

## To speak or not to speak, and what to speak, when doing task actions collaboratively

Jauwairia Nasir, Aditi Kothiyal, Haoyu Sheng, Pierre Dillenbourg

### Angaben zur Veröffentlichung / Publication details:

Nasir, Jauwairia, Aditi Kothiyal, Haoyu Sheng, and Pierre Dillenbourg. 2023. "To speak or not to speak, and what to speak, when doing task actions collaboratively." In *Proceedings of the 16th International Conference on Educational Data Mining, 11-14 July 2023, Bengaluru, India*, edited by Mingyu Feng, Tanja Käser, and Partha Talukdar, 194–206. Massachusetts: International Educational Data Mining Society. <https://doi.org/10.5281/zenodo.8115689>.

### Nutzungsbedingungen / Terms of use:

CC BY-NC-ND 4.0



# To speak or not to speak, and what to speak, when doing task actions collaboratively\*

Jauwairia Nasir<sup>†</sup>  
University of Augsburg  
jauwairia.nasir@uni-a.de

Aditi Kothiyal<sup>†</sup>  
Indian Institute of Technology  
Gandhinagar  
aditi.kothiyal@iitgn.ac.in

Haoyu Sheng  
École polytechnique fédérale  
de Lausanne  
haoyu.sheng@epfl.ch

Pierre Dillenbourg  
École polytechnique fédérale  
de Lausanne  
pierre.dillenbourg@epfl.ch

## ABSTRACT

Transactive discussion during collaborative learning is crucial for building on each other’s reasoning and developing problem solving strategies. In a tabletop collaborative learning activity, student actions on the interface can drive their thinking and be used to ground discussions, thus affecting their problem-solving performance and learning. However, it is not clear how the interplay of actions and discussions, for instance, how students performing actions or pausing actions while discussing, is related to their learning. In this paper, we seek to understand how the transactivity of actions and discussions is associated with learning. Specifically, we ask what is the relationship between discussion and actions, and how it is different between those who learn (gainers) and those who do not (non-gainers). We present a combined differential sequence mining and content analysis approach to examine this relationship, which we applied on the data from 32 teams collaborating on a problem designed to help them learn concepts of minimum spanning trees. We found that discussion and action occur concurrently more frequently among gainers than non-gainers. Further we find that gainers tend to do more reflective actions along with discussion, such as looking at their previous solutions, than non-gainers. Finally, gainers discussion consists more of goal clarification, reflection on past solutions and agreement on future actions than non-gainers, who do not share their ideas and cannot agree on next steps. Thus this approach helps us identify how the interplay of actions and discussion could lead to learning, and the findings offer guidelines to teachers

<sup>†</sup>Both authors contributed equally to this work and share first authorship. A majority of the work was conducted while the authors were employed at CHILI Lab, EPFL

\*(Does NOT produce the permission block, copyright information nor page numbering). For use with edm\_article.cls.

J. Nasir, A. Kothiyal, H. Sheng, and P. Dillenbourg. To speak or not to speak, and what to speak, when doing task actions collaboratively. In M. Feng, T. Käser, and P. Talukdar, editors, *Proceedings of the 16th International Conference on Educational Data Mining*, pages 194–206, Bengaluru, India, July 2023. International Educational Data Mining Society.

© 2023 Copyright is held by the author(s). This work is distributed under the Creative Commons Attribution NonCommercial NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.  
<https://doi.org/10.5281/zenodo.8115689>

and instructional designers regarding indicators of productive collaborative learning, and when and how, they should intervene to improve learning. Concretely, the results suggest that teachers should support elaborative, reflective and planning discussions along with reflective actions.

## Keywords

transactivity, sequence mining, content analysis, collaborative learning

## 1. INTRODUCTION

When students collaborate to learn from computer supported collaborative learning (CSCL) environments, their learning depends not only on the quality of their interaction with each other, but also with the learning activity [5]. In other words, students need to align both in terms of their activities (such as problem-solving steps or co-writing) and their discussion (of strategies and knowledge) [23]. Specifically, in CSCL environments that involve problem-solving to build conceptual understanding, for the activity to be effective for learning, team members need to develop a joint problem space and construct knowledge through the process of explanation, negotiation and mutual regulation [24, 5]. To achieve this, their actions must either move them towards the solution, or provide them some information that generates potential and motivates future problem-solving actions [25]. Actions can thus help ground collaboration [2], if they are followed with the right kind of discussion, i.e., students discussions should then build on and leverage this information or potential to further understand the problem, decide on next steps and construct meaning from the problem-solving experience [24, 6]. Thus, discussion and actions together play critical roles in problem-solving CSCL environments, as it is through both these means that students obtain and share task-related information to build a common ground, develop problem-solving or learning strategies and regulate their learning. For instance, as described in [4] children’s body movements and task-related speech evolve together and serve the purposes of communication and co-ordination, and as cognitive tools for knowledge construction. Similarly, research has also shown that acting together on task-related objects accompanied with speech was related to effective collaboration [15].

Within the collaborative learning research space, transactivity or student’s discussion that builds on each other’s reasoning by interpreting team member’s statements, asking questions, extending, critiquing and integrating has been used as a metric to evaluate the effectiveness of collaboration [29, 28]. We argue that in synchronous problem-solving CSCL environments, the notion of transactivity should be extended to discussion and actions together, i.e., actions that build on students’ discussion, and discussions that build on actions. For instance, students should explicate the information gained from an action such that team members can then discuss about what this suggests for the next problem-solving steps [6]. However, not all actions need be accompanied with discussions. For instance, students may plan and perform a set of actions, or they may perform and reflect on each action [18]. The key question then is, should teams discuss while performing actions, i.e., building on each other’s ideas ‘on the go’ or should they ‘stop and pause’ their actions to discuss their ideas, or both? Further, how are each of these behaviours related to learning? Finally, which kind of discussion accompanying actions is productive? The answers of these questions are necessary to support teachers in intervening at the right time to guide students actions and discussions, or in the design of feedback built into CSCL environments.

Previous research that analysed students’ discussion and actions together in an attempt to identify joint discussion and action indicators of collaboration followed one of two approaches. The first one was considering whether actions and discussion occur together, but not the nature of the actions and the discussion that occur together [15]. The other approach analysed the synchronicity of actions and the transactivity of the discussions separately [23]. In this work, we bring together these two approaches and propose a combined differential sequence mining and qualitative content analysis approach to examine the transactivity of discussion and actions. Specifically, we ask the following questions:

- RQ1: What is the relationship between the discussion and actions, and how is this relationship different between gainers and non-gainers?
- RQ2: What is the qualitative nature of verbal interactions that happen along with a specific action of interest?

We begin with the data of 32 teams working on a collaborative robot-mediated problem-solving activity where *actions* refer to any interaction with the activity interface and *discussions* refer to quantity and quality of communication between the two team members. To answer RQ1 we performed differential sequence mining on the combined speech and action sequence to identify the relationship between actions and speech, which actions are accompanied by speech, and how this differs between gainers and non-gainers. Next, to answer RQ2 we perform content analysis of the discussion occurring around one particular action of interest and examine the nature of discussion and how it varies between gainers and non-gainers. Our two part approach helps us illustrate the notion of action-discussion transactivity that is conducive to learning and we find that reflective actions

accompanied with elaboration, reflection, negotiation and planning regarding next steps, are related with learning. The main contributions of this work are the notion of action-discussion transactivity and a methodology to examine the productivity of collaborative learning with this lens.

## 2. RELATED WORK

Research on collaborative learning has shown the key role of verbal interaction in advancing thinking and learning [26, 3]. Groups that are successful in problem-solving usually discuss and accept the correct proposal and their discussions are more coherent [3]. Conversation is the process by which students build and maintain a joint problem space [24]. Transactive verbal interaction, which is characterized by partner’s building on each other’s reasoning, can improve learning as peers can generate more complex understanding of the problem quickly through such verbal interactions [26]. When they generate explanations during collaboration peers construct shared representations and this may be one of the mechanisms that results in knowledge co-construction [12]. Actions done within a CSCL environment can also create shared representations, which can be then referred to during the discussion and thus improve the quality of collaboration [6, 2].

