysis of student engagement and knowledge-sharing.

Our observations reveal that T5 variants, even without dual contrastive fine-tuning, show commendable predictive accuracy in identifying student dialogue acts. However, these models often exhibit higher discordance with human annotators' judgments, raising concerns about the reliability of their assessments. Implementing dual contrastive fine-tuning with context and label-aware data augmentation markedly improves the T5 models' performance, achieving an accuracy rate of 0.76 and a Cohen's kappa score of 0.65. This improvement surpasses the critical kappa threshold of 0.6, underscoring the effectiveness of this approach in aligning model predictions with human annotations and accurately identifying students' epistemic and topic related contributions.

A detailed examination of model performance indicates substantial disagreement between "Socio-emotional" and "Other" labels, attributed to the higher ambiguity of utterances under these categories compared to "Content" or "Task." Similarly, the epistemic labels "Social Organization" and "Other" cause confusion, impacting overall performance. The observed disagreements suggest that while large language models have advanced capabilities, there remain challenges in accurately distinguishing between nuanced dialogue acts. This indicates a need for further refinement and possibly more granular labeling schemes to improve clarity and reduce ambiguity. Additionally, the improvement over baseline methods suggests that our approach can offer more reliable automated annotation, which could reduce the burden on human annotators and allow for more scalable analysis of collaborative learning environments. However, the persistent confusion between certain labels underscores the importance of ongoing model development and the incorporation of more sophisticated context-aware techniques to better capture the complexities of student dialogue in collaborative learning contexts.

Overall, our results suggest that T5 models, especially when enhanced through dual contrastive fine-tuning and data augmentation, excel in predicting students' epistemic and topic-related contributions. This capability is crucial for educational settings, as accurate recognition of student engagement and knowledge-sharing is essential for fostering effective collaborative learning experiences. Additionally, this suggests that T5 models may outperform baseline and BERT-based methods, providing a valuable tool for researchers and educators to reduce the burden of manual annotation.

Comparative Analysis of Language Models

RQ2: *What qualitative and quantitative differences emerge when using traditional machine learning and BERT-based approaches versus the T5 model for recognizing students' dialogue acts in collaborative discourse?*

For the second research question (RQ2), we compare traditional BERT-based approaches to our fine-tuned T5 model in recognizing students' dialogue acts in collaborative discourse. While BERT-based methods perform similarly to logistic regression, achieving accuracy rates between 68% and 69%, they only marginally outperform simpler methods, highlighting the limitations of these models without extensive customization. Our novel approach, which incorporates label-aware data augmentation, context awareness, and dual contrastive fine-tuning with the T5 model, achieves more substantial improvements, particularly in handling context-dependent dialogue acts.

Quantitatively, fine-tuned T5 models show notable gains in Cohen's kappa scores, with a 12% improvement for topic labels and 5% for epistemic labels over the baseline, reinforcing the superiority of large language models. Notably, T5 outperforms BERT-based models despite having fewer parameters (77M vs. 110M), a result we attribute to its generative architecture and the use of diverse pre-training tasks. Unlike BERT's masked language modeling objective, T5's text-to-text framework allows for more effective context generation, capturing subtleties in student discourse that are critical for recognizing epistemic stance.

Our approach demonstrates that T5, enhanced by label-aware data augmentation, context awareness, and dual contrastive fine-tuning, excels at identifying the qualitative nuances in educational dialogue, which are often missed by traditional models. These findings emphasize the importance of adopting fine-tuning strategies tailored to the task, illustrating how our contributions significantly enhance the model's ability to interpret and respond to complex student interactions. This not only improves automatic dialogue act recognition but also has broader implications for supporting collaborative learning environments through more accurate discourse analysis.

Enhancing the Performance of T5

RQ3: *In what ways can the performance of T5 models be enhanced for student epistemic and topic related dialogue act recognition in collaborative discourse through methods such as dual contrastive fine-tuning, label-aware data augmentation, and incorporation of contextual information?*

Addressing the third research question (RQ3), we explore strategies to enhance the performance of T5 models for student epistemic dialogue act recognition in collaborative discourse, with a focus on providing a viable option for automatic annotation. Specifically, we investigate the contributions of dual contrastive fine-tuning, label-aware data augmentation, and the incorporation of contextual information toward improving model accuracy and effectiveness.