In this direction, research identified productive action patterns during collaborative learning with an interactive tabletop by analysing action logs with and without verbal interaction [7, 15, 23, 8]. [8] found that the number of touches allowed (single or multiple) on the table did not affect the level or symmetry of physical or verbal participation, but the nature of the discussion, which was more task-focussed in the multi-touch condition. [15] found that while the level or symmetry of participation of each team member in terms of action and speech did not relate to collaboration quality, certain sequences of actions and speech were related to the quality of collaboration. Concretely, more collaborative groups have more patterns of verbal discussion accompanied with actions, less concurrency of actions and less parallel actions. On the other hand, less collaborative groups had actions with limited verbal interactions, high concurrency and parallelism. This suggests that students in less collaborative groups were not as aware of their peers actions and did not discuss about the actions. On the other hand, in a chat-based collaborative learning environment, researchers found that neither synchrony of students actions nor transactivity of students’ chats was related to performance on the task, but other factors such as group dynamics and prior knowledge had a more crucial role [23]. Thus, the role of symmetry, synchrony and transactivity of actions and discussion during collaborative learning appears to depend on the context.

In CSCL environments, several metrics of student dialogue (speech or chats) have been identified which are indicative of good collaboration. These include quantity (eg, number and length of utterances, and talk time) and heterogeneity and transactivity of verbal participation (eg, turn taking and building on each other’s reasoning), along with features of speech such as voice inflection [29, 27]. Going further, research employed a combination of audio and action features to measure the quality of collaboration and collaborative learning and found that classifiers using a combination of

audio and action features always perform better than those classifiers using audio or actions alone [27, 22]. This suggests that combining conversation and action metrics together can offer a better understanding of the quality of collaboration. In this work, we build on this line of research by specifically examining the role of action-discussion transactivity in collaborative learning, i.e., how actions and discussions can build on each other to lead to learning.

### 3. METHODS

#### 3.1 Learning Activity and Dataset

In order to understand action-discussion transactivity we propose an approach which combines differential sequence mining and qualitative content analysis, and choosing one type of action - reflective actions - as an example, we show how our analysis can identify what type of actions and discussion occurring together can lead to learning. We use the speech and log action data from a multimodal temporal dataset [17], and log actions and transcripts from its corresponding dialogue corpus [20] collected from a robot mediated collaborative learning activity called JUSThink [19]. In JUSThink, two children play as a team to solve a minimum spanning tree problem where the goal is to build railway tracks to connect gold mines on a fictional Swiss map with a minimum cost, as shown in Figure 1. The corpus comprises data from 64 children aged 9 to 12 years, grouped into 32 teams, from international schools in Switzerland. The children were familiar with collaborative activities and robots as part of school activities, but did not have prior experience with QTrobot. The study was not part of a regular classroom activity. Two different views are provided in this activity, namely figurative and abstract as shown in Figure 2, and each child in a team only has one view at a time. In the figurative view, one can add or remove tracks while in the abstract view, one can see the cost associated with building a track and review the team’s previous solutions. Thus at a time, one child can do solution building actions while the other can do reflective actions, so they have to discuss with each other to plan the next steps. The views of the team members are swapped every two edits. Hence, with these collaborative script choices such as partial information and role switching, only one member can perform an action at a time, therefore every action is a team action. Teams are allowed to submit solutions multiple times until the time limit runs out. They can also check descriptions of activity functionality and rules on the help page, which has been elaborated for them by the robot before the activity starts. More details of the activity can be found in [19].

#### 3.2 Feature Selection and Encoding

The original multimodal temporal dataset consists of 56 features including log features, audio features, video features etc. Our analysis only focuses on the log features and speech features. Therefore, we selected 5 features from the multimodal temporal dataset including T\_add, T\_remove, T\_hist, T\_speech, T\_overlap\_over\_speech. With 32 teams, we have a total of 4676 time windows in our analysis where each time window corresponds to 10 seconds of activity. For each time window, we have three descriptive features additionally, which are team number, time\_in\_secs, and window number. The log and speech features as well as the three descriptive features are shown in Table 1.

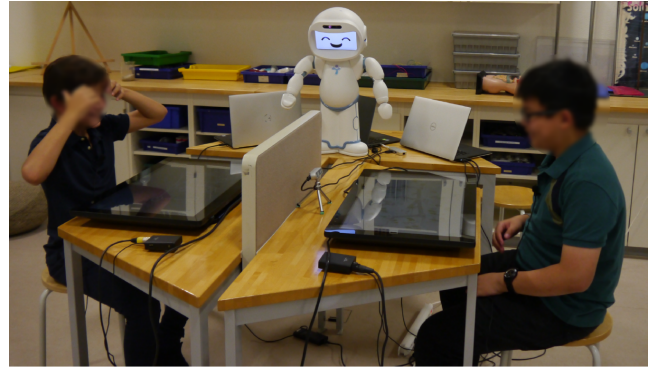
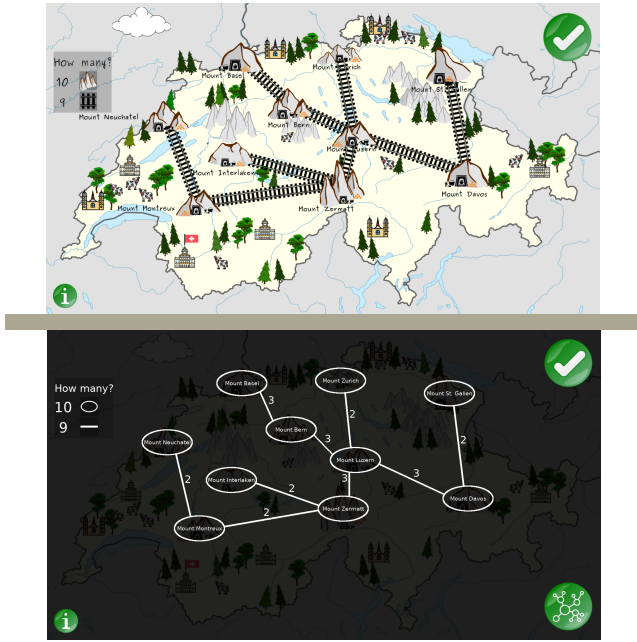


Figure 1: The experimental set up of JUSThink

Table 1: Multi-modal Features

| Feature                      | Meaning                                                                                                    |
|------------------------------|------------------------------------------------------------------------------------------------------------|
| <b>team</b>                  | The team to which the window belongs to                                                                    |
| <b>time_in_secs</b>          | Time in seconds until that window                                                                          |
| <b>window</b>                | The window number                                                                                          |
| <b>T_add</b>                 | The number of times a team added an edge on the map                                                        |
| <b>T_remove</b>              | The number of times a team removed an edge from the map                                                    |
| <b>T_hist</b>                | The number of times a team opened the sub-window with history of their previous solutions                  |
| <b>T_speech</b>              | The average of the two team member’s speech activity in that window/(until that window)                    |
| <b>T_overlap_over_speech</b> | The average percentage of time the speech of the team members overlaps in that window/(until that window). |

We begin by encoding the log and speech features in each 10s time window so that we can get a sequence representing the action+speech of each student in 10s increments. In the data set, the choice of 10 seconds as the unit of analysis is set considering the need to balance between too few and too many robot interventions. Before diving into the encoding details, we briefly elaborate on the terminology of gainers and non-gainers that we will be using from here onwards. In previous work based on this study [17], authors clustered the teams in two ways, once on the multimodal behaviors and once on task performance and learning gains (calculated as the normalized difference between pre and post test scores). Then, comparing the clusters using a similarity metric, they found higher learning gains associated with two sets of multimodal behaviors while lower learning gains were associated with another set of behaviors. They named the former set of 26 teams as gainers (those who gain knowledge) and the latter 6 teams as non-gainers (those who do not gain knowledge). For speech, it was found in [18] that speech behaviors are different for gainers and non-gainers, in terms of both quantity of speech and overlapping speech. So we define three levels - low, medium and high level of speech/speech overlap in each window on the basis of low and high thresholds of speech/speech overlap defined by considering the average of the 25th (for the low thresholds) and the 75th (for



**Figure 2: The interface for JUSThink where the two screens, separated by a barrier, show two different views: a figurative view (above) that allows for interactions such as additions, deletions of rail tracks, and an abstract view (below) that showcases the associated costs.**

the high thresholds) percentiles across the gainers and non-gainers participants. Then we encode  $T_{speech}$  or  $T_{overlap\_over\_speech}$  in each window by using the low and high thresholds as shown in Table 2.

**Table 2: Definition of Speech/Speech Overlap levels where  $x$  is the continuous value of Speech/Speech Overlap in a time window**

| speech/<br>speech overlap level | condition                                             |
|---------------------------------|-------------------------------------------------------|
| LS/LS <sub>o</sub>              | $x \leq \text{low threshold}$                         |
| MS/MS <sub>o</sub>              | $\text{low threshold} < x \leq \text{high threshold}$ |
| HS/HS <sub>o</sub>              | $x > \text{high threshold}$                           |

For action logs, we only consider add edges ( $T_{add}$ ), remove edges ( $T_{remove}$ ) and click solution history button ( $T_{hist}$ ) within each time window as these are the meaningful actions that have been found to contribute to learning in this context[18]. We identify which of the meaningful actions happen in a time window and if an action happened at least once in a time window, it is encoded as being present. Then we have eight action combinations because each of the three actions can be either present or absent. Finally there are 24 combinations of action + speech in each time window (combinations of three levels of speech and eight action combinations) and we encode those combinations to 24 numbers as shown in the following Table 3. Note that the same encoding process is also applied to combinations of speech overlap levels and meaningful actions.

After encoding features in each time window of each team, we obtained two datasets of encoded features - one is en-

coded combinations of speech levels and actions, another is encoded combination of speech overlap levels and actions. Each dataset contains 32 teams' sequences of activities in ten-second time windows and we further separate each dataset into gainer sequences and non-gainer sequences for analysis.

### 3.3 Differential Sequence Mining

To answer our RQ1 by differentiating action+ speech sequences between gainers and non-gainers, we applied differential sequence mining algorithm (DSM)[13]. DSM algorithm mainly uses the following two sequential pattern mining frequency measures.

1. *sequence support (s-support)*: For a set of sequences, the number of sequences in which the pattern occurs, regardless of how frequently it occurs within each sequence.
2. *instance support (i-support)*: For a given sequence, the number of times the pattern occurs, without overlap, in this sequence.

The algorithm firstly finds all patterns that meet the predefined s-support threshold. Then the algorithm selects only those patterns that have statistically significantly different i-support values between the two groups. Concretely, the algorithm filters frequent patterns based on the p-value of a t-test comparing the i-support values of patterns in each sequence, between the groups to find patterns whose p-value is less than 0.05. Finally, the algorithm compares the mean i-support value for each pattern between groups to identify the patterns that occur more often in one group than the other.

Before applying the DSM, we separated the two datasets we get after feature selection and encoding into four datasets. For each of the original two datasets as described in the previous sub-section, we divide the dataset (which contain 32 teams in total) into a sub-dataset that contained sequences of 26 gainer teams and another sub-dataset which contained sequences of 6 non-gainer teams.

Firstly, we set the minimum threshold of s-support to 0.6 and consider patterns that occur in at least 60% of sequences as s-frequent patterns within a group. We employ a simple sequential mining algorithm SPAMc [9] to find frequent patterns for both gainers and non-gainers with the LASAT tool [16]. Then we calculate the i-support of each frequent pattern in each team sequence in both gainers and non-gainers. For each frequent pattern, we generate a vector that contains i-support for each team sequence. Then we apply Welch's t-test with 0.05 p-value threshold to filter frequent patterns that are significantly different between gainers and non-gainers. After the filtering, we compare the mean i-support value for each frequent pattern between gainers and non-gainers so that we could compare patterns that occur more often in one group than the other. Finally, we get four categories of frequent patterns - two categories in which the patterns are s-frequent in only one group, and two categories in which the patterns are frequent in both groups but occurred more often in one group than the other.

**Table 3: Encoding at specific level**

| Speech Level Code  | Speech Overlap Level Code | Meaning                                                                                                                                   |
|--------------------|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| LS_Add             | LSo_Add                   | Low level of speech(S)/speech overlap (So), at least add one edge                                                                         |
| MS_Add             | MSo_Add                   | Medium level of speech(S)/speech overlap (So), at least add one edge                                                                      |
| HS_Add             | HSo_Add                   | High level of speech(S)/speech overlap (So), at least add one edge                                                                        |
| LS_Remove          | LSo_Remove                | Low level of speech(S)/speech overlap (So), at least remove one edge                                                                      |
| MS_Remove          | MSo_Remove                | Medium level of speech(S)/speech overlap (So), at least remove one edge                                                                   |
| HS_Remove          | HSo_Remove                | High level of speech(S)/speech overlap (So), at least remove one edge                                                                     |
| LS_Hist            | LSo_Hist                  | Low level of speech(S)/speech overlap (So) and at least click history button one time                                                     |
| MS_Hist            | MSo_Hist                  | Medium level of speech(S)/speech overlap (So) and at least click history button one time                                                  |
| HS_Hist            | HSo_Hist                  | High level of speech(S)/speech overlap (So) and at least click history button one time                                                    |
| LS_Add_Remove      | LSo_Add_Remove            | Low level of speech(S)/speech overlap (So), at least add one edge, at least remove one edge                                               |
| MS_Add_Remove      | MSo_Add_Remove            | Medium level of speech(S)/speech overlap (So), at least add one edge, at least remove one edge                                            |
| HS_Add_Remove      | HSo_Add_Remove            | High level of speech(S)/speech overlap (So), at least add one edge, at least remove one edge                                              |
| LS_Add_Remove_Hist | LSo_Add_Remove_Hist       | Low level of speech(S)/speech overlap (So), at least add one edge, at least remove one edge and at least click history button one time    |
| MS_Add_Remove_Hist | MSo_Add_Remove_Hist       | Medium level of speech(S)/speech overlap (So), at least add one edge, at least remove one edge and at least click history button one time |
| HS_Add_Remove_Hist | HSo_Add_Remove_Hist       | High level of speech(S)/speech overlap (So), at least add one edge, at least remove one edge and at least click history button one time   |
| LS_Remove_Hist     | LSo_Remove_Hist           | Low level of speech(S)/speech overlap (So), at least remove one edge and at least click history button one time                           |
| MS_Remove_Hist     | MSo_Remove_Hist           | Medium level of speech(S)/speech overlap (So), at least remove one edge and at least click history button one time                        |
| HS_Remove_Hist     | HSo_Remove_Hist           | High level of speech(S)/speech overlap (So), at least remove one edge and at least click history button one time                          |
| LS_Add_Hist        | LSo_Add_Hist              | Low level of speech(S)/speech overlap (So), at least add one edge and at least click history button one time                              |
| MS_Add_Hist        | MSo_Add_Hist              | Medium level of speech(S)/speech overlap (So), at least add one edge and at least click history button one time                           |
| HS_Add_Hist        | HSo_Add_Hist              | High level of speech(S)/speech overlap (So), at least add one edge and at least click history button one time                             |
| LS_NA              | LSo_NA                    | Low level of speech(S)/speech overlap (So) and no useful action happens                                                                   |
| MS_NA              | MSo_NA                    | Medium level of speech(S)/speech overlap (So) and no useful action happens                                                                |
| HS_NA              | HSo_NA                    | High level of speech(S)/speech overlap (So) and no useful action happens                                                                  |

### 3.4 Identifying relevant episodes of interest

To gain more insights into the transactivity of speech and log actions, we need to begin by identifying “patterns of interest” within the frequent patterns. From literature and previous research on the same dataset[18], we know that reflective speech and actions differentiate between gainers and non-gainers. Therefore, we are interested in reflection related actions (particularly clicking history button) and want to find what kind of verbal interaction happened along with it. So we consider frequent patterns which have speech along with open history action as episodes of interest for further qualitative analysis.

To find the exact content of dialogues in the relevant episodes, we matched the time window of the relevant episodes of each team to their transcript datasets in JUSThink Dialogue and Actions Corpus[20]. Due to the imprecision in matching time windows and the fact that it is difficult to extract

meaningful information from very short segments (less than 60 seconds) of the dialogue, we have also included the conversation within 20 seconds before and after the matching time window.

### 3.5 Content Analysis on Dialogues

To answer RQ2 and examine the qualitative nature of dialogue during the episodes of interest, we decided to perform content analysis [14] on the selected dialogues. Content analysis is a qualitative analysis approach to code a corpus of data according to certain existing categories with the goal of doing statistical analysis on the numbers and identifying certain trends or providing evidence for/against a hypothesis. To look deeper into reflection behaviours, we focus on the following three aspects of problem-solving discussions and code the dialogue for these aspects:

1. What do teams observe from past actions?
2. What decisions do they take about future actions on the basis of these observations?
3. Do they reach any agreement on the future actions, and if yes, how?