Our findings, as shown in Tables 3, 4, 5 and 6, reveal a compelling trend: dual contrastive fine-tuning methods consistently outperform baseline results. The most promising outcomes are observed with the proposed method utilizing a context window of 3 (T5 W3 DC), achieving improved accuracy rates of 76% for topic labels and 63% for epistemic labels. This represents a substantial performance increase of 8% for topic recognition and 4% for epistemic recognition.

Moreover, dual contrastive fine-tuning has only a minor impact on training and inference times, making it a practical strategy for automatic annotation. This approach

enhances the predictive performance of large language models in recognizing student epistemic dialogue acts without significant computational overhead, making it feasible for real-world educational applications.

Our research highlights dual contrastive fine-tuning, label-aware data augmentation, and contextual information as highly effective techniques for improving the accuracy and effectiveness of T5 models in automatic annotation of students' epistemic dialogue acts during collaborative discourse. These strategies present viable options for enhancing educational assessments and facilitating more effective collaborative learning experiences, providing deeper insights into the underlying processes of student interaction and learning.

Performance

Answering (**RQ1**), we highlighted the effectiveness of pre-trained large language models, specifically T5 models with dual contrastive fine-tuning, in accurately discerning student epistemic responses and their topic-related contributions within collaborative discourse. This finding offers valuable insights for enhancing student engagement and knowledge-sharing in educational settings. Implementing dual contrastive fine-tuning with context and label-aware data augmentation significantly improves the T5 models' performance, surpassing baseline methods and achieving reasonable accuracy and Cohen's kappa scores. This underscores the potential of these methods to alleviate the burden of manual annotation, making large-scale analysis of collaborative learning environments more feasible.

Our results suggest that the T5 model, utilizing various context window lengths, dual contrastive loss, and label-aware data augmentation, shows substantial improvement over baseline machine learning methods (**RQ3**). Moreover, T5 represents a state-of-the-art methodology with high versatility, capable of being applied to various tasks. It has a fast inference time and can operate with fewer resources (60 million parameters) than other models of similar performance (e.g., GPT-3, BERT). This makes it particularly attractive to practitioners seeking to deploy high-performance dialogue act recognition models in problem-based collaborative learning environments (**RQ2**). Additionally, dual contrastive learning does not significantly increase training or inference time while significantly improving the model's predictive performance, further elucidating the benefits of our proposed approach (**RQ3**).

Our findings also indicate that increasing the context window of previous utterances reaches a point of diminishing returns as the context window grows larger. This may be caused by multiple conversation threads occurring simultaneously, which could result in larger context sizes introducing noise. This observation suggests that while context is beneficial, there is an optimal window size that balances additional context with the introduction of irrelevant information.

Notably, the proposed modifications affect the models differently depending on the granularity of the labels. We observe a small (1% accuracy) improvement using our approach on the fine-grained epistemic labels, while showing a larger (4% accuracy, 6% kappa) increase for the coarse-grained topic-based labels. This discrepancy may be due to the increased variance in the model's confidence across the fine-grained labels.

Dual constrastive learning may struggle with inter-cluster similarity for a subset of the fine-grained classes with fewer samples, while achieving sufficient separability for classes with more samples due to the imbalance in the dataset. Additionally, the model may overfit a subset of classes where the contrastive learning approach creates high intra-cluster similarity between input samples.

We also observe that decreasing the temperature of the dual contrastive loss helps increase the model's performance on the topic labels. A lower temperature allows the dot product between the anchor and target to be larger, encouraging a larger margin for comparing similarities between samples and labels, and thus allowing for a larger number of samples to be clustered under well-represented class categories.

Despite seeing marked improvement for the T5 model on the topic-related labels, we see limited improvement on the epistemic labels. This may indicate that while the models perform well in differentiating students' topic-related contributions, they struggle to differentiate how knowledgeable a student appears to be. The nuanced differences between the epistemic annotations and the short, informal nature of student chat responses, which differ significantly from the training data of large language models, contribute to this challenge. Typically, utterances are only a few words long, providing very little context for the models during inference time. For example, it may be difficult for the model to distinguish between "reply"/"other" utterances due to needing higher amounts of context to differentiate between two semantically related sentences. Overall, more work needs to be done to adapt large language models to unique styles of informal language.

In summary, our study underscores the potential of T5 models, especially when enhanced through dual contrastive fine-tuning and data augmentation, in automating the annotation process and reducing the pressures of manual annotation. These capabilities are crucial for educational settings, as accurate recognition of student engagement and knowledge-sharing is essential for fostering effective collaborative learning experiences. Our approach offers a reliable and scalable solution for analyzing collaborative learning environments, providing valuable tools for researchers and educators.