In order to code their negotiation and agreement on their future actions, we applied the “refine” strategy in the negotiation framework[1] to analyse all dialogues. The “refine” strategy means that an agent decides to make another offer that somehow “refines”, “builds” or “modifies” the original offer proposed by another agent. In this strategy, the initiating move includes an offer which is proposed by a speaker for agreement. The reactive moves include acceptance, ratification and rejection. Ratification refers to an acceptance which follows an acceptance by the other. Additionally, after an initial coding of the data, we defined additional categories in the negotiation framework, i.e., “goal clarification” and “sharing understanding” because they are relevant for this collaborative educational setting. The goal of this activity is to build tracks with the minimum cost – 22 francs. Team members must clarify this goal and share their understanding of the problem with each other due to the fact that they have two complementary views of the problem.

To illustrate the content analysis we conduct, we show some representative dialogues from both gainers and non-gainers in Table 12. For gainers’ dialogue with index 1, team 8 set a wrong goal that they need to achieve the cost of 34 after the submission since they do not seem to understand the meaning of the word minimum. They make a decision to start in the middle and go around it as future problem solving steps. Then team 8 correct their previous wrong goal clarification in the dialogue with index 4 and decide to find the route that costs a lot as they perhaps want to remove the route with high cost.

## 4. RESULTS

### 4.1 *RQ1: What is the relationship between the discussion and actions, and how is this relationship different between gainers and non-gainers??*

The results of the DSM between the gainers and non-gainers action and speech sequences show that there are 12 patterns that are only frequent among gainers as shown in Table 4 and 17 patterns that are only frequent among non-gainers as shown in Table 5. Four patterns are frequent among both gainers and non-gainers, but occur more often among gainers as shown in Table 6 and five patterns are frequent among both gainers and non-gainers, but occur more often among non-gainers as shown in Table 7. Note that in the interest of the space, where there are several patterns we report only the top-10 patterns here and the full list is available in the appendix. In the following we elaborate on the obtained patterns; while there were several interesting patterns, we focus only on patterns which contain speech and actions together as our interest is on action-discussion transactivity.

As shown in Table 4, 92% (11/12) frequent patterns of gainers start with high speech, and 45% (5/11) of them are actions of adding edges with high speech level. The mean

i-support of HS\_Hist (clicking on the review history button with high level of speech) of gainers is 4.00 which is almost 6 times as much as that of non-gainers (0.67). This suggests that compared with non-gainers, gainers tend to review history along with long periods of discussion more frequently. Further, the mean i-support of HS\_Remove (remove an edge with high levels of speech) of gainers is 3.54 which is more than twice that of non-gainers (1.50). This indicates that gainers remove an edge (a reflection action) along with high discussion more frequently than non-gainers.

For speech level related patterns that are frequent only among non-gainers as shown in Table 5 or more frequent among non-gainers (Table 7), barring 2 patterns in all the other patterns either the action of adding an edge or no action happens along with low level of speech. This suggests that compared with gainers, non-gainers tend to add edges more frequently and don’t do as many reflection related actions (click review history button and remove edge) frequently, and that their actions are accompanied by low/medium levels of speech.

For the DSM of speech overlap level sequences, there are 16 patterns that are only frequent among gainers (Table 8) and 25 patterns that are only frequent among non-gainers (Table 9). Besides, seven patterns are frequent among both gainers and non-gainers, but occur more often among gainers as shown in Table 10 and four patterns are frequent among both gainers and non-gainers, but occur more often among non-gainers as shown in Table 11.

From the patterns that are only frequent among gainers (Table 8), we see that the mean i-support of the pattern high level of speech overlap while clicking review history button (HSo\_Hist) among gainers (4.08) is more than twice the mean i-support among non-gainers (1.5). This pattern indicates that gainers have high level of speech overlap (intersecting speech) with the action of clicking the review history button. High level of speech overlap along with removing an edge (HSo\_Remove) is also frequent only among gainers. This indicates that gainers more frequently have high level of overlapping speech while doing reflective actions such as reviewing history or removing an edge.

There is however one frequent pattern with reflective behaviours seen only among non-gainers in Table 15: medium level of speech overlap with clicking review history button, followed by low level of speech overlap without any meaningful action ([MSo Hist, LSo NA]). Compared with the pattern [HSo Hist, HSo NA] (see Table 14) that is only frequent among gainers, the difference is the level of speech overlap. We may infer that because non-gainers communicate less when they click the review history button, they perhaps take away less information from the history (reflection) than gainers and we examine this in depth in the next section.

To summarize the above findings, gainers perform more reflection related actions (review history, remove edges) along with higher level of speech/speech overlap compared with non-gainers. Therefore, gainers reflected more via discussion and improved their solutions based on previous solutions continuously.

**Table 4: Patterns related to speech level and meaningful actions frequent only in gainers**

| frequent_pattern               | p_value   | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support |
|--------------------------------|-----------|-----------------------|---------------------------|---------------------|
| ['HS_NA', 'HS_NA']             | 0.003     | 9.96                  | 1.67                      | 8.3                 |
| ['HS_Add', 'HS_Add']           | 3.4E-04   | 5.69                  | 0.83                      | 4.86                |
| ['HS_NA', 'HS_Add']            | 1.960E-04 | 5.19                  | 0.5                       | 4.69                |
| ['HS_Add', 'HS_NA']            | 1.6E-04   | 4.39                  | 0.67                      | 3.72                |
| ['HS_NA', 'HS_NA', 'HS_NA']    | 0.017     | 4.0                   | 0.67                      | 3.3                 |
| ['HS_Hist']                    | 3.4E-03   | 4.0                   | 0.67                      | 3.3                 |
| ['HS_Remove']                  | 0.048     | 3.54                  | 1.5                       | 2.03                |
| ['HS_Add', 'MS_Add']           | 1.7E-04   | 3.23                  | 0.67                      | 2.56                |
| ['MS_Add', 'HS_Add']           | 1.7E-03   | 3.1                   | 0.67                      | 2.41                |
| ['HS_NA', 'HS_NA', 'MS_NA']    | 8.5E-03   | 1.73                  | 0.33                      | 1.4                 |
| ['HS_Add_Hist']                | 0.028     | 1.69                  | 0.5                       | 1.19                |
| ['HS_Add', 'HS_Add', 'MS_Add'] | 7.7E-05   | 0.96                  | 0.0                       | 0.96                |

**Table 5: Top 10 patterns related to speech level and meaningful action frequent only in non-gainers**

| frequent_pattern                      | p_value   | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support |
|---------------------------------------|-----------|-----------------------|---------------------------|---------------------|
| ['LS_Add', 'LS_NA']                   | 5.441E-03 | 0.92                  | 5.17                      | -4.25               |
| ['LS_Add', 'LS_Add']                  | 1.173E-03 | 1.8                   | 5.0                       | -3.2                |
| ['LS_NA', 'LS_Add']                   | 1.341E-02 | 1.0                   | 4.17                      | -3.17               |
| ['LS_Add', 'LS_NA', 'LS_NA']          | 4.124E-02 | 0.19                  | 2.5                       | -2.31               |
| ['LS_NA', 'LS_NA', 'MS_NA']           | 2.759E-02 | 0.46                  | 2.33                      | -1.87               |
| ['LS_Add', 'LS_Add', 'LS_Add']        | 9.626E-03 | 0.62                  | 2.17                      | -1.55               |
| ['LS_NA', 'LS_NA', 'MS_Add']          | 1.513E-02 | 0.73                  | 2.0                       | -1.27               |
| ['LS_NA', 'LS_NA', 'LS_Add']          | 2.554E-02 | 0.35                  | 1.83                      | -1.48               |
| ['LS_NA', 'LS_Add', 'LS_NA']          | 3.035E-02 | 0.23                  | 1.5                       | -1.27               |
| ['LS_NA', 'LS_NA', 'LS_NA', 'LS_Add'] | 1.418E-02 | 0.27                  | 1.5                       | -1.23               |

#### 4.1.1 Discussions while performing actions or when pausing actions

We are specifically interested in whether actions are performed with speech or whether actions are paused during speech and the difference between gainers and non-gainers in this regard. So, we calculate the percentage of no useful action (NA) happening among frequent patterns in each category we get from DSM. For speech level related patterns, 38.36% of patterns that are only frequent among gainers do not have any meaningful action (NA) and 30.00% that are more frequent among gainers do not have any useful action (NA). On the other hand, 57.28% of patterns that are only frequent among non-gainers are without any useful action (NA) and 61.54% that are more frequent among non-gainers are without any useful action (NA). For speech overlap level related patterns, 33.62% of patterns that are only frequent among gainers do not have any useful action (NA) and 36.00% that are more frequent among gainers do not have any useful action (NA). While 54.24% of patterns that are only frequent among non-gainers are without any useful action (NA) and 57.89% that are more frequent among non-gainers are without any useful action (NA). These results indicate that speech and action occurring concurrently is more frequent among gainers than non-gainers.

#### 4.2 RQ2: What is qualitative nature of speech that occurs along with actions of interest?

##### 4.2.1 Identifying relevant episodes of interests

In [18], it was suggested that 1) speech overlap was one of behaviours which discriminated gainers from non-gainers in

this context 2) students dialogue during episodes of speech overlap helped them build an understanding towards a solution and 3) gainers had more reflective actions than non-gainers. Therefore to explore the difference in the nature of the speech that occurs with actions, we choose reflective actions (specifically, reviewing history) as our action of interest. In our analysis above, we have identified some frequent, history related patterns that can serve as the relevant episodes of interest to perform the content analysis. Among those patterns, we pick [MSo\_Hist, LSo\_NA] which is only frequent among non-gainers and [HSo\_Hist, HSo\_NA] which is only frequent among gainers as the relevant episodes of interest for non-gainers and gainers, respectively (see appendix). In the current analysis, we focus on specific instances of when such speech overlap behaviors occur in conjunction with a action of reviewing history. The aforementioned two frequent patterns have the same time window length and similar actions but with different levels of speech overlap. An in-depth analysis of the content of dialogues happening during and around those episodes can help us better understand reflection behaviours of gainers and non-gainers, and any difference between them.

##### 4.2.2 Content Analysis of Dialogues

We only have dialogue transcripts for a subset of teams (10) in JUSThink Dialogue and Actions Corpus[20]. After matching transcripts with relevant episodes, we get 13 dialogues from four gainer teams (teams 7, 8, 9, and 47) and 4 dialogues from two non-gainer teams (teams 18, 20). Out of these 17 dialogues, 4 dialogues (24%) were analysed together by the first three authors of the paper until there

**Table 6: Frequent patterns related to speech level and exact meaningful action among both groups, but occurring more often in gainers**

| frequent_pattern    | p_value   | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support |
|---------------------|-----------|-----------------------|---------------------------|---------------------|
| ['HS_NA']           | 4.858E-03 | 2.9                   | 1.1                       | 1.8                 |
| ['HS_Add']          | 1.475E-06 | 20                    | 4.5                       | 15.5                |
| ['HS_Add_Remove']   | 1.678E-02 | 3.5                   | 1.17                      | 2.33                |
| ['MS_NA', 'HS_Add'] | 1.138E-02 | 2.1                   | 0.8                       | 1.3                 |

**Table 7: Frequent patterns related to speech level and exact meaningful action among both groups, but occurring more often in non-gainers**

| frequent_pattern     | p_value   | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support |
|----------------------|-----------|-----------------------|---------------------------|---------------------|
| ['LS_NA']            | 1.735E-02 | 10.15                 | 36.50                     | -26.35              |
| ['LS_Add']           | 4.617E-03 | 6.69                  | 17.67                     | -10.98              |
| ['LS_NA', 'LS_NA']   | 4.460E-02 | 2.77                  | 16.17                     | -13.4               |
| ['MS_Hist']          | 1.393E-02 | 3.19                  | 5.33                      | -2.14               |
| ['MS_Hist', 'MS_NA'] | 7.960E-03 | 0.92                  | 2.17                      | -1.25               |

was complete agreement on the coding scheme. After these the remaining dialogues were analysed by one of the three researchers.

The codes for two exemplar gainer and non-gainer team dialogues are shown in Table 12. We began the analysis by summarizing the content of the dialogue. This was followed by coding for the negotiation mechanisms based on Baker’s model of negotiation [1]. Finally, we coded for specific instances of reflection on past actions, planning for future actions and agreement because it is known that these shared regulation processes are necessary for collaborative problem-solving and learning [10].

The difference between the gainer and non-gainer teams is seen from their dialogues in Table 12. For instance, from the dialogues of non-gainer team 18 we see that each team member talks less compared with gainer team’s dialogues. In dialogue with index 0, team 18 compares their previous solution with the current solution, but only speaker A performs some reflection and no one proposes any further steps to solve the problem. In the dialogue with index 1, team 18 discusses the result they get from their submission and decide to start over directly. Speaker A is still the only one who proposes ideas and B just follows A’s requests. Further speaker A does not give any reason why he/she proposes those routes.

On the other hand in gainer team 8, two team members talk about their ideas actively. They always reflect on the previously submitted solution and clearly state a current problem solving strategy. For non-gainer team 18, they do not share their ideas as only one team member talks about his/her idea, and they reflect minimally on their previous solution. Compared with gainer team 8, non-gainer team 18 does not specify any further step to take in any episode.

Apart from these four representative dialogues, we conducted content analysis on all the available dialogues around relevant episodes for gainers (13 in total) and for non-gainers (4 in total). We found that for gainers’ dialogues around relevant episodes, 58.3% (7/12) of them contain goal clarifications. Apart from dialogues around episodes 10 and 11,

all the other dialogues -83.3% (10/12) dialogues - show some reflections from the past solutions. 91.7% (11/12) dialogues include making some decisions to take further steps based on past solutions. Offer-Acceptance happens more than twice in nearly half (5/12) of the dialogues.

In contrast, 25% (1/4) non-gainers’ dialogues include goal clarifications. Only one non-gainer team takes some reflections from the previous solution and it is a wrong reflection. Only half of the dialogues contain decisions to take some steps for the future. There is no episode where offer-acceptance happens more than twice among non-gainers.

Our results are limited because of a skew in terms of much fewer numbers of non-gainer team dialogues than gainer team dialogues. Still, to summarize the findings of our content analysis, we note that gainers on average have more productive communication along with actions, because approximately half of their dialogues reached more than two agreements within 60 seconds as compared to none of the non-gainer dialogues. Gainers also tend to reflect more on past solutions and make timely decisions for future actions as compared to non-gainers.

## 5. DISCUSSION AND CONCLUSION

In this paper, we investigate the relationship between speech and actions, as well as the qualitative nature of the speech that occurs with the actions. Our first RQ was related to identifying the relationship between speech and actions. To answer this RQ, we applied differential sequence mining (DSM) to differentiate frequent patterns between gainers and non-gainers. We found that gainers and non-gainers demonstrate different relationships between speech and actions. Gainers perform all types of actions (solution building and reflective) along with high levels of speech/overlapping speech more frequently than non-gainers. While previous research indicated that gainers speak more [18], our findings nuance those findings by suggesting that speaking *while* performing actions is productive for learning. Our findings align with previous findings related to the concurrency of actions and speech among more collaborative groups [15]; however our findings extend to groups which learned more and are in a different collaborative scripted context. Gainers also show

**Table 8: Top 10 patterns related to speech overlap level and exact meaningful action frequent only in gainers**

| frequent_pattern                | p_value  | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support |
|---------------------------------|----------|-----------------------|---------------------------|---------------------|
| ['HSO_Add', 'HSO_Add']          | 7.9e-05  | 6.7                   | 0.5                       | 6.2                 |
| ['HSO_NA', 'HSO_NA', 'HSO_NA']  | 0.006    | 5.3                   | 1.0                       | 4.3                 |
| ['HSO_NA', 'HSO_Add']           | 1.1e-05  | 4.58                  | 0.3                       | 4.2                 |
| ['HSO_Hist']                    | 0.04     | 4.08                  | 1.5                       | 2.58                |
| ['HSO_Add', 'MSO_Add']          | 0.002    | 3.08                  | 0.5                       | 2.58                |
| ['HSO_Add', 'MSO_NA']           | 9.09e-09 | 2.73                  | 0.17                      | 2.56                |
| ['HSO_NA', 'MSO_Add']           | 0.0005   | 2.66                  | 0.67                      | 1.99                |
| ['MSO_NA', 'HSO_Add']           | 3.08e-06 | 2.66                  | 0.33                      | 2.32                |
| ['HSO_Add_Hist']                | 5.28e-05 | 2.15                  | 0.0                       | 2.15                |
| ['HSO_Add', 'HSO_NA', 'HSO_NA'] | 0.0004   | 1.46                  | 0.0                       | 1.46                |