Design Implications

The proposed approach utilizing T5 models enhanced with dual contrastive fine-tuning and label-aware data augmentation offers several practical applications that can benefit educational settings. One of the primary applications is in the automation of dialogue act recognition within collaborative learning platforms. By accurately identifying students' epistemic and topic-related contributions, the system can provide real-time feedback to both students and educators, helping to facilitate more interactive and responsive learning environments. This automation reduces the reliance on manual annotation, which is often labor-intensive and time-consuming, thereby enabling large-scale analysis of student interactions and improving the scalability of educational assessments.

Moreover, the improved accuracy and alignment with human annotations provided by our approach can enhance the effectiveness of adaptive learning technologies. These technologies can personalize learning experiences by adjusting content and pedagogical strategies based on the detected dialogue acts. For instance, recognizing when students are struggling with a concept (epistemic responses) versus when they are engaging in off-topic or socio-emotional interactions allows for targeted interventions that can keep students on track and improve learning outcomes.

Additionally, the insights gained from analyzing student discourse can inform the development of more sophisticated educational tools. For example, virtual teaching assistants can be designed to recognize and respond appropriately to different types of student contributions, fostering a more supportive and engaging online learning environment. These tools can also provide educators with detailed analytics on student engagement and participation, enabling more informed instructional decisions and identifying areas where students may need additional support.

In research contexts, the ability to accurately annotate and analyze large datasets of student interactions can advance our understanding of collaborative learning dynamics. This can lead to the development of new theories and practices in educational psychology and pedagogy. Researchers can leverage the automated annotations to conduct more comprehensive and nuanced studies of how students interact, collaborate, and learn in group settings.

In sum, our proposed approach to automated dialogue act recognition introduces the opportunity to enhance AI-enabled learning environments, improve learning outcomes, streamline research processes, and ultimately foster more effective and engaging learning experiences for students.

Limitations

There are a few limitations that should be considered in our research approach. Our models have only been tested on the EcoJourneys game-based learning environment containing activity-specific data and may not generalize well to other educational contexts or collaborative learning environments. Moreover, the nature of student interactions within game-based learning environments may differ drastically from the traditional classroom setting, further exacerbating generalization concerns. Model scalability is a potential barrier to the application of this approach in real-time while ensuring consistent performance and managing computational load. Additionally, the reliance on substantial computational resources may limit the feasibility of deploying these models in resource-constrained educational settings. The additional complexity introduced by using dual contrastive fine-tuning can make it difficult to update and maintain the model in the case that new data becomes available or the learning constructs are changed. Finally, the annotation process is susceptible to biases stemming from human annotators' subjective interpretations, which can influence the model's training and outcomes.

Conclusion & Future Work

We have introduced an approach to recognizing epistemic and topic related dialogue acts by identifying epistemic stances and topic-related contributions in multi-party dialogue of students interacting with a collaborative game-based learning environment. We additionally applied a dual contrastive learning approach with context awareness and label aware data augmentation to increase the fine-tuning performance of the T5 model. We compared our approach with baseline machine learning and BERT-based techniques utilizing two different annotation schemes representative of students' epistemic responses and the quality of their collaborative behaviors. The results suggest that utilizing the T5 model and our proposed approach can perform dialogue act recognition and achieve an improvement in accuracy and Cohen's kappa score over baseline methods. Moreover, we showed that transfer learning of T5 models to multi-party educational dialogue is possible.

There are many promising avenues for future work that can increase the model's predictive performance as well as extend its current capabilities. In the work presented in this article, we evaluated base BERT, T5, and traditional machine learning methods to assess the effectiveness of our proposed enhancements to the T5 architecture. However, future work should explore the performance of other commonly used models, such as TextCNN and RoBERTa, to further strengthen the generalizability of our findings. Additionally, comparing our approach against larger language models like GPT-based models (GPT-3.5, GPT-4, etc.) or LlaMA-based approaches (Llama 2, Llama 3, etc.) would provide valuable insights into how billion-parameter models perform in dialogue act recognition, particularly in low-resource educational environments.

To address the issue of generalization, future works should explore the models in different educational contexts beyond the EcoJourneys game-based learning environment. This involves testing the models in traditional classroom settings and other collaborative learning environments to evaluate their robustness and adaptability. Additionally, domain adaptation techniques and transfer learning could be employed to fine-tune models for new settings with minimal data.