**Table 9: Top 10 patterns related to speech overlap level and exact meaningful action frequent only in non-gainers**

| frequent_pattern                         | p_value | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support |
|------------------------------------------|---------|-----------------------|---------------------------|---------------------|
| ['LSO_NA', 'LSO_NA', 'LSO_NA']           | 0.04    | 1.07                  | 6.66                      | -5.58               |
| ['LSO_Add', 'LSO_Add']                   | 0.00    | 1.34                  | 5.83                      | -4.48               |
| ['LSO_Add', 'LSO_NA']                    | 0.00    | 0.65                  | 5.66                      | -5.01               |
| ['LSO_Remove']                           | 0.03    | 1.0                   | 4.5                       | -3.5                |
| ['LSO_NA', 'LSO_Add']                    | 0.01    | 0.84                  | 4.5                       | -3.65               |
| ['LSO_NA', 'MSO_Add']                    | 0.00    | 0.65                  | 3.33                      | -2.67               |
| ['LSO_NA', 'LSO_NA', 'LSO_NA', 'LSO_NA'] | 0.03    | 0.5                   | 3.33                      | -2.83               |
| ['MSO_NA', 'LSO_Add']                    | 0.00    | 0.88                  | 2.83                      | -1.94               |
| ['MSO_Add', 'LSO_NA']                    | 0.01    | 0.69                  | 2.66                      | -1.97               |
| ['LSO_Add', 'LSO_NA', 'LSO_NA']          | 0.02    | 0.19                  | 2.33                      | -2.14               |

longer patterns of continuous speech indicating that they communicate more actively since they usually have medium and high levels of speech/speech overlap while non-gainers often have low and medium levels of speech/speech overlap.

To look deeper into the speech that happens along with actions (RQ2), we performed content analysis on dialogues around the episodes of interest which are identified based on the DSM results. We already know that gainers tend to access the history (reflect) more [18], however here we additionally find that gainers share the information and understanding obtained from the reflective actions to a greater degree as they review their history with higher level of speech or speech overlap compared with non-gainers, and decide on future steps towards the goal. Perhaps non-gainers are unable to extract the needed information from their past solutions i.e, their reflective actions, which is why they do not discuss as much during the episodes of interest and do not arrive at a consensus regarding next steps. This suggests that some additional scaffolding is needed within the environment to point out to students what they should observe from the history. Another significant point of difference between gainers and non-gainers dialogue is that gainers clarify the goal of the task more frequently, which is a sign that the action of reviewing history is being used to ground their shared understanding of the task [2]. While DSM only shows us that some patterns are more frequent in one group vs the other, the qualitative analyses elaborate on how the patterns are different between the groups in terms of the content of their dialogue. The above findings, together with previous literature which suggests that elaborative discussions lead to learning in a collaborative scenario [26, 24], points to the fact that it is the nature of discussions during these differentiating patterns that could be reason for the difference in

the learning between the groups.

Our presented approach combining DSM and qualitative analysis allows us to illustrate the importance of action-discussion transactivity in collaborative learning, and identify the nature of the discussion that can build on certain actions and make them productive. The qualitative analyses of the patterns unpacks the nature of the action-discussion patterns in each group. Specifically, we identify that the gainer groups do more elaborative, reflective and planning discussions which build on the history check action, compared to the non-gainer group. In other words, gainers had a greater degree of action-discussion transactivity, because they articulated their ideas and information obtained from doing the history check action, which helped them progress in the solution building. Compared to previous work [15, 23], our findings highlight the nature of the actions and discussions occurring together, how they build on each other and their association with collaborative learning, as opposed to the quality of collaboration or task performance. To summarize, our findings suggest that those who learned had a greater degree of action-discussion transactivity, and that they more frequently articulated their ideas and information obtained from doing actions, which helped them progress in the solution building. Taken together with previous literature in collaborative learning (eg. [21]) which speaks about the necessity of elaboration and transactivity in discussions and actions, our findings indicate that students should be encouraged to articulate their ideas or information obtained from doing actions and a teacher or a scaffold build into the CSCL environment can prompt the students to do so. Regulation of their performance and learning is challenging for students and researchers have proposed technological tools to support students [11]. This work provides sugges-

**Table 10: Patterns related to speech overlap level and exact meaningful action frequent among both groups, but occurring more often in gainers**

| frequent_pattern       | p_value  | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support |
|------------------------|----------|-----------------------|---------------------------|---------------------|
| ['HSo_NA']             | 0.00     | 31.69                 | 12.66                     | 19.02               |
| ['HSo_Add']            | 2.41e-08 | 21.53                 | 3.0                       | 18.53               |
| ['HSo_NA', 'HSo_NA']   | 0.00     | 11.5                  | 3.0                       | 8.5                 |
| ['HSo_Remove']         | 0.00     | 4.38                  | 1.33                      | 3.05                |
| ['HSo_Add', 'HSo_NA']  | 7.17e-06 | 4.26                  | 0.66                      | 3.60                |
| ['HSo_Add_Remove']     | 0.00     | 4.0                   | 1.0                       | 3.0                 |
| ['MSo_Add', 'HSo_Add'] | 0.00     | 3.0                   | 0.83                      | 2.16                |

**Table 11: Patterns related to speech overlap level and exact meaningful actions among both groups, but occurring more often in non-gainers**

| frequent_pattern     | p_value | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support |
|----------------------|---------|-----------------------|---------------------------|---------------------|
| ['LSo_NA']           | 0.00    | 9.0                   | 37.33                     | -28.33              |
| ['LSo_Add']          | 0.00    | 6.26                  | 21.0                      | -14.73              |
| ['LSo_NA', 'LSo_NA'] | 0.03    | 2.57                  | 15.0                      | -12.42              |
| ['MSo_NA', 'LSo_NA'] | 0.04    | 2.5                   | 7.16                      | -4.66               |

tions regarding when such tools can be most productive for students, for instance, prompts for goal clarification after a failed problem-solving attempt.

Our analysis in this paper is limited from the following aspects. The number of non-gainer teams is much lesser than gainer teams, and as a result the number of non-gainer team dialogues is also much lesser than gainer team dialogues. However, the imbalance between gainers and non-gainers is due to the nature of the experiment setting - the experiment was designed to facilitate learning. Secondly, the interaction for each team of around 20-25 minutes is organized in windows of 10 seconds in the multimodal temporal dataset while dialogues start and end time are recorded as exact timestamps in the transcripts dataset. When we pick dialogues within identified relevant episodes, the difference in time features of the two datasets can cause slight inaccuracies of the matched results. To solve this problem, we pick up dialogues before and after 20 seconds of the relevant episode. Finally, the transcripts dataset does not include all teams. We have only analysed a subset of the dialogues from available transcripts. Our future work focuses on obtaining more data to extend this approach to larger sets of gainers and non-gainers, and other actions of interest in collaborative learning.



## 6. REFERENCES

- [1] M. Baker. A model for negotiation in teaching-learning dialogues. *Journal of Interactive Learning Research*, 5(2):199, 1994.
- [2] M. Baker, T. Hansen, R. Joiner, and D. Traum. The role of grounding in collaborative learning tasks. *Collaborative learning: Cognitive and computational approaches*, 31:63, 1999.
- [3] B. Barron. When smart groups fail. *The journal of the learning sciences*, 12(3):307–359, 2003.
- [4] J. Davidsen and T. Ryberg. “this is the size of one meter”: Children’s bodily-material collaboration. *International Journal of Computer-Supported Collaborative Learning*, 12(1):65–90, 2017.
- [5] P. Dillenbourg, S. Järvelä, and F. Fischer. The evolution of research on computer-supported collaborative learning. In *Technology-enhanced learning*, pages 3–19. Springer, 2009.
- [6] P. Dillenbourg and D. Traum. Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *The Journal of the Learning Sciences*, 15(1):121–151, 2006.
- [7] A. C. Evans, J. O. Wobbrock, and K. Davis. Modeling collaboration patterns on an interactive tabletop in a classroom setting. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 860–871, 2016.
- [8] A. Harris, J. Rick, V. Bonnett, N. Yuill, R. Fleck, P. Marshall, and Y. Rogers. Around the table: Are multiple-touch surfaces better than single-touch for children’s collaborative interactions? In *CSCL (1)*, pages 335–344, 2009.
- [9] J. Ho, L. Lukov, and S. Chawla. Sequential pattern mining with constraints on large protein databases. In *Proceedings of the 12th international conference on management of data (COMAD)*, pages 89–100, 2005.
- [10] S. Järvelä and A. F. Hadwin. New frontiers: Regulating learning in cscl. *Educational psychologist*, 48(1):25–39, 2013.
- [11] S. Järvelä, P. A. Kirschner, A. Hadwin, H. Järvenoja, J. Malmberg, M. Miller, and J. Laru. Socially shared regulation of learning in cscl: Understanding and prompting individual-and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning*, 11(3):263–280, 2016.
- [12] H. Jeong and M. T. Chi. Construction of shared knowledge during collaborative learning. 1997.
- [13] J. S. Kinnebrew, K. M. Loretz, and G. Biswas. A contextualized, differential sequence mining method to derive students’ learning behavior patterns. *Journal of Educational Data Mining*, 5(1):190–219, 2013.
- [14] K. Krippendorff. Content analysis. 1989.
- [15] R. Martinez-Maldonado, Y. Dimitriadis, A. Martinez-Monés, J. Kay, and K. Yacef. Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. *International Journal of Computer-Supported Collaborative Learning*, 8(4):455–485, 2013.
- [16] S. Mishra, A. Munshi, M. Rushdy, and G. Biswas. Lasat: learning activity sequence analysis tool. In *Technology-enhanced & evidence-based education & learning (TEEL) workshop at the 9th international learning analytics and knowledge (LAK) conference, Tempe, Arizona, USA*, 2019.
- [17] J. Nasir, B. Bruno, and P. Dillenbourg. PE-HRI-temporal: A Multimodal Temporal Dataset in a robot mediated Collaborative Educational Setting, Nov. 2021.
- [18] J. Nasir, A. Kothiyal, B. Bruno, and P. Dillenbourg. Many are the ways to learn identifying multi-modal behavioral profiles of collaborative learning in constructivist activities. *International Journal of Computer-Supported Collaborative Learning*, 16(4):485–523, 2021.
- [19] J. Nasir, U. Norman, B. Bruno, and P. Dillenbourg. When positive perception of the robot has no effect on learning. In *2020 29th IEEE international conference on robot and human interactive communication (RO-MAN)*, pages 313–320. IEEE, 2020.
- [20] U. Norman, T. Dinkar, J. Nasir, B. Bruno, C. Clavel, and P. Dillenbourg. Justthink dialogue and actions corpus, Mar. 2021.
- [21] O. Noroozi, A. Weinberger, H. Biemans, M. Mulder, and M. Chizari. Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. 61, 2013.
- [22] J. K. Olsen, K. Sharma, N. Rummel, and V. Aleven. Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology*, 51(5):1527–1547, 2020.
- [23] V. Popov, A. van Leeuwen, and S. C. Buis. Are you with me or not? temporal synchronicity and transactivity during cscl. *Journal of Computer Assisted Learning*, 33(5):424–442, 2017.
- [24] J. Roschelle and S. D. Teasley. The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning*, pages 69–97. Springer, 1995.
- [25] J. R. Segedy, J. S. Kinnebrew, and G. Biswas. Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal of Learning Analytics*, 2(1):13–48, 2015.
- [26] S. D. Teasley. The role of talk in children’s peer collaborations. *Developmental psychology*, 31(2):207, 1995.
- [27] S. A. Viswanathan and K. VanLehn. Using the tablet gestures and speech of pairs of students to classify their collaboration. *IEEE Transactions on Learning Technologies*, 11(2):230–242, 2017.
- [28] F. Vogel, I. Kollar, S. Ufer, E. Reichersdorfer, K. Reiss, and F. Fischer. Developing argumentation skills in mathematics through computer-supported collaborative learning: The role of transactivity. *Instructional Science*, 44(5):477–500, 2016.
- [29] A. Weinberger and F. Fischer. A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & education*, 46(1):71–95, 2006.

## 7. APPENDIX

Table 13: Speech level patterns frequent only among non-gainers

| frequent_pattern | p-value   | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support | freq_decode_pattern                           |
|------------------|-----------|-----------------------|---------------------------|---------------------|-----------------------------------------------|
| [0, 21]          | 5.441E-03 | 9.231E-01             | 5.167E+00                 | -4.244E+00          | ['LS_Add', 'LS_NA']                           |
| [0, 0]           | 1.173E-03 | 1.808E+00             | 5.000E+00                 | -3.192E+00          | ['LS_Add', 'LS_Add']                          |
| [21, 0]          | 1.341E-02 | 1.000E+00             | 4.167E+00                 | -3.167E+00          | ['LS_NA', 'LS_Add']                           |
| [0, 21, 21]      | 4.124E-02 | 1.923E-01             | 2.500E+00                 | -2.308E+00          | ['LS_Add', 'LS_NA', 'LS_NA']                  |
| [21, 21, 22]     | 2.759E-02 | 4.615E-01             | 2.333E+00                 | -1.872E+00          | ['LS_NA', 'LS_NA', 'MS_NA']                   |
| [0, 0, 0]        | 9.626E-03 | 6.154E-01             | 2.167E+00                 | -1.551E+00          | ['LS_Add', 'LS_Add', 'LS_Add']                |
| [21, 1]          | 1.513E-02 | 7.308E-01             | 2.000E+00                 | -1.269E+00          | ['LS_NA', 'MS_Add']                           |
| [21, 21, 0]      | 2.554E-02 | 3.462E-01             | 1.833E+00                 | -1.487E+00          | ['LS_NA', 'LS_NA', 'LS_Add']                  |
| [21, 0, 21]      | 3.035E-02 | 2.308E-01             | 1.500E+00                 | -1.269E+00          | ['LS_NA', 'LS_Add', 'LS_NA']                  |
| [21, 21, 21, 0]  | 1.418E-02 | 2.692E-01             | 1.500E+00                 | -1.231E+00          | ['LS_NA', 'LS_NA', 'LS_NA', 'LS_Add']         |
| [0, 0, 21]       | 2.575E-02 | 3.077E-01             | 1.333E+00                 | -1.026E+00          | ['LS_Add', 'LS_Add', 'LS_NA']                 |
| [21, 0, 0, 0]    | 2.913E-02 | 2.308E-01             | 1.000E+00                 | -7.692E-01          | ['LS_NA', 'LS_Add', 'LS_Add', 'LS_Add']       |
| [21, 1, 22]      | 2.006E-02 | 1.538E-01             | 1.000E+00                 | -8.462E-01          | ['LS_NA', 'MS_Add', 'MS_NA']                  |
| [0, 0, 21, 21]   | 4.219E-02 | 0.000E+00             | 8.333E-01                 | -8.333E-01          | ['LS_Add', 'LS_Add', 'LS_NA', 'LS_NA']        |
| [21, 22, 21, 22] | 4.669E-02 | 1.154E-01             | 6.667E-01                 | -5.513E-01          | ['LS_NA', 'MS_NA', 'LS_NA', 'MS_NA']          |
| [21, 22, 22, 22] | 4.669E-02 | 1.154E-01             | 6.667E-01                 | -5.513E-01          | ['LS_NA', 'MS_NA', 'MS_NA', 'MS_NA', 'MS_NA'] |
| [0, 0, 0, 21]    | 3.716E-02 | 7.692E-02             | 6.667E-01                 | -5.897E-01          | ['LS_Add', 'LS_Add', 'LS_Add', 'LS_NA']       |