To mitigate concerns about the nature of student interactions differing across environments, future research should investigate multi-modal learning approaches. This includes integrating data from various sources, such as video recordings, and audio transcripts alongside text data to provide a richer context for dialogue act recognition. This holistic approach may help in better understanding and capturing the nuances of student interactions across different learning settings.

Addressing the complexity of dual contrastive fine-tuning, future work could focus on developing more streamlined and maintainable training protocols. One approach could be the adoption of continuous learning frameworks that allow the model to update incrementally as new data becomes available, thus avoiding the need for complete retraining. Another approach could involve using meta-learning techniques to make the model more adaptable to changes in learning constructs with minimal adjustments.

To overcome the bias in the annotation process, incorporating techniques such as active learning can be useful. This involves iterative training where the model identifies and queries the most informative data points for human annotation, thereby reducing annotation workload and potentially decreasing bias.

Exploring ensemble or hierarchical approaches may improve model performance. For example, using separate models to predict different classes or formulating an ensemble of tertiary predictors to handle class collapses can help address confusion between similar classes. This method allows for more specialized models to handle specific subsets of the data, improving overall accuracy.

In addition to utilizing students' chat information, future work should consider integrating trace log information from students' interactions with the EcoJourneys environment. These logs provide insights into collaborative and individual actions taken by students, which can be invaluable for predicting epistemic dialogue acts. Specifically, aligning gameplay events with dialogue instances can help in identifying patterns and contexts that are indicative of certain epistemic stances.

Finally, future work should investigate how these interrelationships influence collaborative performance. This could involve network analysis to understand social dynamics within groups and develop predictive models that link individual dialogue acts to group outcomes. Understanding these processes can aid the development of interventions and support systems to support collaborative learning experiences.

Appendix A Descriptive Statistics

Table 7 Descriptive statistics for topic based annotations

Frequency	Percent	Valid	Percent	Cumulative percent
Topic				
Content	2396	27.4	27.4	27.4
Other	1276	14.6	14.6	42.0
Socio-emotional	1368	15.7	15.7	57.7
Task	3701	42.3	42.3	100.0
Total	8741	100.0	100.0	
Epistemic				
K-	1503	17.2	17.2	17.2
K+	1180	13.5	13.5	30.7
Other	1368	21.4	21.4	52.1
Reply-knowledge	1618	18.5	18.5	70.6
Reply-simple	2572	29.4	29.4	100.0
Total	8741	100.0	100.0	

Author Contributions All authors contributed to the study conception and design. Data analysis was performed by Halim Acosta. The first draft of the manuscript was written by Halim Acosta, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This research was supported by funding from the National Science Foundation (NSF) under Grants DRL-2112635, DRL-1561486, and DRL-1561655. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Availability of Data and Material Data and materials created for this research are available upon request. Please direct all inquiries to the corresponding author.

Code Availability Code created for this research is available upon request. Please direct all inquiries to the corresponding author.

Declarations

Conflicts of Interest/Competing Interests No potential conflicts of interest.

References

Baker, I. M., Andriessen, J., & Järvelä, S. (2013). Feeling and meaning in the social ecology of learning: Lessons from play and games. In: Affective Learning Together, pp. 79–102. Routledge.

Berkowitz, M.W., & Gibbs, J.C. (1983). Measuring the developmental features of moral discussion. Merrill-Palmer Quarterly (1982-), 399–410

Blache, P., Abderrahmane, M., Rauzy, S., Ochs, M., & Oufaida, H. (2020). Two-level classification for dialogue act recognition in task-oriented dialogues. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 4915–4925.

Cai, Y., Liu, H., Ou, Z., Huang, Y., & Feng, J. (2022). Advancing semi-supervised task-oriented dialog systems by jsa learning of discrete latent variable models. arXiv:2207.12235.

Chen, Z., Yang, R., Zhao, Z., Cai, D., & He, X. (2018a). Dialogue act recognition via crf-attentive structured network. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 225–234.

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018b). The role of collaboration, computer use, learning environments, and supporting strategies in cscl: A meta-analysis. *Review of Educational Research., 88*(6), 799–843.

Chen, Q., Zhang, R., Zheng, Y., & Mao, Y. (2022). Dual contrastive learning: Text classification via label-aware data augmentation. arXiv:2201.08702.

Chi, M. T., & Wylie, R. (2014). The icap framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist., 49*(4), 219–243.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education., 59*(2), 661–686.