Table 14: Speech overlap level patterns only frequent among gainers

| frequent_pattern | p-value                | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support | freq_decode_pattern              |
|------------------|------------------------|-----------------------|---------------------------|---------------------|----------------------------------|
| [2, 2]           | 7.928934765535499e-05  | 6.730769230769231     | 0.5                       | 6.230769230769231   | ['HSO_Add', 'HSO_Add']           |
| [23, 23, 23]     | 0.006002839704649775   | 5.3076923076923075    | 1.0                       | 4.3076923076923075  | ['HSO_NA', 'HSO_NA', 'HSO_NA']   |
| [23, 2]          | 1.1371592327013828e-05 | 4.576923076923077     | 0.3333333333333333        | 4.243589743589744   | ['HSO_NA', 'HSO_Add']            |
| [8]              | 0.036533665925983824   | 4.076923076923077     | 1.5                       | 2.5769230769230766  | ['HSO_Hist']                     |
| [2, 1]           | 0.0022205587982335228  | 3.076923076923077     | 0.5                       | 2.576923076923077   | ['HSO_Add', 'MSO_Add']           |
| [2, 22]          | 9.087126987994823e-09  | 2.730769230769231     | 0.16666666666666666       | 2.5641025641025643  | ['HSO_Add', 'MSO_NA']            |
| [23, 1]          | 0.0005621716567395286  | 2.6538461538461537    | 0.6666666666666666        | 1.9871794871794872  | ['HSO_NA', 'MSO_Add']            |
| [22, 2]          | 3.082910493793244e-06  | 2.6538461538461537    | 0.3333333333333333        | 2.3205128205128203  | ['MSO_NA', 'HSO_Add']            |
| [20]             | 5.2819931032191796e-05 | 2.1538461538461537    | 0.0                       | 2.1538461538461537  | ['HSO_Add_Hist']                 |
| [2, 23, 23]      | 0.00043067995679562895 | 1.4615384615384615    | 0.0                       | 1.4615384615384615  | ['HSO_Add', 'HSO_NA', 'HSO_NA']  |
| [23, 23, 2]      | 0.0019868197228214823  | 1.4615384615384615    | 0.0                       | 1.4615384615384615  | ['HSO_NA', 'HSO_NA', 'HSO_Add']  |
| [2, 2, 23]       | 0.0003332574321561345  | 1.4615384615384615    | 0.0                       | 1.4615384615384615  | ['HSO_Add', 'HSO_Add', 'HSO_NA'] |
| [23, 2, 2]       | 0.0006605272364293755  | 1.3076923076923077    | 0.0                       | 1.3076923076923077  | ['HSO_NA', 'HSO_Add', 'HSO_Add'] |
| [8, 23]          | 0.0004105089572129998  | 1.2692307692307692    | 0.0                       | 1.2692307692307692  | ['HSO_Hist', 'HSO_NA']           |
| [5, 23]          | 0.0030286337078061225  | 1.1538461538461537    | 0.16666666666666666       | 0.9871794871794871  | ['HSO_Remove', 'HSO_NA']         |
| [22, 2, 2]       | 2.101257038827904e-05  | 0.8846153846153846    | 0.0                       | 0.8846153846153846  | ['MSO_NA', 'HSO_Add', 'HSO_Add'] |

Table 15: Speech overlap level patterns only frequent among non-gainers

| frequent_pattern    | p-value                | mean_gainer_i_support | mean_non_gainer_i_support | diff_mean_i_support | freq_decode_pattern                                 |
|---------------------|------------------------|-----------------------|---------------------------|---------------------|-----------------------------------------------------|
| [21, 21, 21]        | 0.04939554474004129    | 1.0769230769230769    | 6.666666666666667         | -5.58974358974359   | ['LSO_NA', 'LSO_NA', 'LSO_NA']                      |
| [0, 0]              | 0.0018586234032742692  | 1.3461538461538463    | 5.833333333333333         | -4.487179487179487  | ['LSO_Add', 'LSO_Add']                              |
| [0, 21]             | 0.00016706039378552398 | 0.6538461538461539    | 5.666666666666667         | -5.012820512820513  | ['LSO_Add', 'LSO_NA']                               |
| [3]                 | 0.035109050581296      | 1.0                   | 4.5                       | -3.5                | ['LSO_Remove']                                      |
| [21, 0]             | 0.013026456729087523   | 0.8461538461538461    | 4.5                       | -3.6538461538461537 | ['LSO_NA', 'LSO_Add']                               |
| [21, 1]             | 0.008836545666620073   | 0.6538461538461539    | 3.3333333333333335        | -2.6794871794871797 | ['LSO_NA', 'MSO_Add']                               |
| [21, 21, 21, 21]    | 0.03860139794859445    | 0.5                   | 3.3333333333333335        | -2.8333333333333335 | ['LSO_NA', 'LSO_NA', 'LSO_NA', 'LSO_NA']            |
| [22, 0]             | 0.006750888971567344   | 0.8846153846153846    | 2.8333333333333335        | -1.948717948717949  | ['MSO_NA', 'LSO_Add']                               |
| [1, 21]             | 0.01443329824984804    | 0.6923076923076923    | 2.6666666666666665        | -1.9743589743589742 | ['MSO_Add', 'LSO_NA']                               |
| [0, 21, 21]         | 0.023209371486681146   | 0.19230769230769232   | 2.3333333333333335        | -2.141025641025641  | ['LSO_Add', 'LSO_NA', 'LSO_NA']                     |
| [0, 0, 0]           | 0.009755427161716556   | 0.2692307692307692    | 2.0                       | -1.7307692307692308 | ['LSO_Add', 'LSO_Add', 'LSO_Add']                   |
| [0, 0, 21]          | 0.0056742454761588585  | 0.19230769230769232   | 1.6666666666666667        | -1.4743589743589745 | ['LSO_Add', 'LSO_Add', 'LSO_NA']                    |
| [7, 21]             | 0.013914091446642511   | 0.11538461538461539   | 1.3333333333333333        | -1.2179487179487178 | ['MSO_Hist', 'LSO_NA']                              |
| [21, 21, 1]         | 0.013914091446642511   | 0.11538461538461539   | 1.3333333333333333        | -1.2179487179487178 | ['LSO_NA', 'LSO_NA', 'MSO_Add']                     |
| [21, 1, 21]         | 0.014075354800622239   | 0.038461538461538464  | 1.1666666666666667        | -1.1282051282051282 | ['LSO_NA', 'MSO_Add', 'LSO_NA']                     |
| [0, 0, 21, 21]      | 0.011724811003954628   | 0.0                   | 1.0                       | -1.0                | ['LSO_Add', 'LSO_Add', 'LSO_NA', 'LSO_NA']          |
| [21, 0, 0, 0]       | 0.013173766481180184   | 0.038461538461538464  | 1.0                       | -0.9615384615384616 | ['LSO_NA', 'LSO_Add', 'LSO_Add', 'LSO_Add']         |
| [0, 0, 0, 21]       | 0.04085940385929584    | 0.0                   | 1.0                       | -1.0                | ['LSO_Add', 'LSO_Add', 'LSO_NA', 'LSO_NA']          |
| [22, 22, 0]         | 0.020519815735647172   | 0.2692307692307692    | 0.8333333333333334        | -0.5641025641025641 | ['MSO_NA', 'MSO_NA', 'LSO_Add']                     |
| [21, 1, 22]         | 0.0041047159800533225  | 0.0                   | 0.8333333333333334        | -0.8333333333333334 | ['LSO_Add', 'LSO_NA', 'MSO_NA']                     |
| [21, 21, 1, 21]     | 0.004307836785291385   | 0.038461538461538464  | 0.8333333333333334        | -0.7948717948717949 | ['LSO_NA', 'LSO_NA', 'MSO_Add', 'LSO_NA']           |
| [21, 21, 1, 21, 21] | 0.02503101581845297    | 0.0                   | 0.6666666666666666        | -0.6666666666666666 | ['LSO_NA', 'LSO_NA', 'MSO_Add', 'LSO_NA', 'LSO_NA'] |
| [21, 21, 22, 21]    | 0.04669295353054086    | 0.11538461538461539   | 0.6666666666666666        | -0.5512820512820512 | ['LSO_NA', 'LSO_NA', 'MSO_NA', 'LSO_NA']            |
| [21, 1, 21, 21]     | 0.02503101581845297    | 0.0                   | 0.6666666666666666        | -0.6666666666666666 | ['LSO_NA', 'MSO_Add', 'LSO_NA', 'LSO_NA']           |
| [0, 21, 21, 21]     | 0.030127010101375896   | 0.038461538461538464  | 0.6666666666666666        | -0.6282051282051282 | ['LSO_Add', 'LSO_NA', 'LSO_NA', 'LSO_NA']           |