Dascalu, M.-D., Ruseti, S., Dascalu, M., McNamara, D. S., & Trausan-Matu, S. (2022). Dialogism meets language models for evaluating involvement in cscl conversations. In: Ludic, Co-design and Tools

Supporting Smart Learning Ecosystems and Smart Education: Proceedings of the 6th International Conference on Smart Learning Ecosystems and Regional Development, pp. 67–78 . Springer
Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.
Dielmann, A., & Renals, S. (2008). Recognition of dialogue acts in multiparty meetings using a switching dbn. *IEEE Transactions on Audio, Speech, and Language processing., 16*(7), 1303–1314.
Dillenbourg, P., & Hong, F. (2008). The mechanics of cscl macro scripts. *International Journal of Computer-Supported Collaborative Learning., 3*, 5–23.
Greene, J. A., & Azevedo, R. (2007). A theoretical review of winne and hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research., 77*(3), 334–372.
Hainey, T., Connolly, T. M., Stansfield, M., & Boyle, E. A. (2011). Evaluation of a game to teach requirements collection and analysis in software engineering at tertiary education level. *Computers & Education., 56*(1), 21–35.
Heritage, J. (2012a). Epistemics in action: Action formation and territories of knowledge. *Research on Language & Social Interaction., 45*(1), 1–29.
Heritage, J. (2012b). The epistemic engine: Sequence organization and territories of knowledge. *Research on Language & Social Interaction., 45*(1), 30–52.
Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review., 16*, 235–266.
Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research., 67*(1), 88–140. https://doi.org/10.3102/00346543067001088
Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A meta-analysis of cscl in stem education during 2005–2014. *Educational Research Review., 28*, 100284.
Jesus, Â.M., & Silveira, I.F. (2019). A collaborative game-based learning framework to improve computational thinking skills. In: 2019 International Conference on Virtual Reality and Visualization (ICVRV), pp. 161–166. IEEE
Kääntä, L. (2014). From noticing to initiating correction: Students' epistemic displays in instructional interaction. *Journal of Pragmatics., 66*, 86–105.
Katuka, G.A., Bex, R.T., Celepkolu, M., Boyer, K.E., Wiebe, E., Mott, B., & Lester, J. (2021). My partner was a good partner: Investigating the relationship between dialogue acts and satisfaction among middle school computer science learners. In: Proceedings of the 14th International Conference on Computer-supported Collaborative Learning-cscl 2021. International Society of the Learning Sciences
Kumaran, V., Rowe, J., Mott, B., Chaturvedi, S., & Lester, J. (2023). Improving classroom dialogue act recognition from limited labeled data with self-supervised contrastive learning classifiers. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (eds.) Findings of the Association for Computational Linguistics: ACL 2023, pp. 10978–10992. Association for Computational Linguistics, Toronto, Canada. https://doi.org/10.18653/v1/2023.findings-acl.698 . https://aclanthology.org/2023.findings-acl.698
Li, S., & Chen, X. (2023). Multiple information-aware recurrent reasoning network for joint dialogue act recognition and sentiment classification. *Information., 14*(11), 593. https://doi.org/10.3390/info14110593
Liu, B., & Lane, I. (2016). Attention-based recurrent neural network models for joint intent detection and slot filling. arXiv:1609.01454.
McLeod, S., Kruijff-Korbayova, I., & Kiefer, B. (2019). Multi-task learning of system dialogue act selection for supervised pretraining of goal-oriented dialogue policies. In: Proceedings of the 20th Annual SIGDIAL Meeting on Discourse and Dialogue, pp. 411–417.
Mezza, S., Wobcke, W., & Blair, A. (2022). A multi-dimensional, cross-domain and hierarchy-aware neural architecture for iso-standard dialogue act tagging. In: Proceedings of the 29th International Conference on Computational Linguistics, pp. 542–552.
Muis, K. R., & Franco, G. M. (2009). Epistemic beliefs: Setting the standards for self-regulated learning. *Contemporary Educational Psychology., 34*(4), 306–318.
Ni, J., Ábrego, G.H., Constant, N., Ma, J., Hall, K.B., Cer, D., & Yang, Y. (2021). Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. arXiv:2108.08877.
O'Connor, C., & Michaels, S. (2019). Supporting teachers in taking up productive talk moves: The long road to professional learning at scale. *International Journal of Educational Research., 97*, 166–175.
Pengfei, G., & Yinglong, M. (2022). A universality-individuality integration model for dialog act classification. arXiv:2204.06185.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research., 21*(1), 5485–5551.

Rienties, B., & Alden, B. (2014). *Emotions used in learning analytics: A state-of-the-art review* (p. 2). Measuring and Understanding Learner Emotions: Evidence and Prospects.

Saleh, A., Feng, C., Bae, H., Hmelo-Silver, C. E., Glazewski, K.D., Lee, S., Mott, B., & Lester, J. (2021). Negotiating accountability and epistemic stances in Middle-School collaborative discourse. In: Hmelo-Silver, C.E., De Wever, B., Oshima, J. (eds.) Proceedings of the 14th International Conference on Computer-Supported Collaborative Learning - CSCL 2021, Bochum, Germany, pp. 197–200.

Saleh, A., Hmelo-Silver, C. E., Glazewski, K. D., Mott, B., Chen, Y., Rowe, J. P., & Lester, J. C. (2019). Collaborative inquiry play: A design case to frame integration of collaborative problem solving with story-centric games. *Information and Learning Sciences., 120*(9/10), 547–566.

Scardamalia, M., & Bereiter, C. (2006). *Fcl and knowledge building: A continuing dialogue*. Institute for Knowledge Innovation and Technology: University of Toronto.

Solem, M. S. (2016). Displaying knowledge through interrogatives in student-initiated sequences. *Classroom Discourse., 7*(1), 18–35.

Stahl, G. (2006). Group cognition: Computer support for building collaborative knowledge (acting with Technology), pp. 431–469. The MIT Press, Cambridge.

Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema, C. V., & Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics., 26*(3), 339–373.

Sung, H.-Y., & Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & education., 63*, 43–51.

Suthers, D. D., & Hundhausen, C. D. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *The Journal of the Learning Sciences., 12*(2), 183–218.

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education., 89*, 53–74.

Tan, W., Lin, J., Lang, D., Chen, G., Gašević, D., Du, L., & Buntine, W. (2023). Does informativeness matter? active learning for educational dialogue act classification. In: International Conference on Artificial Intelligence in Education, pp. 176–188. Springer

Vielsted, M., Wallenius, N., & Goot, R. (2022). Increasing robustness for cross-domain dialogue act classification on social media data. In: Proceedings of the Eighth Workshop on Noisy User-Generated Text (W-NUT 2022), pp. 180–193.

Wang, D., Li, Z., Zheng, H., & Shen, Y. (2020). Integrating user history into heterogeneous graphs for dialogue act recognition. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 4211–4221.

Wang, A.-H., Song, L., Jiang, H., Lai, S., Yao, J., Zhang, M., & Su, J. (2021). A structure self-aware model for discourse parsing on multi-party dialogues. In: International Joint Conference on Artificial Intelligence. https://api.semanticscholar.org/CorpusID:237100544

Wei, K., Knox, D., Radfar, M., Tran, T., Muller, M., Strimel, G.P., Susanj, N., Mouchtaris, A., & Omologo, M. (2022). A neural prosody encoder for end-to-end dialogue act classification. In: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Singapore, Singapore.

Yu, N., Fu, G., & Zhang, M. (2022). Speaker-aware discourse parsing on multi-party dialogues. In: Proceedings of the 29th International Conference on Computational Linguistics, pp. 5372–5382.

Authors and Affiliations

Halim Acosta[1] · Seung Lee[1] · Haesol Bae[3] · Chen Feng[2] · Jonathan Rowe[1] · Krista Glazewski[4] · Cindy Hmelo-Silver[2] · Bradford Mott[1] · James C. Lester[1]

✉ Halim Acosta
hacosta@ncsu.edu

Seung Lee
sylee@ncsu.edu

Haesol Bae
hbae4@albany.edu

Chen Feng
carrfeng@iu.edu

Jonathan Rowe
jprowe@ncsu.edu

Krista Glazewski
kdglazew@ncsu.edu

Cindy Hmelo-Silver
chmelosi@indiana.edu

Bradford Mott
bwmott@ncsu.edu

James C. Lester
lester@ncsu.edu

1 Center for Educational Informatics, North Carolina State University, Raleigh, NC 27606, USA

2 Center for Research on Learning and Technology, Indiana University, Bloomington, IN 47405, USA

3 Department of Educational Theory and Practice, University at Albany, Albany, NY 12222, USA

4 Friday Institute for Educational Innovation, North Carolina State University, Raleigh, NC 27606, USA